Are perceptuo-motor decisions really more optimal than cognitive decisions? ☆

Andreas Jarvstad [a,c,*], Ulrike Hahn [a,d], Paul A. Warren [b], Simon K. Rushton [a]

[a] School of Psychology, Cardiff University, United Kingdom
[b] School of Psychological Sciences, The University of Manchester, United Kingdom
[c] School of Experimental Psychology, University of Bristol, United Kingdom
[d] Department of Psychological Sciences, Birkbeck, University of London, United Kingdom

ARTICLE INFO

ABSTRACT

Human high-level cognitive decisions appear sub-optimal (Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1979). Paradoxically, perceptuo-motor decisions appear optimal, or nearly optimal (Trommershäuser, Maloney, & Landy, 2008). Here, we highlight limitations to the comparison of performance between and within domains. These limitations are illustrated by means of two perceptuo-motor decision-making experiments. The results indicate that participants did not optimize fundamental performance-related factors (precision and time usage), even though standard analyses may have classed participants as 'optimal'. Moreover, simulations and comparisons across our studies demonstrate that optimality depends on task difficulty. Thus, it seems that a standard model of perceptuo-motor decision-making fails to provide an absolute standard of performance. Importantly, this appears to be a limitation of optimal models of human behaviour in general. This, in conjunction with non-trivial evaluative- and methodological differences, suggests that verdicts favouring perceptuo-motor, or perceptual, systems over higher-level cognitive systems in terms of level of performance are premature.

© 2013 The Authors. Published by Elsevier B.V. All rights reserved.

1. Introduction

There appears to be a striking dissociation between human perceptuo-motor- and cognitive decision-making performance. Cognitive decision-making ability is widely viewed as distinctly less than optimal, because it conflicts with the normative prescriptions of decision theory that set out how 'rational' decision makers should behave (Birnbaum, 2008; Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1979). Perceptuo-motor decision-making, on the other hand, appears well described by the same theory (for a review see Trommershäuser et al., 2008; see Whiteley & Sahani, 2008 for a similar conclusion in a perceptual domain). This apparent dissociation has been highlighted repeatedly. Trommershäuser, Landy and Maloney, for example, note that "...in marked contrast to the grossly sub-optimal performance of human subjects in traditional economic decision-making experiments, our subjects' performance was often indistinguishable from optimal." (2006, p. 987; see also e.g., Maloney, Trommershäuser, & Landy, 2007; Trommershäuser et al., 2008).

This performance dissociation is puzzling. Few reasons are evident for why perceptuo-motor decision-making should be optimal, while cognitive decision-making is sub-optimal (but see e.g., Chater & Oaksford, 2008; Evans & Over, 1996). Furthermore, little progress appears to have been made in explaining the difference.

There are at least three possible sources for the apparent dissociation: (1) competence may be modality dependent (2) performance may be task dependent and (3)

differences may result from the way performance is evaluated. If competence were indeed modality dependent this would be a striking finding. However, as pointed out by Trommershäuser and colleagues (e.g., Maloney et al., 2007), experimental paradigms across the two fields differ along a number of methodological dimensions. Perceptuo-motor studies generally involve repeated decisions with outcome feedback and internalized probabilities. Cognitive decision tasks, on the other hand, generally involve one-shot decisions without feedback and exact probabilities stated on paper (see e.g., Birnbaum, 2008; Kahneman & Tversky, 1979, but see e.g., Hertwig, Barron, Weber, & Erev, 2004; Thaler, Tversky, Kahneman, & Schwartz, 1997). Thus, a less interesting explanation is that one, or more, of these methodological differences give rise to the apparent dissociation.

Not only are there methodological differences in tasks, performance is also evaluated differently in the two fields. Although both perceptuo-motor- and cognitive studies draw on normative theories to provide performance standards, adherence to these norms is assessed in different ways. Generally, the perceptual and perceptuo-motor literature asks how closely human performance matches that of an ideal agent (see e.g., Barlow, 1962; Geisler, 2003; Trommershäuser, Maloney, & Landy, 2003a, 2003b). Broadly, an ideal agent is a model that performs a given task maximally well. Constraints under which the system is assumed to operate are typically built into the model. The cognitive literature, on the other hand, typically asks if a system violates one or more of the axioms of decision theory (assumed fundamental principles of rational choice such as the transitivity of preferences, or independence of irrelevant alternatives, see e.g., Birnbaum, 2008; Hertwig et al., 2004; Kahneman & Tversky, 1979). Experiments are designed so that certain response patterns would constitute violations of these axioms, thereby indicating irrationality (i.e., sub-optimality). Thus, assessment of performance typically differs in two ways across cognitive and perceptual/perceptuo-motor studies: quantitative versus qualitative violations of normative theories and presence or absence[1] of system constraints.

Given these non-trivial differences between cognitive- and perceptuo-motor studies, comparisons of human performance across the two domains need to be made with care. In this paper we highlight difficulties associated with such comparisons using two perceptuo-motor decision-making experiments. The experiments demonstrate that minor changes in task parameters, specifically changes that do not affect an optimal agent's performance, influence whether participants are actually viewed as optimal or sub-optimal. We follow up these empirical results by illus-

trating through simulations how specific changes in task parameters can cause participants hitherto classified as optimal to be classed as sub-optimal. Our experiments also suggest that people's perceptuo-motor decisions are sub-optimal in ways not captured by Trommershäuser et al.'s (2003a, 2003b) model. Together these results, we think, suggest that claims of greater optimality for perceptual systems over higher-level cognitive systems may be premature.

2. Perceptuo-motor decisions & decision performance assessment

The recent interest in comparing the relative optimality of cognitive and perceptuo-motor decisions stems from Trommershäuser et al.'s (2003a, 2003b) elegant perceptuo-motor decision paradigm. Their paradigm has made it possible to translate into perceptuo-motor tasks the kinds of decision problems given to participants in cognitive psychological studies. We begin with a brief introduction to Trommershäuser et al.'s paradigm.

Because the perceptuo-motor system is noisy, speeded pointing towards a target will result in a response distribution dispersed around a chosen aim point (cross, Panel A, Fig. 1). Trommershäuser et al. exploit this noisiness to create perceptuo-motor decisions that are mathematically equivalent to standard cognitive decisions (such as those of e.g., Kahneman & Tversky, 1979).

In Trommershäuser et al.'s (2003a, 2003b) paradigm, participants point towards stimulus configurations (Panel B, Fig. 1) under time pressure, with the goal of earning as many points as possible. Participants accrue points if they hit a reward region (full line, Panel B), lose points if they hit a penalty region (dashed line, Panel B), and incur both penalty and reward if they hit the intersection of both regions.

Different aim points (different symbols, Panel B) will result in different probabilities of hitting each region (*hit probabilities,* Panel C). Different hit probabilities, in turn, will result in different numbers of points earned. Given that there are many aim points, participants are in effect choosing between many different options of the form: reward with probability $p = X$, penalty with $p = Y$, both reward and penalty with $p = Z$. This is easily recognized as a traditional decision-making problem (see e.g., Kahneman & Tversky, 1979).

If participants are to maximize the number of points they earn, they have to find the aim point that will allow them to do so. Trommershäuser et al. (2003a) propose that people's behaviour in these tasks can be explained by a process model that assumes that people solve this optimization problem and make optimal decisions.

As noted in the introduction, the standard way of assessing performance in paradigms such as Trommershäuser et al.'s (2003a, 2003b) is to compare participants' performance to that of an ideal agent. An ideal agent is an agent that performs the task maximally well. Of course, we should not expect participants (even if optimal) to *precisely* match optimal performance (e.g., because our estimates of their behaviour are noisy). Instead, the typical question is whether people are statistically distinguishable from optimal. Next, we describe how this is determined.

---

[1] Studies of higher-level decision making and judgment typically are not concerned with constraints when evaluating participant performance. Instead it is assumed that the experimental task is sufficiently easy that any system that adhered to the studied axioms would in principle be able to perform the necessary computations (Evans, 1993). This is not to say that constraints have gone unstudied. Kahneman and Tversky (1996), for example, have argued that when extensional cues are given to participants, performance improves. This effect is presumed to be due to extensional cues triggering a slow and effortful processing system that would otherwise not have been used (Kahneman & Frederick, 2002).
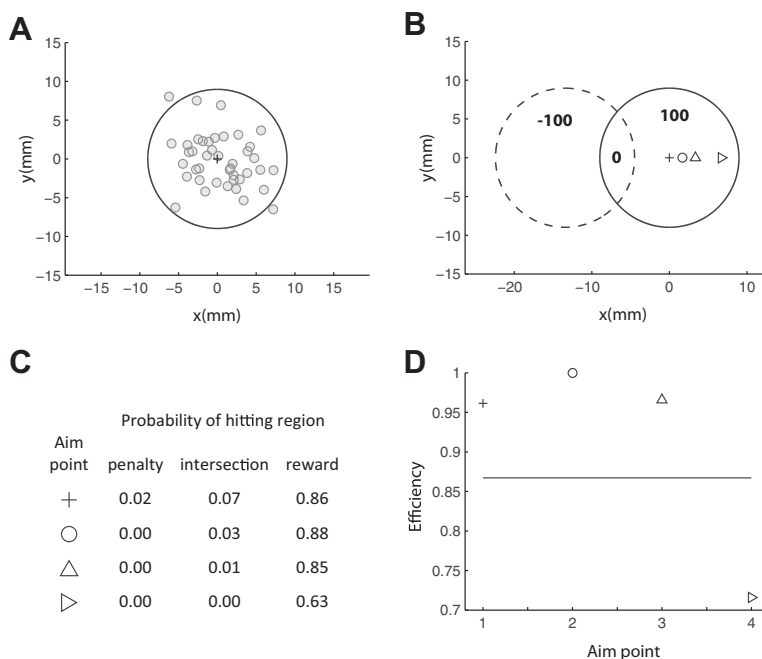
Fig. 1. Perceptuo-motor gambles and performance assessment. Panel A: A simulated response distribution (grey discs) from one participant (σ^2 = 14.78, Participant 2, Experiment 2, in Trommershäuser et al., 2003a) aiming at the centre of a target (cross, Panel A). Panel B: Example of one stimulus configuration and reward structure employed by Trommershäuser et al., with example aim points (symbols) and region-specific rewards and penalties (numbers). Panel C: Hit probabilities for the aim points in panel B. Panel D: Efficiencies (expected gain normalized by optimal expected gain) for the aim points in panel B. The optimal aim point (circle) has an efficiency of 1. The horizontal line represents the lower 95 percentile of optimal performance. Efficiencies below this line are lower than expected by chance and hence sub-optimal.

First we need to determine the optimal choice for a given stimulus configuration (e.g., Panel B, Fig. 1). Given the response distribution of a specific participant (e.g., Panel A, Fig. 1), it is relatively straightforward to find what would be the optimal aim point. One simple method is to systematically move the response distribution around and evaluate the expected gain at each point, until a maximum is found. This simple method, and other more complex methods based on maximum likelihood estimation and numerical integration (Trommershäuser et al., 2003a), work because participants are assumed to control only the position of their response distribution – not its shape.[2]

Once the optimal aim point is found, the hypothetical optimal agent, having inherited the participant's perceptuo-motor variability, "performs" the experiment many times (e.g., 100,000 times). The agent always chooses the optimal aim point. Effectively, one asks: "if this participant performed the same experiment again – but this time always chose the best possible aim point – how well would they do?"

For each simulated experiment, the average earnings of the optimal agent are computed. This procedure results in a distribution of average (expected) optimal earnings (the agent's earnings will vary due to the inherited perceptuo-motor variability), describing how many points an optimal agent is expected to earn in this particular experiment. As described below, this distribution can be used to infer whether participants are performing sub-optimally.

To facilitate comparisons across experiments and participants, the earnings distribution of the optimal agent is divided by its mean. This generates an efficiency distribution, where an efficiency of 1 corresponds to the expected earnings of the optimal agent. Because of the modelled sensori-motor noise, even an optimal agent is likely to deviate from an efficiency of 1 in a particular experiment. Efficiencies below 1 mean that the agent performed worse than expected and efficiencies above 1 mean that the agent performed better than expected. By the same token, if *participants* were in fact optimal, their efficiency scores would likewise be distributed around an efficiency of 1.

The above procedure estimates the distribution of earnings for the optimal agent. To infer whether a specific

---

[2] Harris and Wolpert (1998) showed that Fitts law (Fitts, 1954) is well accounted for by a model that assumes that movements are executed so that, given a specific movement duration, precision is maximized (or equivalently: movement duration is minimized given a precision constraint). Thus, to the extent that people maximize time use in perceptuo-motor decision-making tasks (as e.g., Trommershäuser et al., 2003a found, but see Gepshtein, Seydell, & Trommershäuser, 2007) it can be assumed that participants also maximize precision and therefore do not actively control the shape of their response distribution. If this assumption holds, precision (and equivalently - movement time) should be unaffected by, for example, the size of targets or verbal instruction in paradigms such as Trommershäuser et al.'s. This prediction applies because movement time is restricted, and using nearly all the available time will maximize precision, which in turn will allow targets to be hit more often, which in turn results in more money earned. However, for every-day reaching, in which the goal is to pick up an object, without a strong constraint on movement time, people are likely to reach to objects such that they can interact with them successfully (e.g., use more time, and so be more precise, when reaching towards a key than a pillow, see also e.g., Fitts, 1954).

participant behaved sub-optimally, the efficiency of that participant is computed by dividing their earnings with the expected earnings of the optimal agent. The participant's efficiency can then be compared to the 95% confidence interval on the optimal agent's efficiency distribution. If the participants' earnings lie below the lower bound of the confidence interval the participant is classed as sub-optimal (i.e., statistically distinguishable from optimal).[3] In other words, when participants earnings are sufficiently unlikely to have been generated by an optimal agent they are classed as sub-optimal.

Again, the example in Fig. 1 demonstrates this. Panel B shows a number of potential aim points: because of the cost incurred by hitting the penalty region, the optimal aim point is not the centre of the reward region (here represented by a cross), but rather a point shifted slightly away from the penalty region, specifically the point marked by the circle. The two aim points marked by triangles are shifted further from the penalty region than is optimal; aiming here further reduces the chances of hitting the penalty region, but makes it more likely that the reward region is missed altogether. However, only one of these – the rightward facing triangle – would result in the participant being classed as sub-optimal (see Panel D), because only it lies outside the lower confidence bound [straight line, Panel D]). Thus, due to the limited precision with which we estimate optimal performance, and because deviations from optimality might not be particularly costly, participants can deviate from the optimal strategy and still be classed as optimal.

3. Experimental investigation

Using the paradigm just outlined, or variants thereof, Trommershäuser, Maloney and Landy have explored perceptuo-motor decision-making extensively (see Trommershäuser et al., 2008). We were interested in one of the distinctions they make: that of implicit and explicit decisions. Seydell, McCann, Trommershäuser, and Knill (2008) note that cognitive paradigms generally involve explicit choices (introspectively one is aware of choosing), whilst perceptuo-motor paradigms, such as the one outlined above, generally involve implicit decisions (introspectively one is unaware of choosing). Trommershäuser et al. have previously explored the explicit/implicit choice dimension in two studies (Seydell et al., 2008; Trommershäuser, Landy, & Maloney, 2006) – and concluded that both explicit and implicit motor choice are optimal, or near-optimal.

In our two experiments, which were designed to explore this distinction further, participants made two choices per trial: an aim point choice (the "implicit" choice) and a target choice (the "explicit" choice). The paradigm is illustrated in Fig. 2 (see Methods for details). All pointing movements originated from a dock (the white disc in Fig. 2) and targets were displayed at different distances (crosses illustrate potential target locations in Fig. 2). On each trial, participants had to choose whether to attempt to hit a small or a large target. Participants' response time was limited, which meant that small targets were more difficult to hit than large targets, and meant that far targets were more difficult to hit than nearer targets (Fitts, 1954; Schmidt, Zelaznik, Hawkins, Frank, & Quinn, 1979).

Hitting a target incurred a reward and missing a target resulted in a penalty. The task goal was to earn as many points as possible. To earn as many points as possible, participants had to trade off the probability of hitting each target with the target's value. The small target was always worth more than the large target. Target hit probabilities depended on participants' aim point choices, their motor variability, the size of the target, and the distance to the target.

A novel aspect of our study was that the expected gain (i.e., the average number of points received) for each target depended on the size of the target as well as its distance to the dock. Thus, a basic question was whether humans are able to trade off target size and target distance in an optimal manner when making perceptuo-motor choices.

The use of two target sizes also enabled an indirect assessment of one of the assumptions built into Trommershäuser et al.'s model (2003a, 2003b): that the perceptuo-motor system minimizes motor variability. This assumption is critical because motor variability is fixed when deriving optimal predictions (as noted above in 'Perceptuo-motor decisions & decision performance assessment').

Previous studies have also probed the question of human time allocation in perceptuo-motor tasks. The general conclusion has been that time allocation is optimal or near-optimal (e.g., Battaglia & Schrater, 2007; Dean, Wu, & Maloney, 2007; Hudson, Maloney, & Landy, 2008, and see Jarvstad, Rushton, Warren, & Hahn, 2012 for a comparison of time allocation across perceptual and cognitive domains). However, participants in those studies were explicitly instructed to optimize time usage. Consequently, this does not answer the question of whether the perceptuo-motor system optimizes time in general. First evidence that it may not can be found in the study of Gepshtein et al. (2007).

Gepshtein et al. (2007) employed near and far targets and a response time criterion that was identical for near and far targets. In other words, participants could potentially spend just as much time reaching for near as for far targets. However, participants used less time when they reached for near targets compared to when they reached for far targets. Given the speed-accuracy trade-off (Fitts, 1954; Schmidt et al., 1979), it would appear that precision was sacrificed as reaches to near targets were faster than necessary. Since Gepshtein et al.'s results suggest that time allocation in motor responding may not be optimal without specific, explicit, instruction, further examination seems important.

We conducted two experiments with the task just outlined. The task parameters differed across Experiments 1 and 2. Specifically, target size, target distance, number of possible target locations and the reward for the large target differed across the experiments. To state that the perceptuo-motor system is optimal (or nearly

---

[3] Of course, the logic of null-hypothesis testing implies that one cannot conclude optimality by this method. Strictly speaking, all one can do is to reject the null hypothesis of optimal performance, i.e., reject as optimal anyone who lies outside this interval.
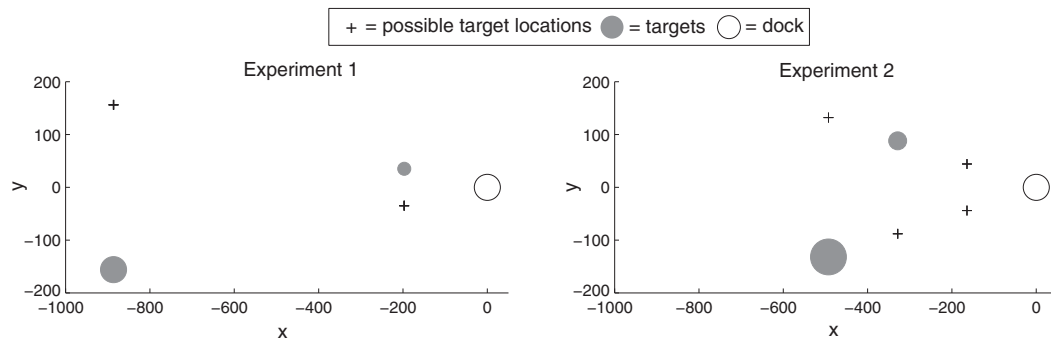
Fig. 2. Design of Experiments 1 and 2. Grey discs and crosses indicate potential target locations. Grey discs represent one possible target configuration. The white disc represents the dock from which all movements originated. The reward for hitting the large target was 50 in Experiments 1 and 75 in Experiment 2. In both experiments, the reward for the small target was 100 points, and the penalty for missing either target was -25. *Note:* targets are not drawn to scale.

so), presumably implies that it can deal with a variety of situations that might occur – not that it is optimal for one particular target size or one specific set of rewards only. That is, if the perceptuo-motor system is optimal, one would expect it to be able to cope with the changing conditions across Experiments 1 and 2.

In the following, we report on both Experiment 1 and Experiment 2 simultaneously. This facilitates comparisons between the two experiments which should produce very similar results if behaviour is indeed optimal. As it turns out, however, seemingly innocuous changes in task parameters can have dramatic effects on whether participants are classed as optimal or sub-optimal.

4. Methods

4.1. Participants and instructions

Sixteen (8 in each experiment) members of the Cardiff University Psychology participant panel were paid an hourly rate of £6 to participate, plus a performance related bonus based on their efficiency (efficiency ∗ £6). The study had approval of the local ethics committee.

Participants were informed of the reward structure in each experiment (that is, how many points could be earned by hitting each target, and how many were deducted for a 'miss'). Participants were told to maximize their total score ("earn as many points as possible"). They were also told that they could receive an additional bonus (ranging between £0 and £6) depending on their performance ("the better you do the more money you will receive"). All participants were naive as to the purpose of the study. All had normal, or corrected to normal, vision and were fully mobile. Participants were fully informed about the experimental protocol.

4.2. Apparatus

The experiments were written in Matlab (Mathworks, Inc.) and run with the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) on a Mac Mini (Apple, Inc.). Participants were seated in front of a pen display (Wacom DTZ-2100, Wacom Co. Ltd.) slanted at 65°. The pen display was used

to display stimuli and record responses. Responses were made with the spring loaded eraser end of a standard Wacom stylus pen. Participants chose their distance and height relative to the display so as to enable natural pointing movements.

4.3. Stimuli, experimental design and procedure

Fig. 3 (Panel A, see also Fig 2) illustrates a sample stimulus configuration. In both experiments, each stimulus configuration contained a dock (radius 16 pixels/ ∼4.3 mm); this dock was the starting position for each trial and was displayed throughout the session. Two discs (potential targets), one large (Experiment1: radius 16 pixels/ ∼4.3 mm; Experiment 2: radius 22 pixels/∼5.9 mm) and one small (Experiment 1: radius 8 pixels/∼2.16 mm; Experiment 2: radius 11 pixels/∼2.9 mm), were displayed to the left of the dock (except for one left handed participant, for whom dock/targets were mirrored).

On each trial, one disc was displayed along the upward diagonal (dotted line Fig. 3 Panel A) and one was displayed along the downward diagonal. In Experiment 1, discs were displayed at one of two distances relative to the dock: near (200 pixels/∼5.4 cm) and far (900 pixels/∼24.3 cm). In Experiment 2, discs were displayed at one of three distances: near (170 pixels/∼4.6 cm), medium (340 pixels/ ∼9.2 cm), or far (510 pixels/∼13.8 cm). A full factorial combination of elevation (up/down), target location and non-target location resulted in 8 unique perceptuo-motor lotteries (i.e., stimulus configurations) for Experiment 1 and 18 unique lotteries for Experiment 2.

Each experiment consisted of one practice session and one experimental session. Each session contained 44 trials per unique stimulus configuration. In the practice session no explicit (target) choice was made. Instead, a disc was designated as the target by the colour green (the non-target was red) and participants merely had to hit it. In the experimental session both discs were yellow and participants chose which disc to aim for.

In Experiment 1, the small target was worth 100 points, the large target was worth 50 points and the background was worth −25 points. In Experiment 2, the reward associated with the large target was raised to 75 points, an
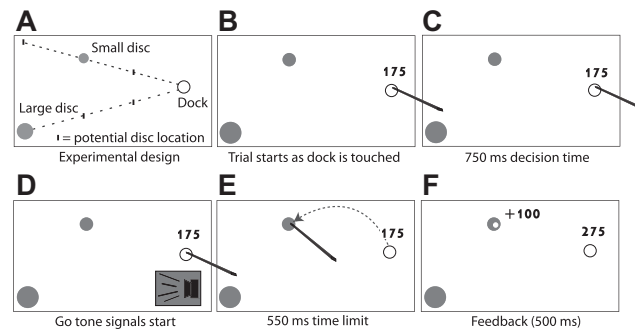
Fig. 3. Experimental procedure. Panel A: Possible stimulus configuration in Experiment 2. Panel B–F: Sequence of events as a trial unfolds. The number above the dock (white disc) represents participants' cumulative score. *Note:* stimuli are not drawn to scale and the background was black and not white.

alteration that should pose no problems for an optimal participant.

Throughout the experiment, the participant's cumulative score was displayed above the dock in blue numerals (Panel B–F, Fig. 3, – exemplified here by "175" and "275"). Participants initiated each trial by touching the dock with the stylus (Panel B), whereupon one of the unique stimulus configurations was displayed. Participants were required to maintain contact with the dock for 750 ms (decision time, Panel C). A 550 Hz tone signalled that movement should begin (Panel D). After the tone, participants had 550 ms to attempt to hit their chosen target (Panel E). Participants received feedback both on where they hit the screen and on the amount of points earned on each trial (Panel F). They could rest at any time during the experiment simply by not initiating a new trial.

Participants were told to respond within the 550 ms interval, but they were free to move as fast as they wished within that upper bound. Responses that exceeded 550 ms were recorded as 'late'. Trials in which the stylus was lifted off the dock before 100 ms had passed since the go signal were recorded as 'anticipatory'. These limits on decision and response time match those of Seydell et al.'s (2008) study. Late and anticipatory responses resulted in feedback to speed up and slow down respectively and were rerun. Late and anticipatory responses were not penalized.

For each trial, reaction time, movement time, response coordinates and the number of points earned were recorded. Reaction time was defined as the time from the go-signal to the lifting of the stylus pen off the dock area. Movement time was defined as the time from lifting the stylus off the dock area to contact with the tablet surface. Total response time was the sum of reaction time and movement time. Response coordinates were defined as the x and y position of the stylus upon first contact with the screen after the stylus had been lifted off the dock.

4.4. Data analysis

The first block in the experimental session was viewed as a warm up block and was deleted prior to any analyses. Late and anticipatory responses were discounted (see e.g., Seydell et al., 2008). For the decision session, the mean proportion of late responses was .07 (SD = .06). The mean proportion of anticipatory responses was .07 (SD = .05).

To assess participants' overall performance (i.e., efficiency), a reliable estimate of movement variability is needed. The free choice component of the decision session meant that some targets (e.g., large near targets) had few or no data points. In order to guarantee a minimum of 20 data points for each precision estimate, data sets for each participant were created by adding the last 20 data points from the practice session to the decision data (for each target location and target size combination). The mean proportion of trials excluded as outliers in the merged data sets was .01 (SD = .016). Outliers were defined as data points further from the target centre than 2.5 times the large target radius (following Gepshtein et al., 2007).

Responses were analysed separately for each participant and each factor. As in previous studies (e.g., Gepshtein et al., 2007; Trommershäuser et al., 2003a, 2003b) three assumptions were made. Firstly, it was assumed that the response distributions were bivariate normal, an assumption that we verified by inspecting chi square plots (Johnson & Wichern, 1998). Secondly, it was assumed that a participant selects a single aim point per target. As in past work, it was assumed that the centroid of each response distribution describes the aim point for that distribution. Any deviation from this aim point was assumed to be due to unexplained variability influencing planning (Churchland, Afshar, & Shenoy, 2006) and execution (van Beers, Haggard, & Wolpert, 2004) of movements. Finally, it was assumed that differences in biomechanical cost between targets were negligible (see e.g., Gepshtein et al., 2007; Trommershäuser et al., 2003a, 2003b).

To describe participants' pointing behaviour we use two metrics: aim point error and movement variability. Given a normal response distribution, circular targets and symmetric penalty regions (as employed here) the optimal aim point is the target centre. Aim point error describes the distance between a participant's aim point (the centroid of each response distribution) and the target centre.[4] The lower the aim point error – the closer to optimal the aim point choice. Movement variability was defined as the mean distance of the movement end points from the centroid of the response distribution (see e.g., Gordon, Ghilardi, & Ghez,

---

[4] Defining aim point as the [x,y] coordinate of a maximum likelihood fitted bivariate Gaussian (cf., Gepshtein et al., 2007) produced equivalent results.
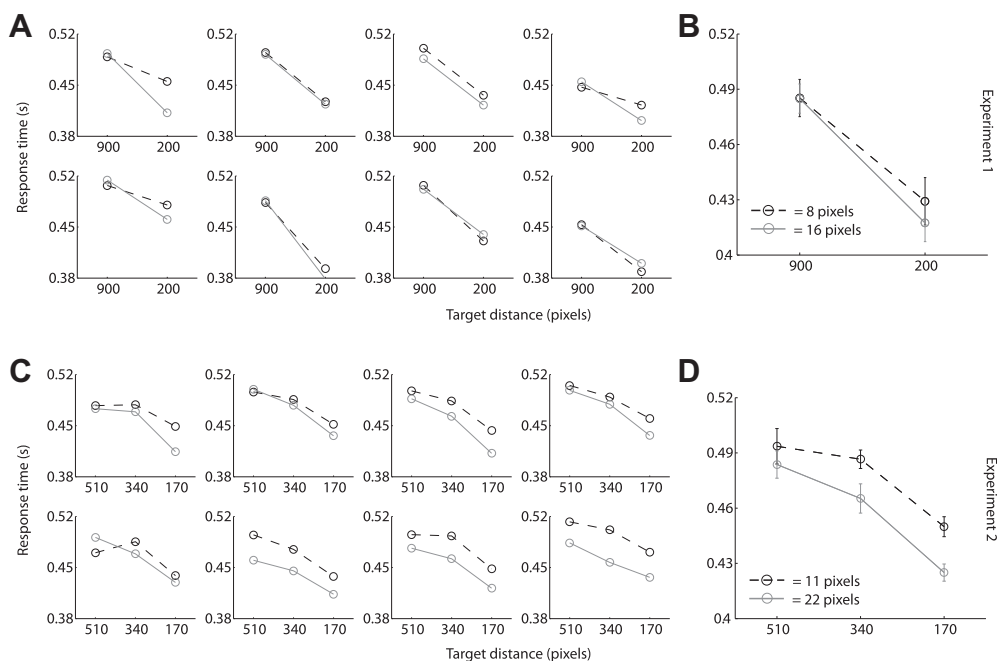
Fig. 4. Response times: group averages (B, D) and individual participants' averages (A and C) as a function of target distance, target size and experiment. The dashed line represents small targets and the full line represents large targets. The legend shows the radius of each target in pixels (1 pixel = .27 mm). Error bars are within-subject 95% confidence intervals.

1994).[5] Movement variability describes how variable participants' pointing movements were (their perceptuo-motor variability). For every participant, we computed aim point error and movement variability separately for each target size and target location combination while collapsing across target elevation.

We present both individual level plots as well as group averages for each analysis. Repeated measures ANOVA's were used to test for group-level effects. When sphericity assumptions were violated, Greenhouse–Geisser corrections were used. We first report on how participants used the available response time. Thereafter we describe how movement variability and aim point choice related to target distance and size. Following this, data describing participants' choices between the two targets (target choice) is presented. Finally, participants overall task performance is compared to that of an optimal agent.

5. Results

5.1. Overall response time

Did participants use all of the available response time as in studies with only one effective reach distance (Trommershäuser et al., 2003a,b), or did they fail to maximize time usage as found in the one previous study, by Gepsh-

tein et al., 2007, that involved different reach distances? As can be seen in Fig. 4, participants used nearly all the available response time (550 ms) when targets were far away.[6] However, for near and medium distance targets, participants used comparatively less of the available time, giving rise to a significant effect of target distance, Experiment 1: $F(1,7) = 85.14$, $p < .001$, $\eta_p^2 = .92$, Experiment 2: $F(2,14) = 247.15$, $p < .001$, $\eta_p^2 = .97$). This suggests that participants may not be maximizing time use, and therefore not be maximizing precision. If they had used the available time as efficiently as in the far condition there would be little difference between near and far targets, and the plots in Fig. 4 would show horizontal lines.

Another trend worth noting is that participants appeared to use more of the available time when they reached towards small targets (dashed lines, Fig. 4) than towards larger targets (full lines, Fig. 4). The difference between response times for small and large targets was marginal in Experiment 1 ($F(1,7) = 4.87$, $p = .063$, $\eta_p^2 = .41$) and significant in Experiment 2 ($F(1,7) = 20.87$, $p = .003$, $\eta_p^2 = .75$). We did not detect an interaction between target size and target distance in Experiment 1 ($F(1,7) = 2.17$, $p = .184$, $\eta_p^2 = .24$), but did so in Experiment 2 ($F(2,14) = 5.92$, $p = .014$, $\eta_p^2 = .46$). For a detailed analysis breaking down the effects of response times into its separate components, reaction time and movement time, see Appendix A.

---

[5] Because movement data was anisotropic, defining movement variability as the standard deviation of the response distribution necessitates multi-variate dependent variables. Using Gordon et al.'s (1994) measure provides a univariate dependent measure, making analyses easier and the exposition clearer. Seydell et al. (2008) likewise adopted a univariate measure (the square root of the determinant of the covariance matrix) to describe the variability of anisotropic data.

[6] Note that, because responses longer than the allocated response time are 'timed out' and discarded, the means must be lower than the maximal response time unless participants were able to perfectly match that response deadline on every trial.
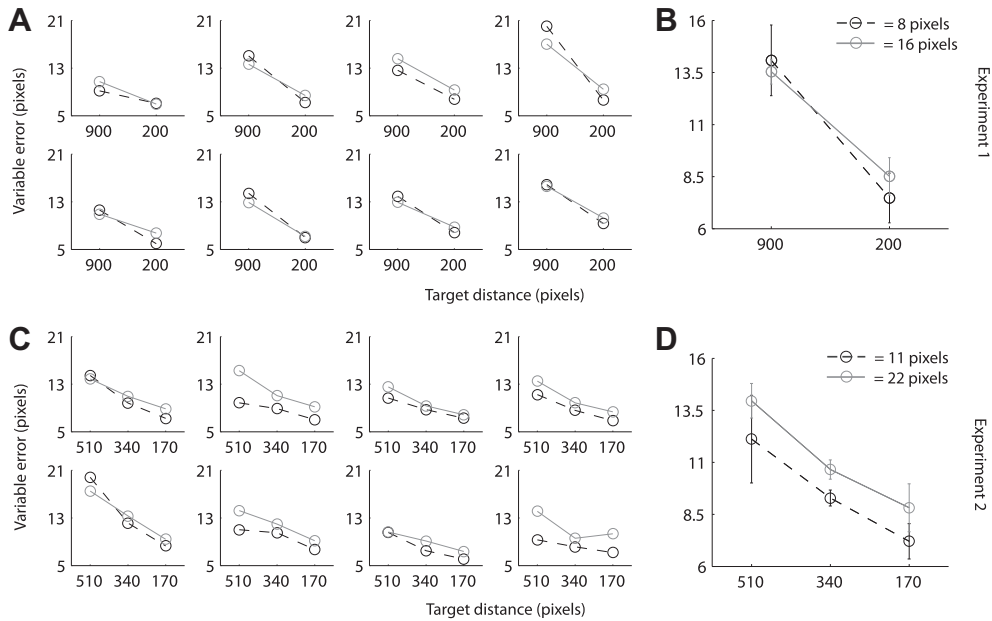
Fig. 5. Movement variability: group averages (B and D) and individual participants' averages (A and C) as a function of target distance, target size and experiment. The dashed line represents small targets and the full line represents large targets. The legend shows the radius of each target in pixels (1 pixel = .27 mm). Error bars are within-subject 95% confidence intervals.

5.2. Movement variability

Movement variability appears related both to target distance and size as shown in Fig. 5. As expected, movements to far targets were more variable than movements to near targets (Experiment 1: $F(1,7) = 63.21$, $p < .001$, $\eta_p^2 = .9$, Experiment 2: $F(1.2, 8.4) = 40.9$, $p < .001$, $\eta_p^2 = .85$). There was also evidence that movements were more variable for large targets than for small targets, with a main effect of size in Experiment 2 ($F(1,7) = 15.47$, $p = .006$, $\eta_p^2 = .69$), though not in Experiment 1 ($F(1,7) = 0.78$, $p = .41$, $\eta_p^2 = .1$) and a (marginal) interaction between target size and target distance in Experiment 1, $F(1,7) = 5.23$, $p = .056$, $\eta_p^2 = .43$ (Experiment 2: $F(1.1,7.7) = .23$, $p = .668$, $\eta_p^2 = .03$). A direct comparison between near small targets and near large targets as the likely source of that marginal interaction (see Panel B, Fig. 5) shows that movements to large near targets were more variable than those to near small targets ($t(7) = -4.14$, $p = .004$); in other words, participants in Experiment 1 aimed with greater precision to small near targets than they did to large near targets.[7] Thus both studies show evidence of effects of size and distance on movement variability (see also e.g., Chua & Elliott, 1993; Fitts, 1954).
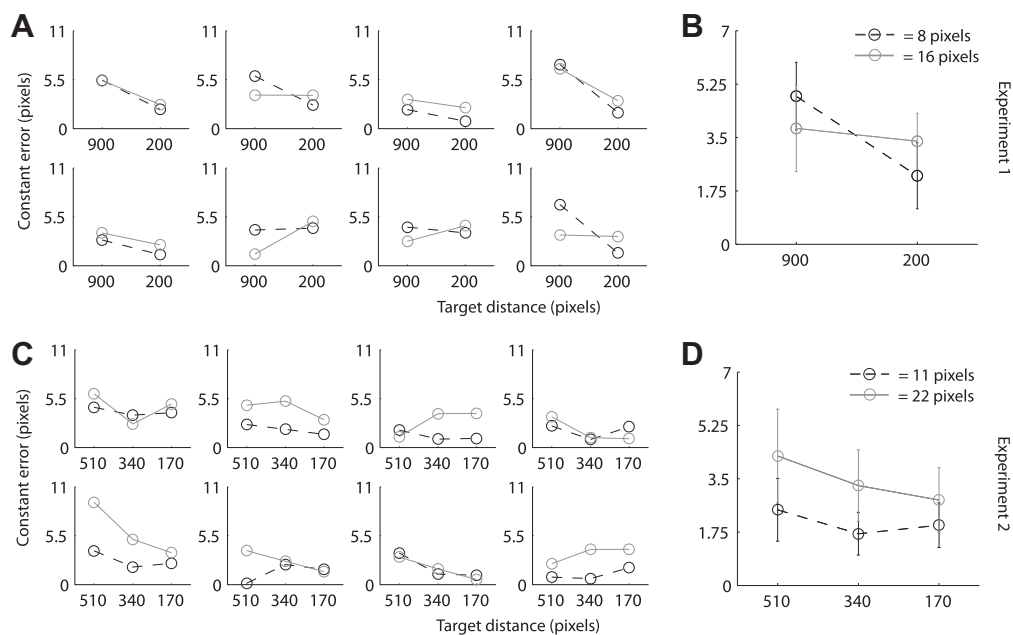
The result that movements to larger targets were more variable than movements to small targets, and that participants did not use all of the available time for near and medium distance targets ('Overall response time'), suggests that the perceptuo-motor system does not always choose the optimal movement strategy as defined in Trommershäuser et al.'s model. Instead, the perceptuo-motor system may satisfice (Simon, 1959) end-point variance or optimise a more complex cost function, an issue we will be returning to below.

5.3. Aim point error

The implicit choice participants made in our task was where to aim. 'Aim point error' (the distance between the participant's aim point and the optimal aim point) indicates how well participants chose aim points. In contrast to movement variability (Fig. 5), aim point choice (Fig. 6) seemed less strongly influenced by target distance and size. Also in contrast to movement variability, there was little evidence of consistent patterns across participants.

More specifically, aim point error showed no main effects of either size or distance in Experiment 1 (size: $F(1,7) = .02$, $p = .9$, $\eta_p^2 < .01$, distance: $F(1,7) = 4.42$, $p = .074$, $\eta_p^2 = .39$), although there was a significant interaction ($F(1,7) = 12.18$, $p = .01$, $\eta_p^2 = .64$). In Experiment 2, there was a significant effect of size only, with aiming at larger targets worse than aiming at smaller targets (size: $F(1,7) = 10.46$, $p = .014$, $\eta_p^2 = .60$; distance: $F(2,14) = 2.45$, $p = .12$, $\eta_p^2 = .25$; size × distance interaction: $F(2,14) = .86$, $p = .45$, $\eta_p^2 = .11$). In general, aim points rarely deviated from the target centre by more than 5 pixels (1.35 mm), suggesting that participants' aiming performance was good.

---

[7] There are trends in the data that suggest that for the furthest distance tested (Experiment 1, 900 pixels distance), the difference may disappear or even reverse (a trend that is also visible in the movement time plots, see Fig. 4). A possible explanation is that at very high difficulties participants relax their precision criteria even further (e.g., "there is no point in trying hard – it's too difficult"). An alternative explanation is that the far distance employed in Experiment 1 was sufficiently far, given the time deadline, as to constrain the possible pointing strategies that could be employed (i.e., it was not possible for subjects to choose different movement times for these targets).

Fig. 6. Aim point error: group averages (B and D) and individual participants' averages (A and C). Aim point errors as a function of target distance, target size and experiment. The dashed lines represent small targets and the full lines represent large targets. The legend shows the radius of each target in pixels (1 pixel = .27 mm). Error bars are within-subject 95% confidence intervals.

5.4. Target choice behaviour

The explicit choice component of the task required participants to choose between a small and a large target. To characterise participants' target choices, we compared the proportion of times the small target was chosen to the number of times it would have been chosen, had participants been optimal. In Fig. 7, the proportion of small target choices is plotted as a function of the difference between the expected value for the small and large target (ΔEV).

Participants should always choose the small target for positive ΔEV (a small choice proportion of 1), and always choose the large target for negative ΔEV (a small choice proportion of 0). Cumulative Gaussians have been fit to the individual data to assist the eye. If participants' choices were perfect, these functions would match step-functions centred on the dashed lines at 0 ΔEV.

The individual data (Experiments 1 and 2, Fig. 7), suggest that participants were sensitive to differences in expected gain, but not perfectly so. They generally picked the higher valued target. However, differences between the experiments are apparent. In Experiment 1, many participants appeared nearly un-biased - they chose small targets when these had higher EV's and large targets when these had higher EV's. In Experiment 2, on the other hand, most participants appeared biased towards the small target – choosing it even if doing so resulted in a loss relative to choosing the larger target. This is apparent in the fact that best-fitting functions appear shifted to the left of the dashed line at $\Delta EV = 0$.

To characterise this apparent bias on a group level, we pooled the data by experiment and fit cumulative Gaussian density functions. As can be seen (Fig. 7), group level fits

confirm the apparent trend and show that the small target bias is stronger in Experiment 2 than in Experiment 1 (as judged by non-overlapping 95 percentile intervals). This is noteworthy as participants appeared to have aimed for the harder-to-hit target even though aiming for the easier-to-hit larger target would have resulted in a higher return.

5.5. Task performance

We used standard methods, briefly outlined below, to assess whether participants were optimal or not. See 'Perceptuo-motor decisions & decision performance assessment' above for a detailed description and see Trommershäuser et al. (2003a, 2003b) for mathematical details.

Overall performance depended on two choices – choice of aim point and choice of target. An optimal agent always picks the best target and aim point. As the response distributions were Gaussian and the penalty region symmetric (missing a target in any direction incurred a penalty), the optimal aim point was always the centre of each target.[8]

---

[8] It has been suggested that, because of optimizing biomechanical cost, the perceptuo-motor system is biased towards undershooting targets (e.g., Lyons, Hansen, Hurding, & Elliott, 2006). In this literature 'undershoot' is used to refer to the spatial location (primary movement endpoint) of the initial (more or less) ballistic phase of movements (Lyons et al., 2006, p. 97). Since our apparatus did not allow for reliable trajectory measurements (i.e., kinematics) it is impossible to say whether primary endpoints undershoot targets. However, we can assess whether the actual endpoints (i.e., where the screen was hit) systematically undershot targets. Though we found some evidence of this type of undershooting, participants did not seem to consistently undershoot the target, which is in line with previous findings (e.g., Chua & Elliott, 1993; Fitts & Petersen, 1964).
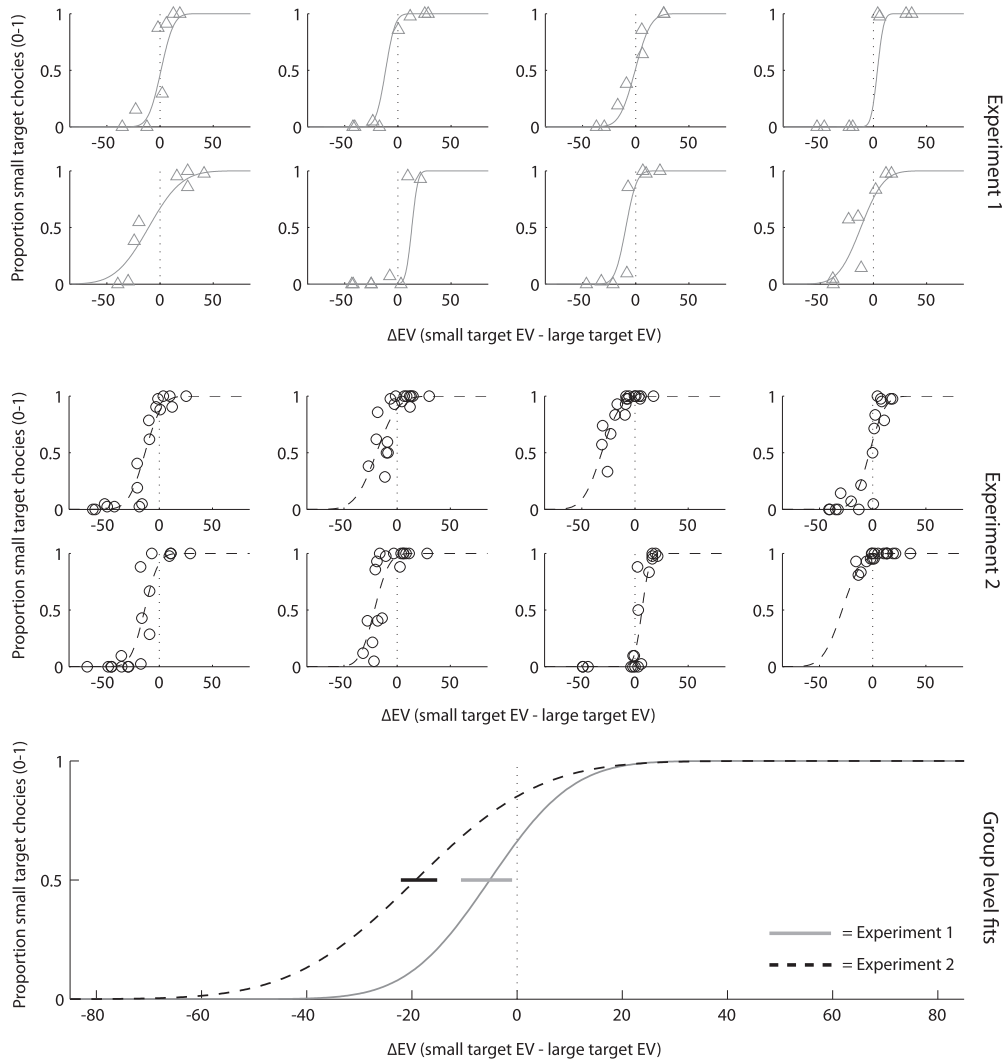
Fig. 7. Explicit target choice behaviour. Plots show the proportion of times the small target was chosen as a function of the value difference between the small and the large target (ΔEV). Positive ΔEV means that the small target was more valuable. Conversely, negative ΔEV means that the large target was more valuable. The top plots (Experiments 1 and 2) show individuals' target choices for each experiment. Cumulative Gaussian density function were fit to facilitate comparisons. The bottom plot (Group level fits) show cumulative Gaussian density functions fit to the pooled data for Experiment 1 (full grey line) and Experiment 2 (dashed black line) respectively. Error bars are bootstrapped 95 percentile intervals.

For each participant, we simulated an optimal agent, who inherited the participant's pointing variability, performing the experiment 100,000 times. The resulting distribution of average gains allowed us to estimate the expected gain of the optimal agent and the confidence in this estimate. If a participant's performance lay outside the lower 95% confidence bound on the expected gain of the optimal agent, they were classed as sub-optimal. If participants performed better than this lower bound, they were classed as statistically indistinguishable from optimal.

Fig. 8 shows participants' efficiencies for Experiment 1 and Experiment 2 (top and bottom panel, respectively). The first thing to note is that participants' efficiencies are not distributed around an efficiency of 1 as would be expected if participants had been optimal, rather they are lower. Nevertheless, six of eight participants in Experiment

1 were within the bounds of optimal performance. In Experiment 2, on the other hand, only one of eight participants' efficiencies was within the 95th percentile of the optimal expected gain. A Fisher's exact test testing for differences in optimal performance rates across the two experiments is significant ($p = .04$). Likewise a Bayesian comparison of rates (Kass & Raftery, 1995; Lee & Wagenmakers, 2005) shows that, compared with the hypothesis that the rates of optimal performance is the same across the two experiments, the hypothesis that the rates differ across experiments is 10.7 times more likely.

However, the *absolute efficiencies* across the two experiments are fairly similar. In other words, relative to the optimal agents, participants earned similar amounts in the two experiments. The lower bound on optimal performance, on the other hand, appears to be substantially
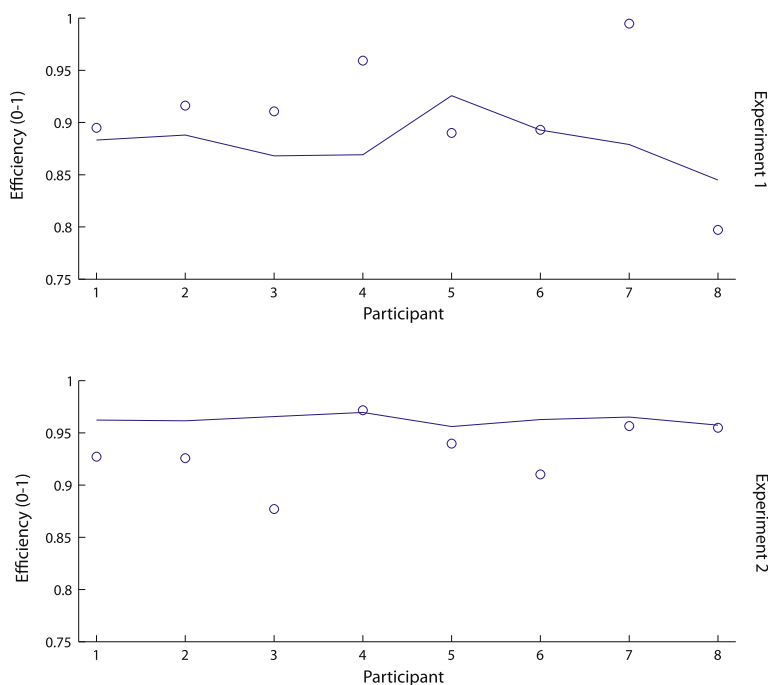
Fig. 8. Overall task efficiency (circles) and the lower bound of optimal efficiency (full line) for each participant in Experiments 1 and 2 respectively.

lower in Experiment 1 than in Experiment 2. Thus, the reason participants are classed as optimal in Experiment 1, but not in Experiment 2, appears to be due to differences in the confidence intervals and not due to differences in absolute efficiency levels. This is supported by statistical analysis.

A Bayesian *t*-test (Rouder, Speckman, Sun, & Morey, 2009) comparing the absolute efficiency levels across the two experiments shows that there is insufficient evidence to conclusively favour either the null hypothesis that they are the same or the alternative hypothesis that absolute efficiencies differ (JZS Bayes Factor in favour of alternative hypothesis = .55, unpaired *t*-test, $t(14) = -1.13$, $p = .28$). However, the same test performed on the lower 95% confidence interval of optimal performance shows overwhelming support for the hypothesis that confidence bounds differ (JZS Bayes Factor = 79 438, unpaired *t*-test, $t(14) = -9.68$, $p < 1e-6$). Thus, the difference in participant optimality across Experiments 1 and 2 appears due to a difference in how variable the optimal agent's earnings were, and not due to different levels of absolute participant performance.

6. Discussion – Experiments 1 and 2

Regardless of whether participants were classed as suboptimal or optimal, they were generally sensitive to the difference in expected gain between small and large targets and generally chose the target with the higher expected value. On the other hand, participants in Experiment 2 were biased towards choosing the small target, the target with the higher but more uncertain gain, *even* when this choice produced lower gains on average.

Both experiments further suggest that participants' perceptuo-motor behaviour may deviate from what is optimal in ways not captured by Trommershäuser et al.'s (2003a, 2003b) model. Firstly, participants appeared to favour speed over precision, producing movements to near targets that were faster than necessary. Given the speed-accuracy trade-off (Fitts, 1954; Schmidt, Zelaznik, Hawkins, Frank, & Quinn, 1979), such movements should decrease precision and therefore participants' ability to hit targets. The model does not capture such apparent satisficing as it assumes that people's movements maximize precision.

Secondly, participants appeared to relax their precision criteria when aiming for larger targets. As Trommershäuser et al.'s model assumes that precision is maximized, participants are not penalised for their less-than-maximal precision, because it is the degree of precision they actually display that is 'inherited' by the optimal model. Hence an optimal model assuming that participants always minimize movement error in perceptuo-motor task may make participants appear more optimal than they really are: Points that are lost, because participants' are hitting the targets less often than they could have done, do not enter into the comparison with the optimal agent.[9]

---

[9] Of course, participants not minimizing motor variability and/or not maximizing movement time does not, in of itself, imply that participants were sub-optimal. The experimental task did not demand that they do either, but that they earn as many points as possible. If participants were hitting the larger target 100% of the time, or were hitting the nearer targets 100% of the time, differences in movement time and variability would be largely irrelevant. However, even large near targets were not hit 100% of the time (see Appendix C for details). This means that participants could theoretically have improved their scores if they had minimized error and maximized movement time.

We followed this up in a control study (see Appendix B for details), where we tested whether participants could reach with equal precision to small and large targets when they were explicitly asked to do so. Under these conditions, three of five tested participants reached with equal precision to small and large targets. For these three participants, the null hypothesis of equal precision was more than three times as likely as the alternative hypothesis that the precision was unequal (JZS Bayes Factors > 3), with the evidence for the two other participants being inconclusive. Consequently, the failure to reach to small and large targets with equal precision in Experiments 1 and 2 does not appear to be due to a capacity limitation. By simulating optimal agents who aim with equal precision to both target sizes, one can also show that the apparent precision-satisficing in Experiments 1 and 2 was consequential. Had participants been compared to such agents, their efficiencies would have dropped significantly relative to the standard analyses presented above (paired t-test, $t(15) = -3.74$, $p = .002$, mean difference = $-.02$).

Most worryingly for claims about optimality, however, was the effect of seemingly minor changes to task parameter. Across Experiment 1 and Experiment 2, the experimental set-up was identical, and both experiments required two kinds of choices: aim point- and target choices. However, the precise stimulus configurations and the reward structure differed across the experiments. Compared to Experiment 2, Experiment 1 had smaller targets, fewer target locations, target distance differences were greater and the difference between reward for the small and the large target was greater. It turns out that these differences in task parameters were highly consequential. Experiment 1 resulted in optimal participants, whereas Experiment 2 resulted in sub-optimal participants. This result implies that optimality standards as commonly employed are not absolute but relative (see also Section 11).

Relative standards mean that classifying systems as optimal or sub-optimal is problematic without further clarification. For which experiment should we use if we wanted to evaluate the optimality of the perceptuo-motor system: Experiment 1 or Experiment 2? We next explore the effects of particular task parameters on the two sub-components of our task in greater detail.

7. The effect of task parameters on performance metrics

The key result of Experiments 1 and 2 was that seemingly innocuous changes in task parameters such as target size result in very different verdicts on optimality. Next, we simulate the effects of changes in task parameters to explore in greater detail how such changes might affect sub-optimal participants.

We focus on three properties of the tasks: hit probability, rewards and the total number of trials experienced. Note that, although we simulate changes in hit probability by manipulating target size, the same effects can be achieved by instead changing target distance (Fitts, 1954; Schmidt et al., 1979). That is, changes in hit probability can be brought about by either changing target distance or target size. However, simulating changes in target distance would require extrapolating beyond the data we have available, whereas simulating a change in target size merely requires changing a task parameter (given the assumptions of the model, i.e., that precision is maximized).

First, we consider how task parameter changes impact on participants' efficiency separately for the implicit and the explicit choice component. Then, we explore which changes across Experiments 1 and 2 are likely to explain why most participants were classed as optimal in the first experiment but not the second.

## 8. Task parameters and the optimality of aim point choices

Fig. 9 illustrates the effect of changing target size on the implicit component. Panel A shows the optimal aim point (cross) with sample hit points (grey discs) as well as two sub-optimal aim points (triangle and square). As target size increases, naturally so does the likelihood of hitting the target (Panel B), whether a participant is optimal (full line) or sub-optimal (triangles and squares). Panel C shows the hit probability for the two sub-optimal aim points as a proportion of the optimal hit probability (i.e., as efficiency). It appears that sub-optimal aiming becomes less costly in terms of absolute efficiency as target size increases.

However, as noted above, whether or not behaviour is considered optimal depends not on absolute efficiency, but the relationship between absolute efficiency and the variability of the optimal agent. Panel D shows the lower 95% confidence bound on the optimal agent's hit efficiency (dashed line). When either of the two sub-optimal aim points (triangles, squares) results in efficiencies above the dashed line, participants are classed as optimal. Conversely, efficiencies below the dashed line imply that participants are sub-optimal. As can be seen in Panel D, smaller targets lead to more variable optimal agents (and thus wider CI's). This means that small targets allow for greater deviation from the optimal aim point before participants are classed as sub-optimal.

How do these simulations fit with the results of Experiments 1 and 2? Targets in Experiment 1 were smaller than targets in Experiment 2. This means that sub-optimal participants should have been more likely to be classed as optimal in Experiment 1. This is precisely the pattern of results obtained. There were significantly more optimal participants in Experiment 1 than in Experiment 2, and this contrast appeared driven by differences in confidence intervals rather than differences in absolute efficiencies. However, we will return to the issue of the confidence interval difference between Experiments 1 and 2 below, and show that changes in target size are likely to have accounted only for a small part of the total effect.

9. Task parameters and the optimality of target choices

We next examine how a change in hit probability affects the explicit choice component of our task: the participant's choice between the small and the large target. Fig. 10 illustrates the effect of changing target size and
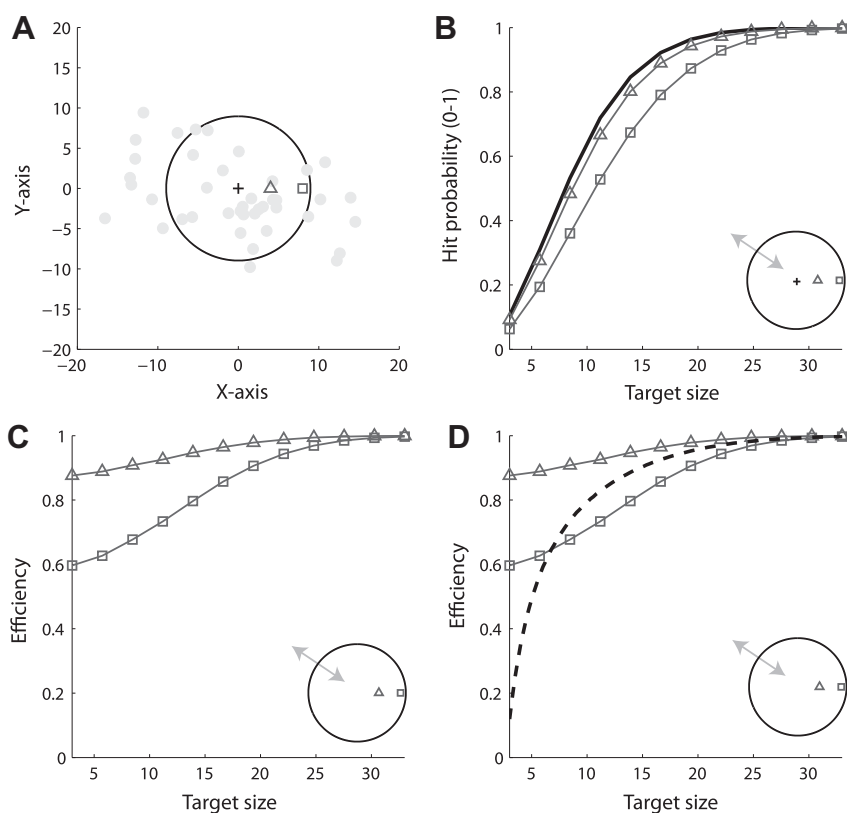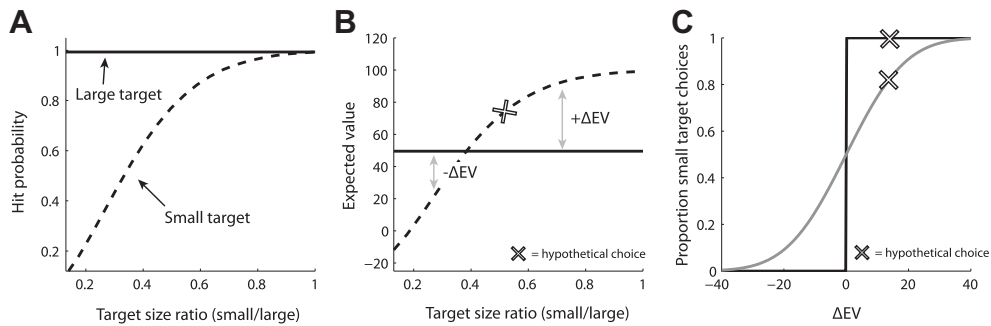
Fig. 9. Effects of changing task parameters on implicit choice. (A) Target with optimal and sub-optimal aim points (with a hypothetical response distribution [grey discs]). (B) Hit probabilities for each of the three aim points: optimal (full line), small deviation (triangles), and a large deviation (squares). (C) Efficiencies (hit probabilities normalized by optimal hit probabilities) for the two sub-optimal aim points in Panel A and B. (D) As Panel (C) but now with the lower 95% CI of optimal performance.

the resultant change in hit probability on target choice. In the simulation, we set the size of the large target such that the hit probability is close to 1 and then vary the size of the smaller target. Panel A illustrates the effect of this manipulation on hit probabilities for the small target (dashed line) relative to the large target (full line). As we increase the small target's size (i.e., increase the small/large target ratio), it becomes increasingly easy to hit (hit probability increases).

Of course, for choosing between the small (dashed line) and the large target (full line), knowing hit probabilities alone is not sufficient; one also needs to know the rewards associated with each target in order to identify the target with the greater expected value. Panel B (Fig. 10) shows the number of points a participant can expect to earn for each target given the reward structure of Experiment 1. ΔEV is the difference in expected value between the small and large target. If it is positive, the smaller target would yield greater gains on average and should be chosen (if negative, it is the large target that promises better returns). As can be seen from the graph, the large target should be chosen for a small-to-large target size ratio of up to about .4, as its expected value is higher in this range. With further increases in the size of the small target, the participant should switch and choose the small target.

Of course, to choose the more highly valued target, the participant must recognize the differences in expected value. The black step-function in Panel C (Fig. 10), illustrates the behaviour of an optimal agent who does this perfectly, and so fully maximizes expected value (as in Trommershäuser et al.'s 2003a, 2003b model). Such an agent will always choose the small target when its expected value is higher (in which case the proportion of small target choices is 1), and choose the large target when its expected value is higher. The shape of the resultant step-function is illustrative of the generally all-or-none prediction of maximization theories such as expected utility. However, people's ability to resolve differences will be limited in practice: some differences will simply be too small for the system to detect, thus making participants less than perfectly sensitive to differences in expected value. Consequently participants are unlikely to exhibit such perfect sensitivity, though one would expect better choices when ΔEV is large (because it should be more readily apparent which of the two targets is better, see e.g., Mosteller & Nogee, 1951; see also Brandstätter, Gigerenzer, & Hertwig, 2008, for this idea applied to model evaluation).

The grey function in Panel C of Fig. 10 illustrates a participant who is less-than-perfectly sensitive to differences in expected value (ΔEV). The cross in Panel B and C,

Fig. 10. Effects of changing task parameters on explicit choice. (A) The effect of the target size ratio on hit probability for the small (dashed line) and large (full line) respectively. (B) Expected value of the small (dashed line) and large (full line) target as a function of target size ratio. ΔEV is the difference in expected value between the small and the large target (see text for explanation). The cross represents a hypothetical choice situation in which the small target should be chosen. (C) Choice predictions (as proportion small target choices) for an optimal agent (black step-function) and a less-than-perfectly sensitive sub-optimal agent (grey function).

illustrates a particular choice situation in which the optimal response is to choose the small target. The less-than-perfectly sensitive participant (grey line) will only pick the optimal target ∼80% of the time – leading to a loss relative to the ideal agent (black line). From the grey function, it should also be clear that as the absolute ΔEV becomes larger, the optimal agent and the sub-optimal agent become increasingly similar.

How does this all relate to the explicit choices in Experiments 1 and 2? Given that people are likely to be less-than-perfectly sensitive to differences in expected value, we would expect better choices in the experiment that had the largest expected value differences. In Experiment 1, the difference between the small and large target reward was greater, and the differences in target distance (and hence hit probabilities) were greater, than those in Experiment 2. This should mean that, on average, Experiment 1 had choice options that were more different than those in Experiment 2. The mean absolute difference in expected value was indeed greater in Experiment 1 than in Experiment 2 (unpaired t-test, $t(13) = 2.74$, $p = .017$, mean difference = 4.23[10]). Thus, for anyone who is less-than-perfectly sensitive to ΔEV differences, Experiment 1 is easier than Experiment 2. One indication, that the explicit choice was indeed easier in Experiment 1, is that the proportion of choices that maximized expected value was greater in Experiment 1 than in Experiment 2 (unpaired t-test, $t(14) = 3.86$, $p = .0017$, mean difference = .218).

10. Task parameters and the bounds of optimal performance in overall evaluation

So far, we have explored separately the effect of individual task parameters on the two choice components, but what about their combination? Do these parameters interact and which is most influential in bringing about the differing verdicts on optimality across the two experiments?

Participants' absolute efficiencies were comparable in the two experiments, yet participants were classed as opti-

mal in Experiment 1 and sub-optimal in Experiment 2. The crucial difference seemed to be the confidence intervals on the optimal agents' performance (see Fig. 8 above). The confidence intervals are used to infer whether or not participants are optimal. The wider confidence intervals in Experiment 1 therefore meant a more lenient standard of optimality. What could account for the different standards of optimality?

The two experiments differed in seemingly innocuous ways. Target distances, target sizes and the reward structure were slightly different across the two experiments. However, one additional factor could be of importance here. A further variation between studies was the number of stimulus configurations (i.e. distinct decision problems). Manipulating the number of distinct choice configurations (here: 8 in Experiment 1 and 18 in Experiment 2), whilst keeping the number of trials for each configuration constant, results in a different number of total trials. Specifically, the total number of trials was substantially greater in Experiment 2 ($44 * 18 = 792$) compared to Experiment 1 ($44 * 8 = 352$). Everything else being equal, a greater sample size leads to tighter confidence intervals, so that this also needs to be considered.

In short, the different widths of the confidence intervals could potentially be accounted for by changes in target size (and/or distance), changes in rewards and/or changes in total sample size. We explored the effect of these factors by simulation. We simulated the eight optimal agents of Experiment 2 (one for each participant) under conditions that were made increasingly similar to those of Experiment 1. The variable of interest is how the lower 95% confidence bound on the optimal agents' efficiency changes when the various task parameters change. In other words – how do changes in task parameters affect the chances that a sub-optimal participant is classed as optimal?

Fig. 11 shows the original tight confidence bound from Experiment 2 (gray discs, 'Exp. 2') for each optimal agent (1–8, x-axis). The other symbols show the result of varying the degree of similarity between Experiment 2 and Experiment 1. The difference, between the original bound (gray discs) and the other bounds, is a measure of the effect size of a particular change. The shaded region in Fig. 11 is the between-subject 95% confidence interval on the lower

[10] One outlier ∼2.5 inter-quartile ranges from the median in Experiment 2 was excluded.
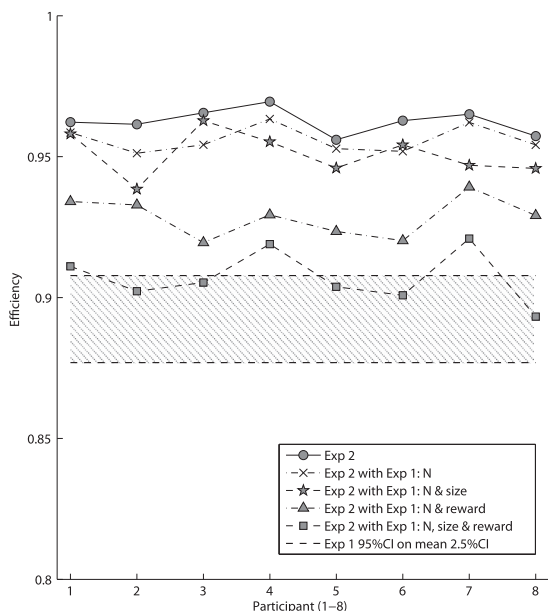
Fig. 11. The lower confidence bound of optimal efficiency as a function of sample size, target size and reward structure. The five different symbols represent the lower 2.5% bound of optimal performance as Experiment 2 is made increasingly similar to Experiment 1. The shaded area between the dashed lines represents the 95% confidence bound (bootstrapped) on the *average* lower 2.5% bound of optimal performance in Experiment 1.

confidence bound from Experiment 1. If the two experiments were identical we would expect the bounds from Experiment 2 (or the simulated variants) to lie in this region.

For the first simulation (crosses, Fig. 11) we equated the number of total trials across the two experiments. This was achieved by excluding from Experiment 2 the mid-distance targets – creating an experiment with four possible target locations (like Experiment 1). This resulted in Experiment 2 having the same number of total trials as Experiment 1, but with different rewards and different target sizes. As can be seen, the crosses that illustrate this change (Fig. 11) deviate only marginally from the original confidence bound – suggesting that the total number of trials is relatively unimportant in explaining the difference in confidence bounds across Experiments 1 and 2.

For the next three simulations, the number of total trials was kept the same as those in Experiment 1. For the second simulation, we additionally changed the target sizes to match those of Experiment 1 (stars, Fig. 11). This also had only a marginal effect on the confidence bounds. For the third simulation, we changed the rewards (but not target size). This appears to have an appreciable effect on the confidence bounds (triangles, Fig. 11).

For the final simulation, in addition to equating the total number of trials, we replaced both the target sizes and the rewards in Experiment 2 with those from Experiment 1. This resulted in the largest drop in the lower confidence bound (squares, Fig. 11). In fact, the lower optimal bound for many participants is now within a range we would expect from Experiment 1 (shaded region, Fig. 11).

The slight underestimation of variability relative to Experiment 1 (shaded area) is possibly due to the fact that far targets were nearer in Experiment 2 (this difference could not easily be simulated[11]). Because targets were nearer, they also were easier to hit (a greater proportion had hit probabilities close to 1), and therefore should result in less variable gains, which lead to tighter confidence intervals.

It is perhaps surprising that the effect of roughly doubling the number of trials has such a relatively minor effect on the width of the confidence bounds. However, what matters is the variability on the overall gain in an entire experiment. This variability depends not only on the confidence intervals for hit-probabilities of particular targets (as illustrated in Fig. 9), but also on the specific combinations of rewards, penalties and hit probabilities across targets. As Fig. 11 illustrates, these factors may interact in ways that are difficult to predict.

To illustrate this further, consider two hypothetical instantiations of our experimental task. In both instantiations, the optimal agent is faced with two target pairs. For each pair, the agent chooses the optimal target 40 times. In Experiment A, the rewards for hitting the two optimal targets are 259 and 10 points respectively. In Experiment B the rewards for the optimal targets are 115 and 100 points respectively. In both experiments the penalty for missing is -5 points, and the higher valued target is harder to hit than the lower valued target (probability of hitting the high value target $p = .5$; probability of hitting the low value target $p = .8$).

An optimal participant would, on average, earn the same number of points in both experiments (134 points). However, the confidence interval on the expected gain will be very different. In fact, Experiment A will result in confidence intervals almost twice the width of Experiment B: namely, a width of approximately 80 points versus a width of approximately 45 points. By imposing a different reward structure, a change that will not affect a strictly optimal participant, we have made it much more likely that a sub-optimal participant be classed as optimal.

The previous simulations illustrate that one can, as we did, illustrate some of the potential problems of categorizing participants as optimal or sub-optimal by breaking down the effects of particular changes in task parameters, but the final verdict on whether people are optimal or not, depends on task parameters that interact in ways not easily foreseen.

11. General discussion

It has been suggested that the perceptuo-motor system makes optimal, or near-optimal, decisions in tasks that require both explicit target choice and implicit aim point choice (Seydell et al., 2008; Trommershäuser et al., 2006). Using a novel perceptuo-motor decision task, we found that this was the case for one set of task parameters

---

[11] As noted in the introduction to this simulation section, we would have to extrapolate beyond the current data to simulate the effect of changes in target distance and so treat such differences as noise here. To the extent that the confidence bounds converge as experiments are made more similar this imperfection is of little relevance.

(Experiment 1), but not for another set of task parameters (Experiment 2). Even in Experiment 1, where participants were mostly classed as optimal, participants' efficiencies were consistently lower than 1. That is, efficiencies did not cluster around 1 as expected if participants had been optimal.

We argued that the likely origin of the difference in performance between Experiments 1 and 2 was a more lax criterion of optimality in Experiment 1. The change in criterion for optimal performance followed innocuous changes in task parameters, such as changes in target size and target reward. These parameter changes do not affect the expected performance of an optimal agent and therefore do not affect participants who are truly optimal. Yet, the changes had dramatic effects on whether participants were classed as optimal. Through simulations, we showed that task parameter dependent optimality is a general problem that extends beyond the specific parameters of our experiments.

Furthermore, departures from optimality were evident in the raw data itself. Participants reached with greater precision to small targets than to large targets, which suggests that humans sometimes satisfice rather than maximize precision. Moreover, we showed that the apparent failure to maximize precision was consequential. Had participants been compared to optimal agents who aimed with equal precision to both target sizes, their efficiencies would have been lower than the ones reported here. We further established that participants seem capable of reaching with equal precision to small and large targets when they are explicitly asked to do so.

Similar to Gepshtein et al. (2007), we also found that participants did not make use of all the available response time when pointing to near targets. Whatever the source of this time under-utilisation, it is likely sub-optimal in the present tasks. Spending more time on harder-to-hit targets should increase the likelihood of hitting those targets (Fitts, 1954; Schmidt et al., 1979), leading in turn to greater earnings.

Both of these limitations in performance suggest that it may be more appropriate to view performance as sub-optimal, and they highlight the model-dependency of optimal agent-based comparisons. Neither the failure to maximise precision when reaching to large targets, nor the failure to use all available response time when reaching to near targets, both which likely impeded participants' performance, are captured when precision-maximization is assumed as in Trommershäuser et al.'s (2003a, 2003b) optimal agent model. Instead, the optimal agent simply inherits the participants' actual variability in hitting targets.

The deeper problem, here, concerns which constraints and cost assumptions to include in one's model. It might be possible to build optimal agents based on independent assessment of how precise participants' pointing behaviour could be if they were performing at their maximal level. Conversely, one might also assume that participants are optimizing a different cost function; if this function (whatever it is) were to be used for analysis instead, one might potentially be tempted to label their actual precision as optimal.

However, there are clear conceptual limitations that arise here: If one includes all the factors influencing behaviour then 'optimality' seemingly follows by definition: limitations in performance are simply translated into system constraints and/or appropriate trade-offs between (subjective) costs. In the limit, findings of optimality become trivial and cease to be of theoretical interest.

Methodologically, it may be invaluable to *assume* optimality and iteratively seek to incorporate constraints as a way of understanding the workings of the system. Ideal observer analysis (Geisler, 2011) and rational analysis (Anderson, 1990) can be used in this way: to, for example, constrain the search for plausible models (see e.g., Schrater & Kersten, 2002) and to facilitate the design of process models (see, e.g., Howes, Lewis, & Vera, 2009).

Such methodological use of optimal models, however, should be distinguished from contexts where substantive statements about optimality per se (or lack thereof) are at stake. Such statements are frequent in the literature, whether they are phrased in terms of system optimality or, as is more common in the cognitive and social literature, in terms of 'rationality'. The statements discussed in the introduction, which contrast the degree of optimality between perception and cognition, provide just one small sample of such claims.

The results presented in this paper, make clear how difficult such claims are to establish and sustain. A given empirical assessment of human decision-making performance necessarily involves a specific task that determines how difficult the decision problem is. Everything else being equal, an easier task will naturally lead to a more optimistic view of human decision-making performance compared with a harder task.

To illustrate task difficulty, consider a choice pattern that violates maximization of expected value. When asked to choose between a gamble that yields $2500 with a .33 probability, $2400 with a .66 probability and $0 with a .01 probability and a gamble that yields $2400 with certainty, most tend to pick the latter (Kahneman & Tversky, 1979, pp. 265–266). The expected value of the former is $2409 and the expected value of the latter is $2400. This may be deemed sub-optimal, yet the expected loss of choosing the modal response is only 0.4%. Moreover, assuming noisy computational processes (Faisal, Selen, & Wolpert, 2008), people might not even be able to distinguish between expected values that differ so little.

It may seem obvious that comparing performance across tasks that vary in difficulty is problematic. Nevertheless, there has been little attempt to equate decision-problems when making comparisons across modalities or cognitive domains. When such attempts have been made, little – if any – difference in performance has been found (see Jarvstad, Hahn, Rushton, & Warren, 2013; Jarvstad et al., 2012; Wu, Delgado, & Maloney, 2009). Furthermore, the issue is not limited to comparison such as that between perception and cognition.

All performance evaluations are inherently relative to the tasks that make up the evaluation. Consequently, the evaluations they provide are relative, not absolute. Optimality analyses that evaluate performance relative to an optimal agent do not circumvent this limitation, rather

they compound it, because such model-dependent comparisons depend also on the constraints included in the optimal agent.[12]

That statements about optimality are specific and conditional in this way – that is, a behaviour is optimal *given* a task of this difficulty, and *given* these capacity constraints included in the optimal agent— may be appreciated by many, however the literatures typically do not make this explicit, and many claims are simply unsustainable once this fact is taken into account.

The fact that performance evaluations are always task relative makes comparative evaluations across systems difficult. The surprising extent to which results reported in this paper further reveal statements about optimality to be sensitive to task parameters that are unlikely to be the focus of attention compounds these difficulties. As we have shown, it is possible to make people appear optimal, or sub-optimal, by seemingly innocuous changes to decision-making tasks, because changes to task parameters have consequences for the confidence intervals delineating optimal performance in ways that are difficult to foresee. Consequently, a comparable level of performance can lead to opposite verdicts on optimality across tasks.

In conclusion, our results suggest that statements about optimality, both within and across domains, are likely to be considerably more fragile than the literature presently assumes.

Acknowledgement

Appendix A.

Here we break down the response times into its two separate components: reaction time and movement time and analyse them separately.

A.1. Movement time

As can be seen in Fig. A1, movement times show approximately the same pattern of results as the overall response times (movement time + reaction time) reported on

---

[12] Sometimes it is argued that some ideal observer models *do* provide absolute standards (e.g., Ma, 2012), provided that the models are defined relative to the information available at the sensory stage (e.g., the retina). However, even these models are conditioned on something – the constraints of the sensors. It seems to us that there is no principled reason for thinking that this particular constraint - the sensors being the way they are - is any different to any other constraint, for example a particular cortical structure (e.g., V1) being the way it is. In fact, recently the opposite argument has been made: that higher cortical structures have constrained the evolution of our eyes (Beck, Ma, Pitkow, Latham, & Pouget, 2012). Regardless, modelling *any* constraint, whether at the sensor- or later levels, implies that performance standards are relative (to those constraints) and not absolute. Nevertheless, fully unconstrained models can perhaps be specified. For light detection, this may involve conditioning on the whole light spectrum – not just the one that is visible to our retina. Similarly, for the motor system such an unconstrained model might have no constraints on precision and speed. In both cases, it seems unlikely that the system would perform optimally given such an unconstrained model.

in the main paper. When participants point to far targets they use more time than when pointing to near targets (Experiment 1: ($F(1,7) = 120.68$, $p < .001$, $\eta_p^2 = .95$, Experiment 2: $F(2,14) = 289.28$, $p < .001$, $\eta_p^2 = .98$). Recall that the time available for each pointing movement was the same regardless of distance. If participants had used nearly all the available time (550 ms) to point at targets, the plots in Fig. A1 would have been horizontal lines. Faster movement times for nearer targets, compared to far targets, suggests that movements to near targets were faster than necessary. Another interesting trend is that participants moved more slowly towards small targets than they did to large targets (Experiment 1: $F(1,7) = 21.31$, $p = .002$, $\eta_p^2 = .75$, Experiment 2: $F(1,7) = 41.81$, $p < .001$, $\eta_p^2 = .86$). Given the speed-accuracy trade-off (Fitts, 1954; Schmidt et al., 1979) this also suggests that people were sacrificing precision for movement speed. In Experiment 1 there was also a significant interaction $F(1,7) = 9.54$, $p = .018$, $\eta_p^2 = .58$), whereas in Experiment 2 it did not reach significance ($F(2,14) = 41.81$, $p = .09$, $\eta_p^2 = .29$).

A.2. Reaction time

Reaction times for most participants show the opposite, albeit weaker, trend to that of movement time. Firstly, reaction time decreases with increases in target distance (Experiment 1: $F(1,7) = 11.32$, $p = .012$, $\eta_p^2 = .62$, Experiment 2: $F(2,14) = 20.01$, $p < .001$, $\eta_p^2 = .74$). Thus, participants to some extent appear to trade-off movement time with response time, initiating movements faster to targets that are far away. Secondly, for some participants reaction times to small targets are slower than reaction times to large targets. This trend was significant in Experiment 2 ($F(1,7) = 8.125$, $p = .025$, $\eta_p^2 = .54$) and marginal in Experiment 1 ($F(1,7) = 4.92$, $p = .06$, $\eta_p^2 = .41$). In Experiment 1 there was also a significant interaction ($F(1,7) = 5.72$, $p = .048$, $\eta_p^2 = .45$), whereas none was detected in Experiment 2 ($F(2,14) = .376$, $p = .693$, $\eta_p^2 = .05$) (Fig. A2).

Appendix B. Is target size-independent precision achievable?

Here we give brief details on the small control study reported on in "Discussion – Experiments 1 and 2". Only significant deviations from Experiments 1 and 2 are noted.

B.1. Method

B.1.1. Participants and instructions
Five participants took part. All participants (except Participant 2 who was the first author) were paid at an hourly rate of £10. Participants were informed that speeded movements are variable and that even if one aims for the same spot on each reach, the actual end point will deviate randomly from trial to trial. They were further informed about the difference between accuracy and precision. Participants were instructed to minimize the distance between each end point and the perceived target centre. All participants (except Participant 2) were naive as to the purpose of the study.
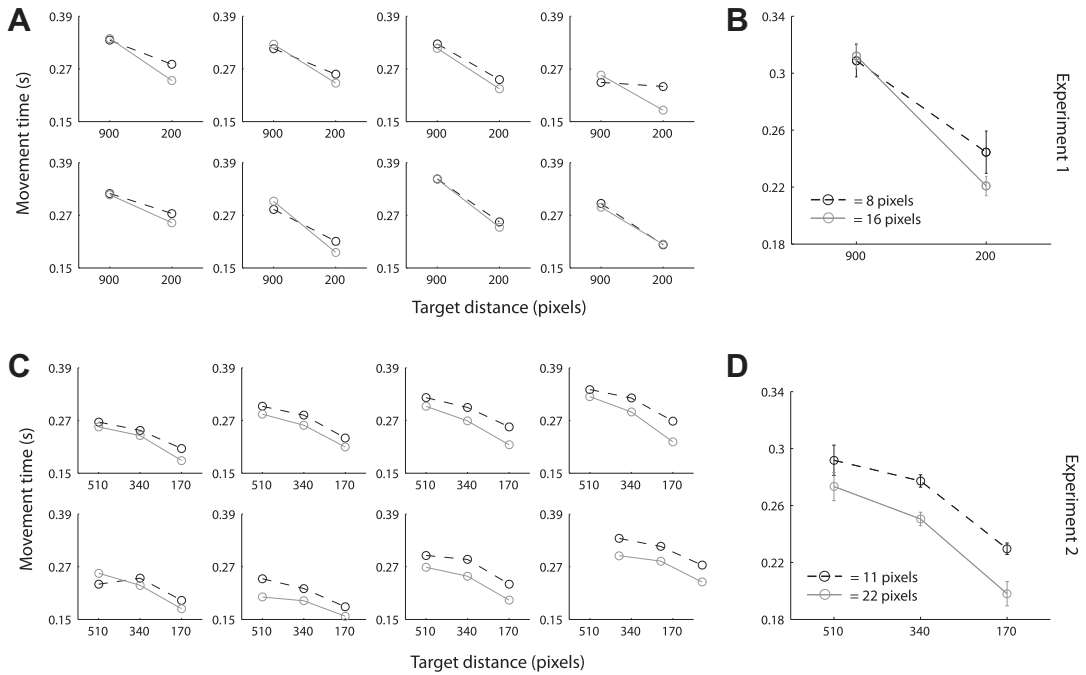
Fig. A1. Movement time: group averages (B and D) and individual movement times (A and C) as a function of target distance, target size and experiment. The dashed line represents small targets and the full line represents large targets. The legend shows the radius of each target in pixels (1 pixel = .27 mm). Error bars are within-subject 95% confidence intervals.
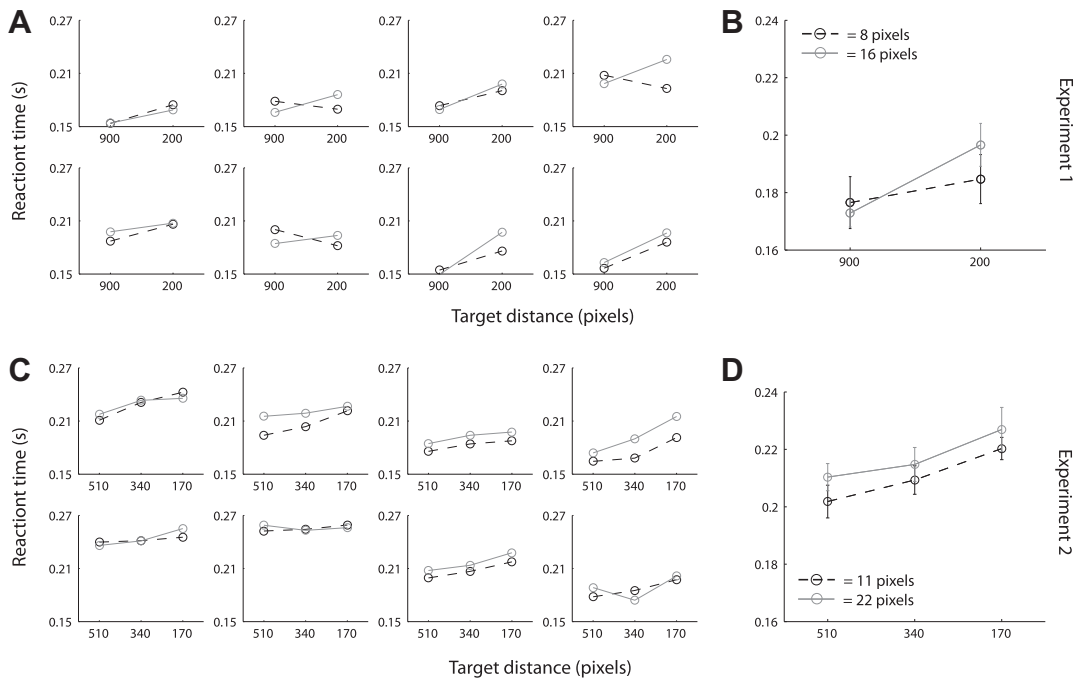

Fig. A2. Reaction time: group averages (B and D) and individual reaction times (A and C) as a function of target distance, target size and experiment. The dashed line represents small targets and the full line represents large targets. The legend shows the radius of each target in pixels (1 pixel = .27 mm). Error bars are within-subject 95% confidence intervals.

B.1.2. Stimuli, experimental design and procedure

On each trial, one target disc was displayed to the left of the dock at a distance of ∼9.2 cm (340 pixels) at one of five angles (−15°, −7.5°, 0°, +7.5°, +15°). Targets were either small (radius ∼2.9 mm/11 pixels) or large (radius ∼5.9 mm/22 pixels) yellow discs. On each trial a random
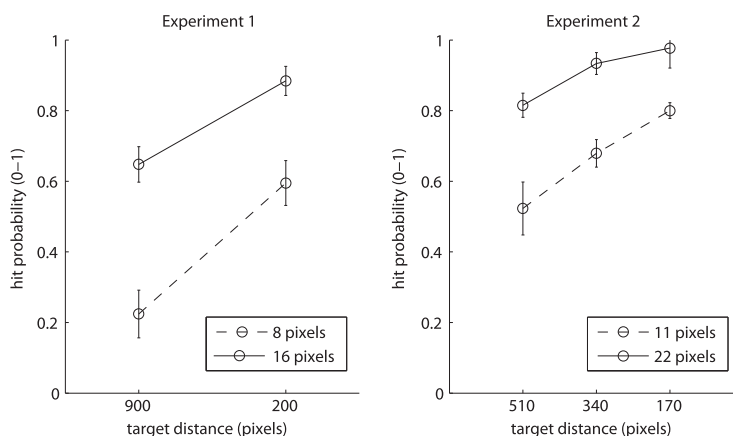
Fig. C1. Hit probabilities for Experiments 1 and 2. Each panel shows average hit probability (across participants) for each experiment as a function of target distance and size (pixel radius). The dashed line represents small targets and the full line represents large targets. The legend shows the radius of each target in pixels (1 pixel = .27 mm). Error bars are within-subject 95% confidence intervals.

target size and target angle was selected for presentation. In total, 300 small and 300 large non-late and non-anticipatory trials were collected.

Participants received feedback identical to that in Experiments 1 and 2 on where they hit the screen (but did not receive any points for hitting the targets as we wanted to minimize the incentive for satisficing).

B.1.3. Data analysis and results

For each participant we collapsed across target angle, creating one small target and one large target distribution (see Gordon et al., 1994). On a group level, there was a very small but detectable effect of target size on movement time ($t(4) = 4.18$, $p = .014$, mean difference = 4.2 ms), indicating that movements to smaller targets were marginally slower than movements to large targets. We did not detect an effect of target size on either response time or reaction time ($t(4) = 1.02$, $p = .37$, mean difference = 1.6 ms; $t(4) = 2.51$, $p = .066$, mean difference = 5.9 ms).

We compared each participant's movement variability for small targets to their movement variability for large targets using un-paired t-tests. The t-statistic was used to derive JZS Bayes Factors (Rouder et al., 2009), which allow inferences in favour of the null hypothesis (movement variability does not differ across target size) as well as in favour of the alternative hypothesis (movement variability does differ across target size).

Three of five participants reached with equal precision to small and large targets (JZS Bayes Factors > 3) and the evidence for two of five participants was inconclusive. If one performs the same analyses on the data for Experiments 1 and 2, the results are markedly different – most participants reached with greater precision to small targets (12 of 16, JZS Bayes factors < 0.33) and only three of sixteen participants reach with equal precision to small and large targets (JZS Bayes Factor > 3). A group-level analysis provides similar evidence, showing that participants had a lower average difference (between small and large targets)

in movement variability compared to those in Experiments 1 and 2 ($t(19) = 3.86$, $p = .001$, mean difference = 1.33).[13]

Appendix C. Target hit probabilities

See Fig. C1.

References

Anderson, J. R. (1990). *The adaptive character of thought*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Barlow, H. B. (1962). Measurements of the quantum efficiency of discrimination in human scotopic vision. *Journal of Physiology, 160*, 159–188.

Battaglia, P. W., & Schrater, P. R. (2007). Humans trade-off viewing time and movement duration to improve visuomotor accuracy in a fast reaching task. *The Journal of Neuroscience, 27*, 6984–6994.

Beck, J. M., Ma, W. J., Pitkow, X., Latham, P. E., & Pouget, A. (2012). Not noisy, just wrong: The role of suboptimal inference in behavioral variability. *Neuron, 75*, 30–39.

Birnbaum, M. H. (2008). New paradoxes of risky decision-making. *Psychological Review, 115*, 463–501.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*, 433–436.

Brandstätter, E., Gigerenzer, G., & Hertwig, R. (2008). Risky choice with heuristics: Reply to Birnbaum (2008), Johnson, Schulte-Mecklenbeck, and Willemsen (2008), and Rieger and Wang (2008). *Psychological Review, 115*, 281–290.

Chater, N., & Oaksford, M. (2008). The probabilistic mind: Where next? In N. Chater & M. Oaksford (Eds.), *The probabilistic mind: Prospects for Bayesian cognitive science* (pp. 501–514). Oxford: Oxford University Press.

---

[13] There are two potential issues with directly comparing Experiment 3 with Experiments 1 and 2. Firstly, Experiment 2 generally had more data than Experiment 3 (where there were 300 large and 300 small target samples). A reviewer also questioned whether precision differences may be affected by target distance (in Experiment 3 one mid-distance was used, whereas Experiments 1 and 2 used two and three different target distances respectively). As a control, we therefore fit bivariate Gaussians to the mid-distance data in Experiment 2 (the same distance as used here). The parameters of these maximum-likelihood fits were used to simulate participants in Experiment 2 reaching, the same number of times as here, to mid-distance targets only. Even when distance and sample size has been equated, the average precision difference between small and large targets is larger in Experiment 2 than here ($t(11) = 3.7$, $p = .003$, mean difference = .96).

Chua, R., & Elliott, D. (1993). Visual regulation of manual aiming. *Human Movement Science, 12*, 365–401.

Churchland, M., Afshar, A., & Shenoy, K. (2006). A central source of movement variability. *Neuron, 52*, 1085–1096.

Dean, M., Wu, S.-W., & Maloney, L. T. (2007). Trading off speed and accuracy in rapid goal-directed movements. *Journal of Vision, 7*, 1–12.

Evans, J. St. B. T. (1993). Bias and rationality. In K. I. Mankeltow, & D. E. Over (Eds.), *Rationality: Psychological and philosophical perspectives* (pp. 6–30). London: Routeledge.

Evans, J. St. B. T., & Over, D. E. (1996). *Rationality and reasoning.* Hove, UK: Psychology Press.

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews: Neuroscience, 9*, 292–303.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movements. *Journal of Experimental Psychology, 47*, 381–391.

Fitts, P. M., & Petersen, J. R. (1964). Information capacity of discrete motor responses. *Journal of Experimental Psychology, 67*, 103–112.

Geisler, W. S. (2011). Contributions of ideal observer theory to vision research. *Vision Research, 51*, 771–781.

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

Gepshtein, S., Seydell, A., & Trommershäuser, J. (2007). Optimality of human movement under natural variations of visual-motor uncertainty. *Journal of Vision, 7*, 1–18.

Gordon, J., Ghilardi, M. F., & Ghez, C. (1994). Accuracy of planar reaching movements: Independence of direction and extent variability. *Experimental Brain Research, 99*, 97–111.

Harris, C. M., & Wolpert, D. M. (1998). Signal-dependent noise determines motor planning. *Nature, 394*, 780–784.

Hertwig, H., Barron, G., Weber, E. U., & Erev, I. (2004). Decisions from experience and the effect of rare events in risky choice. *Psychological Science, 15*, 534–539.

Howes, A., Lewis, R. L., & Vera, A. (2009). Rational adaptation under task and processing constraints: Implications for testing theories of cognition and action. *Psychological Review, 116*, 717–751.

Hudson, T. E., Maloney, L. T., & Landy, M. S. (2008). Optimal movement timing with temporally asymmetric penalties and rewards. *PLoS Computational Biology, 4*, 1–9.

Jarvstad, A., Hahn, U., Rushton, S. K., & Warren, P. A. (2013). Perceptuo-motor, cognitive, and description-based decision-making seem equally good. *Proceedings of the National Academy of Sciences USA, 110*, 16271–16276.

Jarvstad, A., Rushton, S. K., Warren, P. A., & Hahn, U. (2012). Knowing when to move on: Cognitive and perceptual decisions in time. *Psychological Science, 23*, 589–597.

Johnson, R. A., & Wichern, D. W. (1998). *Applied multivariate statistical analysis.* Upper Saddle River, NJ: Prentice Hall.

Kahneman, D., Slovic, P., & Tversky A. (1982). *Judgment under uncertainty: Heuristics and biases.* Cambridge: Cambridge University Press.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49–81). New York: Cambridge University Press.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica, 47*, 263–292.

Kahneman, D., & Tversky, A. (1996). On the reality of cognitive illusions. *Psychological Review, 103*, 582–591.

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*, 773–795.

Lee, M. D., & Wagenmakers, E.-J. (2005). Bayesian statistical inference in psychology: Comment on Trafimow (2003). *Psychological Review, 112*, 662–668.

Lyons, J., Hansen, S., Hurding, S., & Elliott, D. (2006). Optimizing rapid aiming behaviour: movement kinematics depend on the cost of corrective modifications. *Experimental Brain Research, 174*, 95–100.

Ma, W. J. (2012). Organizing probabilistic models of perception. *Trends in Cognitive Sciences, 16*, 511–518.

Maloney, L. T., Trommershäuser, J., & Landy, M. S. (2007). Questions without words: A comparison between decision-making under risk and movement planning under risk. In W. Grey (Ed.), *Integrated models of cognitive systems* (pp. 297–315). Oxford University Press.

Mosteller, F., & Nogee, F. (1951). An experimental measurement of utility. *Journal of Political Economy, 59*, 371–404.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision, 10*, 437–442.

Rouder, J. N., Speckman, P. L., Sun, D., & Morey, R. D. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*, 225–237.

Schmidt, R. A., Zelaznik, H. N., Hawkins, B., Frank, R. S., & Quinn, J. T. (1979). Motor-output variability: A theory for the accuracy of rapid motor acts. *Psychological Review, 86*, 415–451.

Schrater, P. R., & Kersten, D. (2002). Vision, psychophysics, and Bayes. In R. P. N. Rao, B. A. Olshausen, & M. S. Lewicki (Eds.), *Probabilistic models of the brain: Perception and neural function.* MIT Press.

Seydell, A., McCann, B. C., Trommershäuser, J., & Knill, D. C. (2008). Learning stochastic reward distributions in a speeded pointing task. *The Journal of Neuroscience, 28*, 4356–4367.

Simon, H. A. (1959). Theories of decision-making in economic and behavioral science. *The American Economic Review, 49*, 253–283.

Thaler, R. H., Tversky, A., Kahneman, D., & Schwartz, A. (1997). The effect of myopia and loss aversion on risk taking: An experimental test. *The Quarterly Journal of Economics, 112*, 647–661.

Trommershäuser, J., Landy, M. S., & Maloney, L. T. (2006). Humans rapidly estimate expected gain in movement planning. *Psychological Science, 17*, 981–988.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003a). Statistical decision theory and the selection of rapid, goal-directed movements. *Journal of the Optical Society of America A, 20*, 1419–1432.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003b). Statistical decision theory and trade-offs in the control of motor response. *Spatial Vision, 16*, 255–275.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2008). Decision-making, movement planning and statistical decision theory. *Trends in Cognitive Sciences, 12*, 291–297.

van Beers, R. J., Haggard, P., & Wolpert, D. M. (2004). The role of execution noise in movement variability. *Journal of Neurophysiology, 91*, 1050–1063.

Whiteley, L., & Sahani, M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *Journal of Vision, 8*, 1–15.

Wu, S.-W., Delgado, M. R., & Maloney, L. T. (2009). Economic decision-making compared with an equivalent motor task. *Proceedings of the National Academy of Sciences USA, 106*, 6088–6093.