

People adopt optimal policies in simple decision-making, after practice and guidance.

Nathan J. Evans and Scott D. Brown
School of Psychology, University of Newcastle, Australia

Abstract

Organisms making repeated simple decisions are faced with a tradeoff between urgent and cautious strategies. While animals can adopt a statistically optimal policy for this tradeoff, findings about human decision-makers have been mixed. Some studies have shown that people can optimize this “speed-accuracy tradeoff”, while others have identified a systematic bias towards excessive caution. These issues have driven theoretical development and spurred debate about the nature of human decision-making. We investigated a potential resolution to the debate, based on two factors that routinely differ between human and animal studies of decision-making: the effects of practice, and of longer-term feedback. Our study replicated the finding that most people, by default, are overly cautious. When given both practice and detailed feedback, people moved rapidly towards the optimal policy, with many participants reaching optimality with less than one hour of practice. Our findings have theoretical implications for cognitive and neural models of simple decision-making, as well as methodological implications.

Keywords: decision-making; speed-accuracy tradeoff; optimality; reward rate.

Introduction

The statistical optimality of theories of decision-making has been of interest for more than 50 years, since the random walk model was identified with a statistically optimal decision algorithm known as the sequential probability ratio test (Wald & Wolfowitz, 1948). More recently, considerations of statistical optimality have driven theoretical developments in understanding the cognitive and neurophysiological mechanisms of decision-making (Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Simen et al., 2009; Drugowitsch, Moreno-Bote, Churchland, Shadlen, & Pouget, 2012; Thura, Beauregard-Racine, Fradet, & Cisek, 2012; Gold & Shadlen, 2002).

There are many reasonable definitions of optimality. For example, a student taking a multiple choice exam probably aims to optimize the number of correct responses, although they may instead want to optimize the probability of not failing, which is not the same. One useful definition concerns the optimal policy for balancing the competing demands for caution and urgency in decision-making. In a wide range of tasks, an organism can choose to make more careful decisions, which will lead to more correct decisions but which will also take longer, or choose to make urgent decisions, which are faster but less accurate. This creates a “speed-accuracy tradeoff” whereby both very urgent and very cautious decisions lead to poorer outcomes than some intermediate balance. This balance is most often quantified by measuring reward rate, which is just the rate at which correct decisions outcomes are produced.

For many animals, defining optimality according to maximum reward rate makes sense. Consider a squirrel who has one hour to look for acorns (or, equivalently, a laboratory rat who is tested for one hour). The squirrel could choose to examine 20 different hiding spots for acorns each very carefully, and find the acorn hiding in every single one – this policy yields 20 acorns. Alternatively, she could examine 40 different hiding spots each very quickly, and find the acorn hiding in just half of the spots – also yielding 20 acorns. Between these extremes, there may be a better outcome; perhaps examining 30 hiding spots each with moderate care will lead to a success rate of 75%, yielding 22.5 acorns, on average.

Reward rate has been extensively investigated using the dominant theoretical framework for decision-making, evidence accumulation. Evidence accumulation models of decision-making have successfully explained behavioral and neural data in great detail (Ratcliff, 1978; Brown & Heathcote, 2008; Ratcliff & Rouder, 1998). As shown in the top panel of Figure 1, evidence accumulation models, such as the random walk, assume that decisions are made by slowly gathering evidence from the environment about the different possible decision outcomes. When sufficient evidence is gathered in favor of one outcome over the others, a decision is triggered. The amount of evidence required to trigger a decision is known as the “decision threshold”, and is key to explaining the speed-accuracy tradeoff. A cautious decision-maker sets a high threshold, requiring a lot of evidence before making a decision. This means that decisions take a long time, because the decision-maker

must wait for a lot of evidence to arrive, but that many incorrect decisions are avoided. Conversely, a low decision threshold leads to fast decisions which are more frequently incorrect, when they are driven by random fluctuations in the environmental evidence (grey lines in the middle panel of Figure 1). Setting the decision threshold too high or too low will reduce reward rate, and reward rate is optimized by some intermediate setting (lower panel of Figure 1).

Laboratory animals have been found to adopt their speed-accuracy tradeoff strategy to one that optimizes their reward rate. For example, in environments with randomly intermixed decision difficulties, a fixed threshold cannot yield as high a reward rate as a collapsing threshold (Drugowitsch et al., 2012; Thura et al., 2012; Ditterich, 2006), and Hawkins, Forstmann, Wagenmakers, Ratcliff, and Brown (2015) observed that non-human primates successfully implemented such a collapsing threshold. Whether human decision-makers adopt their decision strategies in an optimal manner is less clear. Simen et al. (2009) found that people appropriately identified situations in which the optimal decision threshold changed towards more cautious or more urgent policies, and people followed the direction of these changes in response to manipulations of experiment timing, stimulus contingencies, and rewards. However, changes in task difficulty have shown the opposite trend, with both Starns and Ratcliff (2012) and Balci et al. (2011) observing that participants adopted cautious strategies with increasing difficulty, where the optimal adjustment would dictate being less cautious.

In terms of the quantitative separations between the adopted threshold and the optimal threshold values, Simen et al. (2009) found that participants were optimal in conditions with slow stimulus timing, but not with faster paced tasks that required lower thresholds. This is consistent with the idea that people adopt quite cautious decision thresholds, as those thresholds are optimal in slow tasks, but not fast tasks. Similarly, Balci et al. (2011), observed participants to adopt optimal thresholds for easy decisions, but to be sub-optimally over-cautious in difficult conditions. However, due to interpretation issues associated with multiple difficulties that are explained below, they were unable to conclusively identify whether participants were adopting a single sub-optimal threshold, or multiple thresholds where some were optimal, and some were sub-optimal. Starns and Ratcliff (2012) made similar observations for younger participants and older participants. A difficulty with Starns and Ratcliff's findings is that participants were not explicitly instructed to optimize reward rate, which leaves open the question of whether people cannot perform optimally, or simply do not do so by default. For example, when Hawkins, Brown, Steyvers, and Wagenmakers (2012) did not give participants explicit instructions to optimize reward rate, they found that the participants instead minimized the time they spent in the experiment, while meeting a minimum performance goal.

One possible explanation for these human-animal differences appeals to experimental design. Decision-making experiments with animals are frequently designed in such a way that the animal is directly encouraged to maximize reward rate, but this is not the case for human experiments. For example, decision-making experiments with monkeys (Roitman

& Shadlen, 2002) typically begin with a thirsty animal who is rewarded for each correct decision with a drop of juice. The duration of the experimental session is fixed by the experimenter and so the monkey is most able to satisfy its needs by maximizing the number of correct decisions per unit time – that is, reward rate. In usual human decision-making experiments, the total number of correct decisions is less salient than for the monkey. The participant usually leaves the experiment whenever they have made a fixed number of decisions, rather than when a fixed amount of time elapses. Starns and Ratcliff (2012) investigated this methodological explanation by directly comparing performance in blocks where the duration was controlled by the experimenter versus by the participant. Consistent with animal research, Starns et al. found that experimenter-controlled blocks led participants to adopt speed-accuracy tradeoff settings that were closer to optimal, although these settings were still decidedly sub-optimal in most conditions.

We investigated whether people were able to adopt speed-accuracy tradeoff policies which optimized reward rate. We made three changes from previous studies. Firstly, we provided some participants with longer-term feedback about their performance, designed to mimic the kind of feedback hungry animals might experience. Those animals, and animals in decision-making studies, are rewarded by food and experience appetitive feedback about their foraging success – adopting a sub-optimal policy will lead to a lower reward rate and greater hunger. We informed decision-makers about their reward rate in recent blocks, analogous to the amount of food rewards earned in recent minutes, for animal studies. In addition, to assess whether human participants were capable adjusting to the optimality strategy under the more basic condition of where the amount of adjustment needed was known, we gave some participants feedback about how they could perform better, in addition to feedback about their general performance.

Secondly, we restricted our experimental design to use the same nominal difficulty level for all decision stimuli. The use of multiple difficulty levels presents problems in the theoretical interpretation of optimality. When all decisions are of a fixed difficulty level, the standard random walk or diffusion models of decision-making are statistically optimal (Wald & Wolfowitz, 1948). When decision difficulty varies, the situation is less clear. In those cases, reward rate can be increased beyond the maximum rate attainable with a standard decision-making model via the adoption of time-varying decision boundaries (Thura et al., 2012; Drugowitsch et al., 2012). Starns and Ratcliff (2012) separated the different difficulties into different blocks, which potentially allow different, optimal tradeoff policies to be adopted for each separate difficulty. While possible, this presumably increases the cognitive load, and may have contributed to the difficulty their participants had in setting reward-rate optimal decision policies. In addition, findings from this paradigm have been mixed: sometimes participants adopted multiple thresholds that are all sub-optimal; sometimes researchers were unsure whether participants adopted a single threshold which is sub-optimal compared with multiple thresholds, or multiple thresholds which are optimal for some conditions but not others (Balci et al., 2011); and sometimes participants adopt multiple thresholds that optimize a different goal than reward rate (Hawkins et al., 2012).

Thirdly, we directly investigated the effects of practice on reward rate. Participants in animal studies typically undertake many thousands of practice decisions before data collection begins. This extended practice is used by the experimenter to shape behavior, but it also affords the animals opportunity to learn about the reward-rate consequences of different speed-accuracy tradeoff settings, through trial-and-error. We compared performance across 24 blocks of trials. Both Starns and Ratcliff (2010) and Balci et al. (2011) found that participants moved closer to optimality with practice, from an initial over-emphasis on accuracy. However, the results are difficult to interpret very clearly, because of the above-mentioned inter-mixed difficulty conditions. Nevertheless, both of these results show the promise of investigating the effects of practice on reward rate optimality, suggesting that participants can approximate the best possible fixed threshold with adequate practice, though it is not necessarily the optimal strategy in the scenario.

Finally, we also investigated the effect of having the task duration controlled by either the experimenter (“fixed time blocks”) or by the participant (“fixed trial blocks”). Starns and Ratcliff (2012) also investigated this same manipulation, and we hoped to replicate their finding that fixed time blocks lead to better reward rate optimization.

Method

Participants

Eighty five participants from the University of Newcastle were randomly assigned into one of six groups formed by the factorial combination of two factors: fixed time vs. fixed trial blocks; and three different levels of longer-term feedback detail. After data collection, we excluded all participants whose response accuracy was less than our minimum criterion of 70%, resulting in 26 participants being excluded (see also Table 1). This criterion was chosen before having observed the data, as a means to eliminate participants who did not appear to properly understand or respond to the task requirements. Although this may seem like a relatively large number of participants to exclude, follow-up analyses which included all participants made no substantial difference (see Supplementary Materials for full disclosure).

Task and procedure

We used a motion discrimination task with a random dot kinematogram (Roitman & Shadlen, 2002). On each trial, participants were shown 40 dots in the center of the screen, and asked to judge whether these dots appeared to be moving mostly towards the top-left, or top-right, with each type of trial occurring 50% of the time. The dots were 3×3 pixels and always remained in a circle 150 pixels in diameter – any dot leaving this circle was randomly re-placed within it. For each trial, the coherent motion direction was chosen randomly to be either towards the top-left or top-right. On each frame, occurring every 66.7ms, 4 dots (i.e. 10% coherence) were randomly sampled and moved coherently by 3 pixels upwards and 3 pixels left or right, depending on the coherent motion direction

Table 1: The number of participants in each group and an example of the feedback received by each feedback group after each block. The different block types (fixed-time and fixed-trial) are placed in the columns, and the different feedback types (low, medium, high) are in the rows. As feedback was the same for both fixed-time and fixed-trial participants, the columns are merged for this section. The number in brackets is the number of participants who scored more than 70% accuracy, and so entered the main analyses.

		Block type		
		Time	Trial	
Feedback type	Low	<i>Participants</i>	16 (8)	13 (10)
		<i>Example feedback</i>	"End of block. Please take a short break."	
	Medium	<i>Participants</i>	16 (11)	13 (9)
		<i>Example feedback</i>	"In recent blocks you've attained 120 points in 4.057 minutes, meaning you achieved 29.582 points per minute."	
	High	<i>Participants</i>	14 (10)	13 (11)
		<i>Example feedback</i>	"In recent blocks you've attained 120 points in 4.057 minutes, meaning you achieved 29.582 points per minute. However, you could potentially improve your performance. The average time you spent on each trial was 1.283ms; if you were to go 12.942% faster on each trial, you would only lose 5.522% accuracy, resulting in 0.483 more points per minute."	

for that trial. The other 36 dots all moved in a random direction. Participants responded using the “z” and “/” keys. Participants were given feedback for 300ms on whether each response was correct. Incorrect responses were accompanied by a further 500ms timeout. Feedback was followed by an inter-trial-interval of 100ms before the start of the next trial. The screen resolution was 1920 x 1080 pixels, and the screen size was approximately 58cm.

Participants completed 24 blocks of this task. For the fixed trial groups, there were 40 trials in each block. For the fixed time groups, each block lasted one minute. On average, those in the fixed time groups completed 34.4 trials per block, and those in the fixed trial groups completed each block in 62.3 seconds. Those in the fixed time groups were told that their goal in the task was to earn as many points (i.e., correct answers) as possible. They were also informed that the number of trials would depend on their speed at the task. Those in the fixed trial groups were told that their goal in the task was to earn as many points as possible in each one-minute period, in order to reflect the goal of reward rate optimality.

Participants in the low feedback groups received only the trial-by-trial feedback on correctness outlined above. Participants in the medium feedback groups received the same trial-by-trial feedback as above, and additionally were given information on their average accuracy and response time for the previous 200 trials, after each of blocks 4–23 (or all previous trials, if they had completed fewer than 200). Participants in the high feedback groups were given all the feedback that the medium feedback groups received, plus extra guidance on how they could change their speed-accuracy tradeoff to achieve reward rate optimality. The optimal reward rate was calculated by estimating the parameters of the E-Z diffusion model (Wagenmakers, Van Der Maas, & Grasman, 2007), and then calculating optimal threshold conditional on those parameters (see Supplementary Materials for details). These levels of feedback allowed us to judge whether participants were able to freely achieve optimality with no information about their general performance or the optimal performance strategy (low), with information about only their general performance (medium), or with feedback about both (high). Examples of feedback given to all three groups can be seen in Table 1.

Design and data analysis

For analysis, we treated our study as a 3 (feedback: low, medium, high) by 2 (fixed time vs. fixed trial) by 23 (practice block) design. For each cell and each individual, we calculated mean response time and accuracy. Reward rate was calculated as $\frac{PC}{MRT+ITI+FDT+(1-PC)*ET}$, where MRT and PC refer to the mean correct response time and probability of a correct response, ITI is the inter-trial interval, FDT is the feedback display time, and ET is the error timeout.

We estimated the parameters of a diffusion model (Ratcliff, 1978; Ratcliff & Rouder, 1998), which makes predictions for the reaction time distributions of correct and error responses based on four parameters: a drift rate, which represents the rate of evidence accumulation; a starting point, which represents pre-decision bias; a threshold, which is the amount of evidence needed to trigger a decision; and non-decision time, which is the time required for non-decision processes (see Figure 1). The model also includes between-trial variability parameters, which were not of primary interest here. All parameters were constrained to be equal across blocks, except for the threshold parameter, which was allowed to vary between blocks in order to estimate the changes in the SAT across blocks. In a secondary analysis, we investigated a model variant in which both the threshold and drift rate parameters changed across practice blocks. This model instantiates the psychological assumption that learning on this task changes information processing. Analysis of this model showed no qualitative differences in the results from the main model (see Supplementary Materials).

To estimate the parameters of the diffusion model we used Bayesian hierarchical modeling, where each participant’s individual parameters were constrained to follow group-level truncated normal distributions. A separate hierarchical structure was used for each of the six different groups (feedback x block type), with all group-level parameters allowed to

vary between groups. Our main interest was in the posterior distributions of the threshold parameters for each group, across practice. We also calculated the threshold which optimized reward rate, using the group-level posteriors for the other parameters, and testing all plausible threshold values to see which maximized reward rate.

Lastly, we performed a 3-way Bayesian ANOVA on the differences between observed thresholds and optimal thresholds, which we call “threshold sub-optimality”. This was calculated by subtracting the median of the optimal threshold distribution from the median of the actual threshold distribution for each block for each participant.

The full details for the diffusion modeling, the calculation of the actual threshold distribution, and the calculation of the optimal threshold, can all be found in the Supplementary Materials.

Results

Mean reward rates, reaction time, and accuracy are shown in Figure 2. To investigate the effects of the three factors on reward rate, we calculated a Bayesian ANOVA (Morey, Rouder, & Morey, 2014). The ANOVA model with the greatest marginal likelihood included only the three main effects (block number, block type, and feedback type) and no interactions. The top panel of Figure 2 illustrates the effects. Reward rate increases with practice, for all groups. Reward rate was higher for people in the fixed-trial condition than the fixed-time condition. Participants who received the high level of feedback had higher reward rates than those in the medium level, who in turn had higher levels of reward rate than people who were given low feedback. There was no interaction between feedback type and block number – the no-interaction model above had a much greater marginal likelihood ($BF = 23$). This suggests that all groups increased in reward rate with practice at a similar rate.

The posterior distributions over decision threshold parameters measure the speed-accuracy tradeoff settings adopted by the participants, and so we compared these against posterior predictive distributions for the optimal threshold setting (calculated as described above). These comparisons are shown in Figure 3, with one panel for each group of participants. In each panel, the error bars show the posterior distribution over decision thresholds, across practice blocks. The green bars show the optimal threshold setting – this is constant across practice blocks, because our model fits assumed that all model parameters were fixed cross blocks, except for thresholds. We also analyzed a model that also allowed changes in drift rate with practice; there was no substantive change in the results (see Supplementary Materials).

In the high feedback group, both the fixed time and fixed trials groups showed a clear trend towards the optimal threshold over blocks – participants adopted increasingly close-to-optimal speed-accuracy tradeoff settings with increased practice. In the early blocks, both groups adopted overly-cautious threshold settings (much higher than optimal). In later blocks, the median of the posterior over threshold for both groups was very close

to the optimal threshold, suggesting that most participants had adopted thresholds close to optimal. For the two groups with less specific feedback (“medium” and “low”), there were similar trends but with slightly further-from-optimal threshold settings, and slower movement towards optimality.

An ANOVA on the distance between observed and optimal threshold settings confirmed the trends seen in Figure 3. The model with the highest marginal likelihood included the three main effects (block number, block type, and feedback type), as well as all two way interactions, and the three way interaction. Again, this showed that people move closer to optimal over blocks, and more feedback speeds-up this process. It also showed that consistent with other analyses, but somewhat surprising theoretically, that those in the fixed-trial condition came closer to optimality than those in the fixed-time condition. To test whether the participants had completely achieved optimality, we performed a one-sample Bayesian t-test on the block that appears to be closest to optimality, being block 22 of the high feedback, fixed trial group; however, even this block showed moderate evidence against optimality ($BF = 12.6$), suggesting that although participants were close to optimality, they had not completely achieved it.

It is also important to note that both of these findings based upon the distance between observed and optimal threshold settings, especially the latter finding, should be interpreted with caution, as making such comparisons required using a point estimate of the posteriors, and could only be performed on the individual-level parameters. This ignores both of the key advantages gained through the Bayesian hierarchical modeling methods that we used, being the uncertainty contained in the Bayesian posterior, as well as group-level distribution information.

Discussion

Evaluating human performance against statistically-defined optimality can be an illuminating way to investigate cognition. For simple decision-making, an interesting question is how well people manage to find an optimal balance between the competing demands for caution and urgency. Previous work (Starns & Ratcliff, 2012; Simen et al., 2009; Balci et al., 2011; Starns & Ratcliff, 2010) has provided conflicting evidence about whether people adopt optimal or too-cautious strategies. We found that practice rapidly improved the optimality of the speed-accuracy tradeoff setting. With practice, the group-average decision thresholds became very close to optimal, particularly in groups that received detailed feedback. Increasing feedback detail also led to faster adjustment towards optimality. One particularly interesting finding is that reward rate optimality is apparently attainable, given the right feedback and around one hour’s practice, even for the group average – not just a small subset of participants.

Consistent with Starns and Ratcliff (2012), we found that people who did not receive longer-term feedback (our low feedback group) failed to achieve reward rate optimality. However, all participants, even those without longer-term feedback, still became closer to

optimal with practice. This was consistent with results reported by Simen et al. (2009), who found trends consistent with optimality in participants who had spent at least three sessions previously training on their task.

An important question is why our participants were able to approximate the optimal threshold after such a short amount of practice. Previous research has suggested changes in parameters values over practice (Dutilh, Vandekerckhove, Tuerlinckx, & Wagenmakers, 2009), and movement towards optimality (Balci et al., 2011; Starns & Ratcliff, 2010), but none have reported such major adjustments occurring in such a short time frame. One potential explanation comes from expected utility theory. Shanks, Tunney, and McCarthy (2002) observed that directed feedback, similar to our high feedback condition, can lead to participants achieving optimal performance for expected utility at very fast rate; and much faster than without feedback. Additionally, providing monetary compensation for correct responses leads to a closer, and faster, approximation of optimality than mere correct/error feedback (Shanks et al., 2002; Vulkan, 2000; Siegel & Goldstein, 1959), making the investigation of monetary rewards on reward rate optimality an interesting avenue for future research. However, participants who received directed feedback in the experiment reported by Shanks et al. (2002) still required a great deal of practice to approximate optimality, which casts doubt on this as a complete explanation of our present findings. Another potential explanation is our advanced fitting methods, namely the Bayesian hierarchical modelling, which may have given us a better group-level comparison than previous studies, as well as the added benefit of accounting for uncertainty through the Bayesian posterior. However, both of these potential explanations should be considered speculative, and require further investigation.

Starns and Ratcliff (2012) found that fixed time blocks led to higher reward rates than fixed trial blocks, which was consistent with the idea that experimenter-controlled timing (fixed time blocks) encourage participants to consider reward rates. In our experiment, fixed trial blocks led to a higher reward rate and closer-to-optimal threshold setting than fixed time blocks. The difference between our finding and Starns and Ratcliff's might be explained by different task instructions, which influenced participants' goals. In our study, participants in all groups were given instructions that explicitly encouraged the optimization of reward rate. In contrast, in the fixed trial conditions of Starns and Ratcliff's study, there were no explicit instructions about task goals, or what to optimize (this is consistent with standard practice in decision-making research). It might be, therefore, that the instructions given to participants in studies of optimality have a greater impact than typically appreciated, and those instructions may be important in encouraging participants to adopt optimal speed-accuracy tradeoff settings. An important corollary of this argument is that reward rate optimality may not be a default goal for human participants, which is also what the findings of Hawkins et al. (2012) suggested.

Our study extends on both the studies of Starns and Ratcliff (2012) and Simen et al. (2009) by incorporating feedback on longer-term performance, motivated by the kind of feedback animals might experience in similar studies. Bogacz et al. (2006) discussed the

possibility that participants over-value feedback on individual trials (see also Maddox & Bohil, 1998). If this occurs, then the longer-term feedback could provide an important corrective role. Our experiment showed that providing people with feedback on their general performance led to a threshold that approached reward rate optimality, especially when compared to those who received no general feedback. Such feedback presumably is provided by internal hunger signals in animal experiments on optimality. In addition, adding information about the optimal threshold allowed participants to approach optimality further and faster, showing that human participants are capable of approximating reward rate optimality, at least in ideal conditions. Future research could assess whether such feedback could serve a corrective role in task conditions that have consistently resulted in performance which is extremely far from optimality, such as conditions of high task difficulty.

Although all groups in our experiment moved their speed-accuracy tradeoff settings closer to optimal, no group's threshold estimates completely converged on the optimal value. There are three potential explanations for this. Firstly, the inability to reach the optimal region may have been due to the short length of our experiment; the actual practice time was approximately 24 minutes, which may well be insufficient for full adjustment. Secondly, this could have been due to the nature of reward rate as a function of speed-accuracy tradeoff setting. Decreases in reward rate are not symmetric on either side of the optimal setting, with faster decrease for too-fast thresholds than for too-slow thresholds (Bogacz et al., 2006). If participants were uncertain about the exact value of the optimal threshold setting, the least risky strategy is to set thresholds slower higher than optimal. Thirdly, humans may be simply unable, or unwilling to adopt completely optimal settings in certain conditions, in line with the claims of Starns and Ratcliff (2012) that participants make the wrong threshold adjustment for increasing difficulty. Based upon the previous literature, the ability to achieve, or approximate, reward rate optimality could be potentially related to the cognitive load of the task. This idea is supported by the studies that found older participants to be further from optimality than younger participants (Starns & Ratcliff, 2012, 2010), as well as studies that found participants to be further from optimality with higher task difficulty (Starns & Ratcliff, 2012; Balci et al., 2011) and quicker intervals (Simen et al., 2009). Our task may have had an increased cognitive load due to the relatively quick response-to-stimulus interval (400ms for corrects, 900ms for errors), which could indicate that the feedback structure was able to overcome some of the burden associated with increased load. Future research could more directly explore the impact of cognitive load on achieving optimality.

Conclusions

We investigated the impact of two previously unexplored, yet theoretically important, factors in the study of reward rate optimality, being time spent on the task, and the amount of overall task feedback given to participants. Our study showed that manipulation of both factors led to increased reward rates and closer-to-optimal strategies (Gold &

Shadlen, 2002; Bogacz et al., 2006). These findings suggest that humans are capable of closely approximating reward rate optimality given conditions which mimic the practice and feedback structure that might be seen in naturalistic environments.

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.
- 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.
- Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: linear ballistic accumulation. *Cognitive psychology*, *57*(3), 153–178.
- Ditterich, J. (2006). Evidence for time-variant decision making. *European Journal of Neuroscience*, *24*(12), 3628–3641.
- 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.
- Dutilh, G., Vandekerckhove, J., Tuerlinckx, F., & Wagenmakers, E.-J. (2009). A diffusion model decomposition of the practice effect. *Psychonomic Bulletin & Review*, *16*(6), 1026–1036.
- Gold, J. I., & Shadlen, M. N. (2002). Banburismus and the brain: decoding the relationship between sensory stimuli, decisions, and reward. *Neuron*, *36*(2), 299–308.
- Hawkins, G. E., Brown, S. D., Steyvers, M., & Wagenmakers, E.-J. (2012). An optimal adjustment procedure to minimize experiment time in decisions with multiple alternatives. *Psychonomic bulletin & review*, *19*(2), 339–348.
- Hawkins, G. E., Forstmann, B. U., Wagenmakers, E.-J., Ratcliff, R., & Brown, S. D. (2015). Revisiting the evidence for collapsing boundaries and urgency signals in perceptual decision-making. *The Journal of Neuroscience*, *35*(6), 2476–2484.
- Maddox, W. T., & Bohil, C. J. (1998). Base-rate and payoff effects in multidimensional perceptual categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(6), 1459.
- Morey, R. D., Rouder, J. N., & Morey, M. R. D. (2014). Package ‘bayesfactor’.
- Ratcliff, R. (1978). A theory of memory retrieval. *Psychological review*, *85*(2), 59.
- Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science*, *9*(5), 347–356.
- Roitman, J. D., & Shadlen, M. N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *The Journal of neuroscience*, *22*(21), 9475–9489.
- Shanks, D. R., Tunney, R. J., & McCarthy, J. D. (2002). A re-examination of probability matching and rational choice. *Journal of Behavioral Decision Making*, *15*(3), 233–250.
- Siegel, S., & Goldstein, D. A. (1959). Decision-making behavior in a two-choice uncertain outcome situation. *Journal of Experimental Psychology*, *57*(1), 37.
- Simen, P., Contreras, D., Buck, C., Hu, P., Holmes, P., & Cohen, J. D. (2009). Reward rate optimization in two-alternative decision making: empirical tests of theoretical predictions. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(6), 1865.
- Starns, J. J., & Ratcliff, R. (2010). The effects of aging on the speed–accuracy compromise: Boundary optimality in the diffusion model. *Psychology and aging*, *25*(2), 377.
- Starns, J. J., & Ratcliff, R. (2012). Age-related differences in diffusion model boundary optimality with both trial-limited and time-limited tasks. *Psychonomic bulletin & review*, *19*(1), 139–

- 145.
- Thura, D., Beaugregard-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., Sederberg, P. B., Brown, S. D., & Steyvers, M. (2013). A method for efficiently sampling from distributions with correlated dimensions. *Psychological methods*, *18*(3), 368.
- Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods*, *39*(4), 767–775.
- Vulkan, N. (2000). An economists perspective on probability matching. *Journal of economic surveys*, *14*(1), 101–118.
- Wagenmakers, E.-J., Van Der Maas, H. L., & Grasman, R. P. (2007). An ez-diffusion model for response time and accuracy. *Psychonomic bulletin & review*, *14*(1), 3–22.
- Wald, A., & Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *The Annals of Mathematical Statistics*, 326–339.

Supplementary Material

Calculating the optimal threshold during the decision-making task for the high feedback group.

The optimal reward rate (for the feedback of the high feedback group) was calculated by firstly estimating the parameters of a simple evidence accumulation theory, the E-Z diffusion model (Wagenmakers et al., 2007)¹. This model has three parameters: a drift rate, which measures the ability of the decision-maker at this task; a decision threshold, which measures the speed-accuracy tradeoff; and a non-decision time which measures how long is spent on things other than evidence accumulation, such as executing the motor response. These parameters were estimated using Wagenmakers et al.’s method calculated on the previous 200 trials (or all trials, if fewer than 200 had been completed). Conditional on the drift rate and non-decision time parameters, we calculated the expected mean response time and accuracy rate corresponding to each potential threshold setting over a wide range: from 0.001 to 1, in steps of 0.001. We used the expected mean response time and accuracy for each potential threshold setting to calculate the corresponding expected reward rate, and we then identified the threshold setting which corresponded to the maximum (optimal) reward rate. This was the threshold setting used to provide feedback of the sort illustrated in Table 1.

Diffusion modelling details, and the calculation of the actual and optimal thresholds for data analysis.

The scaling parameter of the diffusion process, which sets the size of the random moment-to-moment variations in evidence, was fixed at 1. The model also assumes trial-to-trial variability in the drift rate, starting point, and non-decision time parameters, leading to three additional variability parameters: the standard deviation of the Gaussian drift

¹Note that obtaining the optimal threshold could have also been achieved by using the analytic solutions provided in (Bogacz et al., 2006).

rate distribution, the width of the uniform distribution of starting points, and the width of the uniform distribution of non-decision times.

We used code extracted from the “fast-dm” toolbox (Voss & Voss, 2007) to calculate the density function of the diffusion model, and we used Markov-chain Monte Carlo with differential evolution proposals (Turner, Sederberg, Brown, & Steyvers, 2013) to sample from the posterior distribution over parameters. For the MCMC, we used 87 chains and drew 3,000 samples from each. We discarded the first 1,500. The hierarchical structure used was as follows:

Data level :

$$(RT_i, resp_i) \sim Diffusion(v_i, z_i, ter_i, a_{2:24,i}, s_{v,i}, s_{z,i}, s_{ter,i})$$

Group level :

$$\begin{aligned} v_i &\sim N(\mu_v, \sigma_v) \\ \frac{z_i}{a_{2:24,i}} &\sim TN(\mu_z, \sigma_z, 0, 1) \\ ter_i &\sim TN(\mu_{ter}, \sigma_{ter}, 0, Inf) \\ a_{2:24,i} &\sim TN(\mu_{a_{2:24}}, \sigma_{a_{2:24}}, 0, Inf) \\ s_{v,i} &\sim TN(\mu_{s_v}, \sigma_{s_v}, 0, Inf) \\ s_{z,i} &\sim TN(\mu_{s_z}, \sigma_{s_z}, 0, Inf) \\ s_{ter,i} &\sim TN(\mu_{s_{ter}}, \sigma_{s_{ter}}, 0, Inf) \end{aligned}$$

Prior distributions :

$$\begin{aligned} \mu_v &\sim N(3, 3) \\ \mu_z &\sim TN(.5, .5, 0, 1) \\ \mu_{ter} &\sim TN(.3, 1, 0, Inf) \\ \mu_{a_{2:24}} &\sim TN(2, 2, 0, Inf) \\ \mu_{s_v}, \mu_{s_z}, \mu_{s_{ter}} &\sim TN(1, 1, 0, Inf) \\ \sigma_v, \sigma_{a_{2:24}} &\sim \Gamma(1, 1) \\ \sigma_z, \sigma_{ter}, \sigma_{s_v}, \sigma_{s_z}, \sigma_{s_{ter}} &\sim \Gamma(.5, .5) \end{aligned}$$

where i indexes participants, and $TN(\mu, \sigma, lower, upper)$ is a truncated Gaussian distribution with mean μ and standard deviation σ , truncated at $lower$ and $upper$.

To calculate the actual threshold, we took 100 samples from the group-level posterior distributions of the μ and σ parameters for the estimated thresholds of each block. From

this, 100 samples were then randomly drawn from a truncated normal distribution for each pair of μ and σ parameters, with mean μ and standard deviation σ . This sampling procedure captures both the uncertainty contained in the posteriors of the parameters, as well as the uncertainty of the group estimate contained in the σ parameter.

To calculate the optimal threshold, 100 samples were taken from the posterior distributions of all parameters other than threshold. Next, all possible threshold between 0.01 and 3 in increments of 0.01 were investigated, and 10,000 posterior predictive data were simulated from the diffusion model using these parameters. From this, the mean reaction time and percentage correct were calculated for the simulated data, and whichever possible value of the threshold created the highest reward rate was declared the optimal threshold. This formed a distribution of optimal thresholds, via the distribution of parameters from the posterior.

For the threshold sub-optimality calculation, the optimal and actual threshold values were calculated using the same method as above, except with the individual-level parameters rather than group-level parameters, in order to get estimates for each individual subject. Also, in order to make the method computationally tractable, only 50 samples were taken from the distributions, and all possible threshold were tested between 0.05 and 2 in increments of 0.05.

Supplementary analysis 1: All subjects.

In order to assess whether our exclusion criteria for subject's performance may have influenced our results, we replicated our key analysis - the estimation of the actual and optimal thresholds via the diffusion model - with all participants included. The model specification was identical to that of the analysis in the main section of the paper, with the only difference being the number of subjects analyzed.

As can be seen in Figure 4, the results with all subjects included - both in terms of the qualitative patterns and the exact quantitative trends - appear to be identical to that of our analysis that excluded subjects whose accuracy was under 70%. The participants in the high feedback group still show a clear trend towards the optimal threshold over blocks, and finished extremely close to the optimal threshold. Those in the medium and low feedback groups again showed similar trends, though to a lesser extent, finishing further from the optimal threshold.

Supplementary analysis 2: Model that included a variation in drift rate over blocks.

On many tasks, participants improve their ability to process the decision-relevant information with practice. Therefore, one may consider the model used in the main analysis, which only allows threshold to differ with practice, to be too restrictive. Although we excluded the initial block in an attempt to allow for transient improvements, this may have been inadequate to cover the majority of task learning, which may have continued throughout the task.

To check whether making such an assumption influenced our results, we replicated our key analysis using a model that contained a variation in both the threshold parameter and the drift rate parameter over blocks. The model specification was identical to that of the analysis in the main section of the paper, except this model contained a separate parameter for the drift rate of each block for each subject, and separate group-level parameters for the drift rate of each block. The priors for these drift rate group-level parameters were all identical, and were the same as that of the drift rate priors in our analysis in the main section of the paper.

As can be seen in Figure 5, the results with a variation included maintain the same qualitative pattern as the results of our analysis in the main section of the paper. The participants in the high feedback group still show a clear trend towards the optimal threshold over blocks, and finished close to the optimal threshold. However, the proximity of actual threshold to the optimal threshold appears to be slightly further than the analysis in the main section of the paper, showing a slight difference in the exact quantitative trend. Those in the medium and low feedback groups again showed similar trends, though to a lesser extent, finishing further from the optimal threshold.

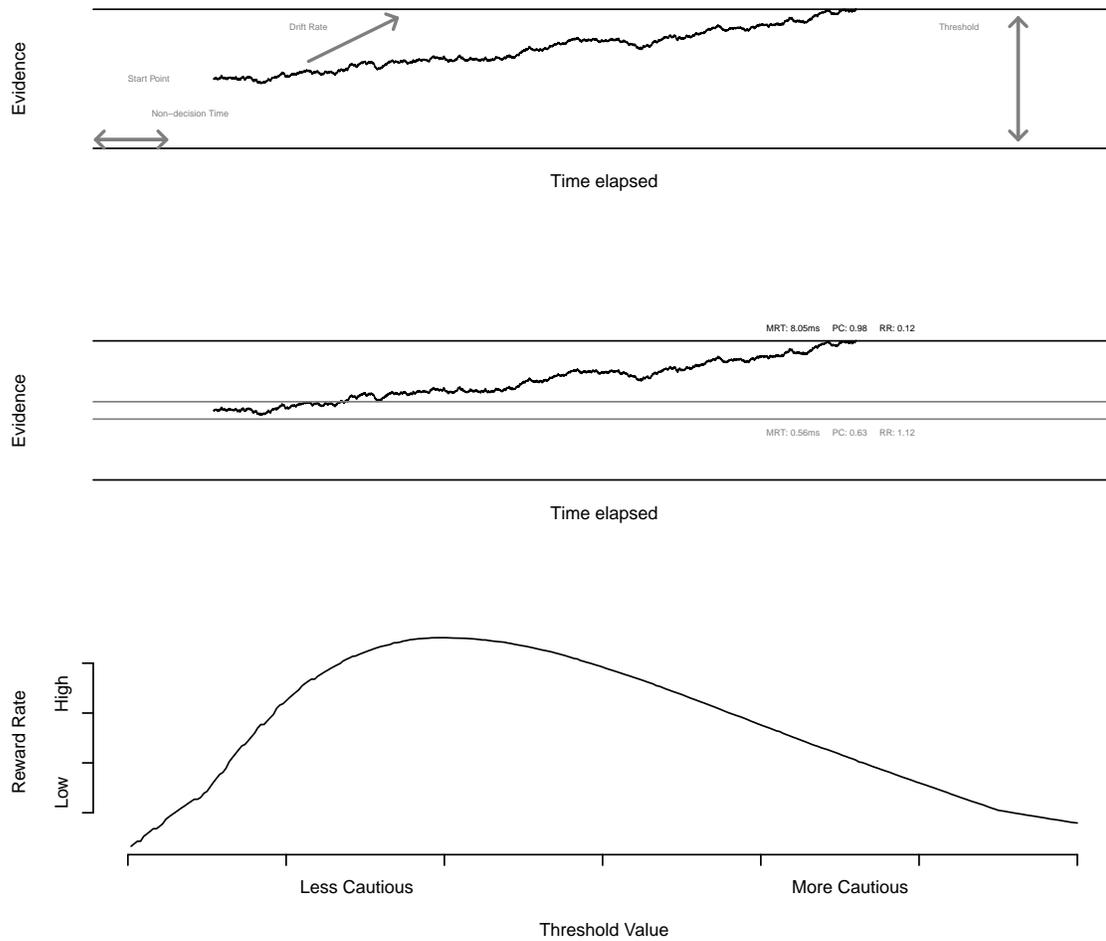


Figure 1. Top panel illustrates the basic components of the diffusion model. The black line displays the path of evidence accumulation over a decision, which is governed by the drift rate and noise. This path begins at the start point, which represents pre-decision bias, after some non-decision time has elapsed, which represents the time taken by non-decision components such as perception and response execution. The process terminates when the evidence path hits either the top or bottom threshold, with each threshold representing a separate response option. The middle panel compares a more and a less cautious threshold setting, with black and grey lines respectively. Each threshold leads to a different mean reaction time, and a different percentage correct, and so a different reward rate. The less cautious, grey threshold results in a much greater reward rate than the more cautious, black threshold, as the amount of time saved for this drop in caution justifies the accuracy lost. The bottom panel shows how reward rate changes with decision threshold. This function is concave downwards, and reward rate drops more quickly from its maximum for too-urgent than too-cautious thresholds.

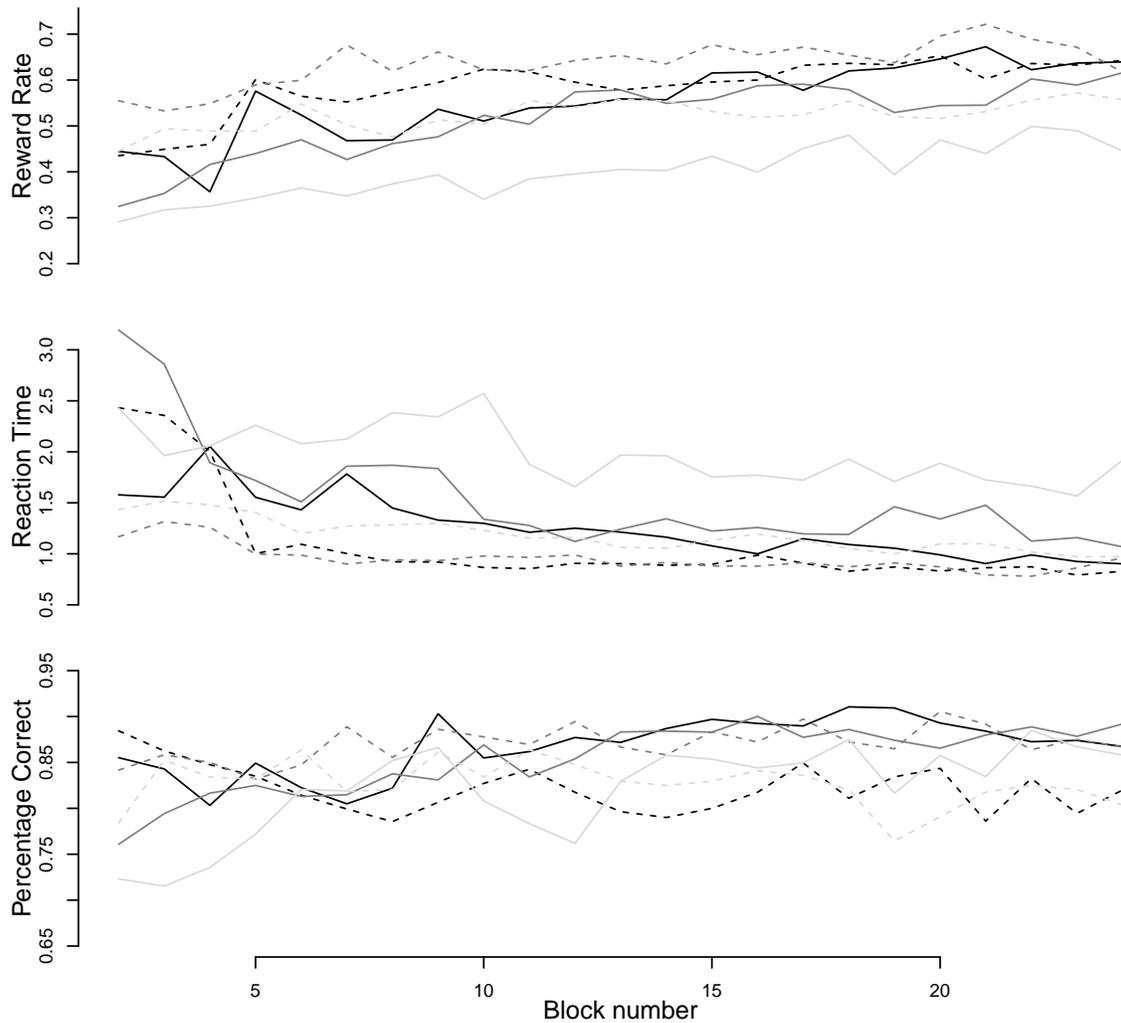


Figure 2. Mean reward rate (top panel), mean reaction time (middle panel), and percentage correct (bottom panel) as functions of practice block, for participants in each of the six groups. The solid lines correspond to the fixed-time groups, and the dashed lines to the fixed-trial groups. The light gray lines show the low feedback group, the dark gray lines the medium feedback group, and the black lines the high feedback group. All groups show an increase in reward rate over blocks, with the fixed trials groups having a generally greater reward rate than the fixed time groups, and the high feedback groups having a greater reward rate than the medium feedback groups, who in turn had a greater reward rate than the low feedback groups. The mean reaction time and percentage correct graphs both reflect this finding, with general increases in speed and accuracy with time.

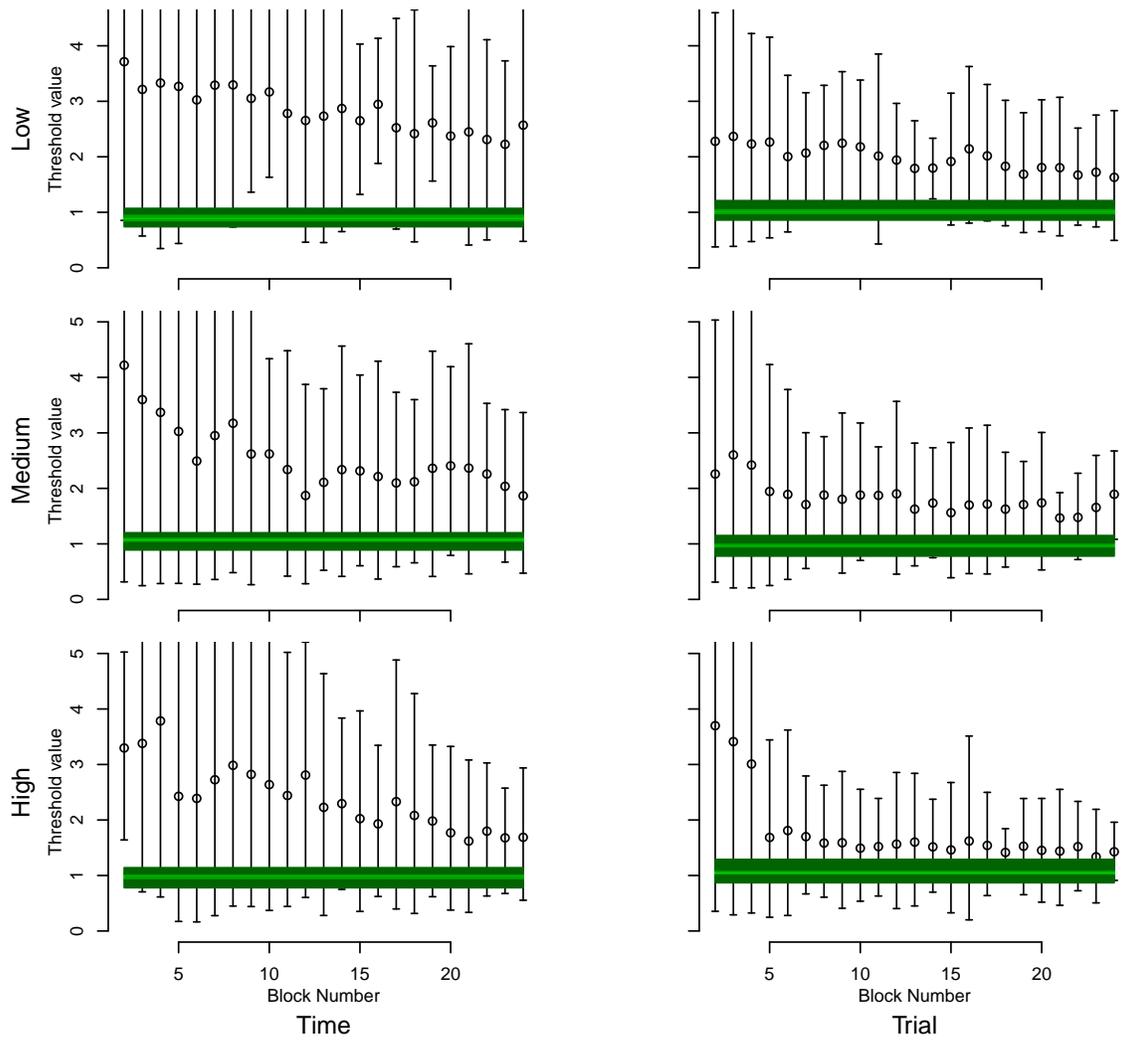


Figure 3. The median of the group-level threshold posterior (circles), the 95% quantiles (error bars), and the optimal threshold distribution (green area), with the optimal distribution being broken into different shades of green for the 40%-60% quantiles (lightest shade), the 20%-40% and 60%-80% quantiles (middle shade), and the 0%-20% and 80%-100% quantiles (darkest shade). The y-axis displays the numerical threshold value, which is around 1 for all groups, and the x-axis displays the block number, starting at block 2 and ending at block 24. The different rows show, from top to bottom, the low, medium, and high feedback groups, and the columns, from left to right, show the fixed-time and fixed-trial groups. Thresholds in all groups move closer to the optimal value over blocks. Groups with more feedback have greater movement in threshold, and end closer to optimal. Fixed trial groups also become closer to optimal than fixed time groups.

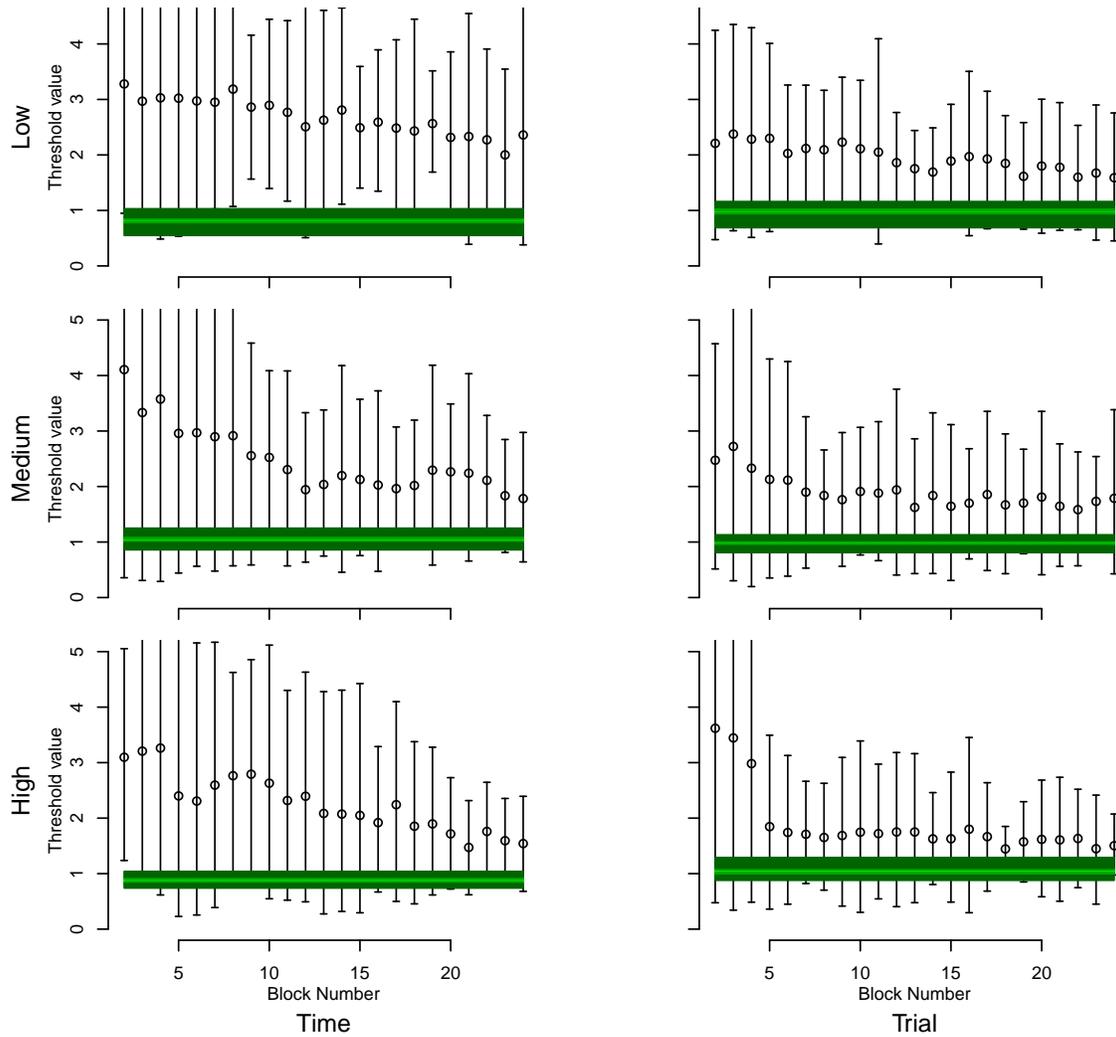


Figure 4. Displays the replication of our main analysis with all subjects included. Median of the group-level threshold posterior (points), the 95% quantiles (error bars), and the optimal threshold distribution (green area), with the optimal distribution being broken into different shades of green for the 40%-60% quantiles (lightest shade), the 20%-40% and 60%-80% quantiles (middle shade), and the 0%-20% and 80%-100% quantiles (darkest shade). The y-axis displays the numerical threshold value, which is around 1 for all groups, and the x-axis displays the block number, starting at block 2 and ending at block 24. The different rows show, from top to bottom, the low, medium, and high feedback groups, and the columns, from left to right, show the fixed-time and fixed-trial groups. Thresholds in all groups move closer to the optimal value over blocks. Groups with more feedback have greater movement in threshold, and end closer to optimal. Fixed trial groups also become closer to optimal than fixed time groups.

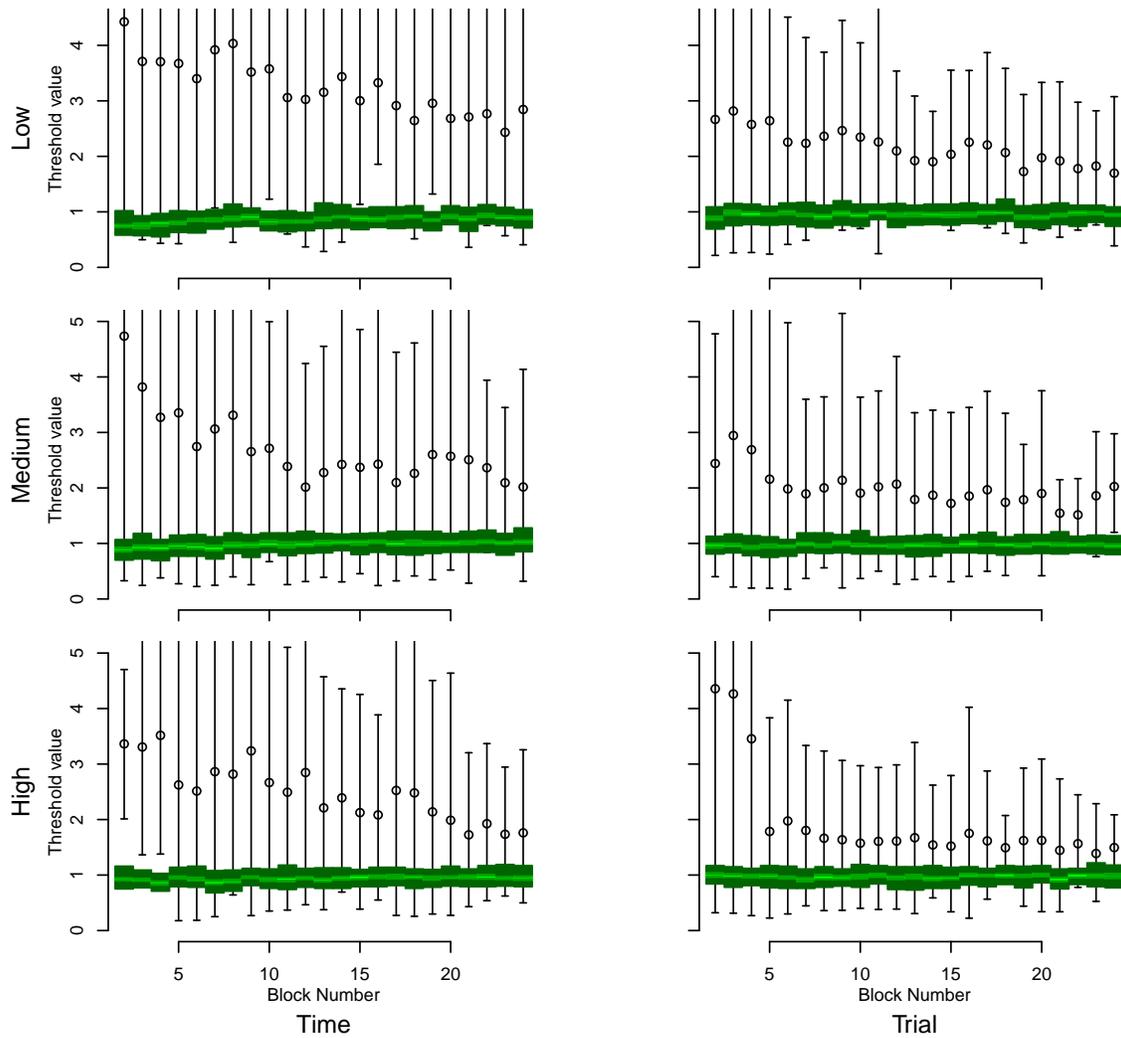


Figure 5. Displays the replication of our main analysis, using a diffusion model that allowed a variation in drift rate over block, as well as the standard variation in threshold over block. Median of the group-level threshold posterior (points), the 95% quantiles (error bars), and the optimal threshold distribution (green area), with the optimal distribution being broken into different shades of green for the 40%-60% quantiles (lightest shade), the 20%-40% and 60%-80% quantiles (middle shade), and the 0%-20% and 80%-100% quantiles (darkest shade). The y-axis displays the numerical threshold value, which is around 1 for all groups, and the x-axis displays the block number, starting at block 2 and ending at block 24. The different rows show, from top to bottom, the low, medium, and high feedback groups, and the columns, from left to right, show the fixed-time and fixed-trial groups. Thresholds in all groups move closer to the optimal value over blocks. Groups with more feedback have greater movement in threshold, and end closer to optimal. Fixed trial groups also become closer to optimal than fixed time groups.