Reward and Prediction Errors in Bayesian Sensorimotor Control

Cedric Galetzka

September 2018

A thesis submitted for the degree of *Master of Science*

MSc Cognitive Science – Embodied Cognition

Faculty of Human Sciences

University of Potsdam

First advisor: Dr. André Krügel

Second advisor: Professor Dr. Ralf Engbert

Published online on the
Publication Server of the University of Potsdam:
https://doi.org/10.25932/publishup-50350
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-503507

Abstract

Midbrain dopamine neurons invigorate responses by signaling opportunity costs (tonic dopamine) and promote associative learning by encoding a reward prediction error signal (phasic dopamine). Recent studies on Bayesian sensorimotor control have implicated midbrain dopamine concentration in the integration of prior knowledge and current sensory information. The present behavioral study addressed the contributions of tonic and phasic dopamine in a Bayesian decision-making task by alternating reward magnitude and inferring reward prediction errors. Twenty-four participants were asked to indicate the position of a hidden target stimulus under varying prior and likelihood uncertainty. Trial-by-trial rewards were allocated based on performance and two different reward maxima. Overall, participants' behavior agreed with Bayesian decision theory, but indicated excessive reliance on likelihood information. These results thus oppose accounts of statistically optimal integration in sensorimotor control, and suggest that the sensorimotor system is subject to additional decision heuristics. Moreover, higher reward magnitude was not observed to induce enhanced response vigor, and was associated with less Bayes-like integration. In addition, the weighting of prior knowledge and current sensory information proceeded independently of reward prediction errors. Taken together, these findings suggest that the process of combining prior and likelihood uncertainties in sensorimotor control is largely robust to variations in reward.

Keywords: Bayesian decision theory, reward prediction error, sensorimotor control, prior-likelihood integration, dopamine

Zusammenfassung

Inwieweit prägen Belohnungen die Integration von vorherigem Wissen und sensorischen Informationen im Kontext der Bayesianischen Entscheidungstheorie? Untersuchungen mit Parkinson-Patienten haben gezeigt, dass die Dopamin-Verfügbarkeit in den Basalganglien Integrationsprozesse in der Sensomotorik beeinflussen. Dopaminerge Neuronen schütten Dopamin tonisch und phasisch aus, wobei diese Modi verschiedenen Funktionen unterliegen, wie dem Signalisieren von Opportunitätskosten oder der Unterstützung assoziativen Lernens. Die Konzentration tonisch freigesetzten Dopamins richtet sich nach Belohnungsgrößen, wogegen phasische Dopamin-Komponenten durch Fehler in der Belohnungserwartung hervorgerufen werden. Die Bedeutung dieser Variablen in sensomotorischem Lernen ist jedoch größtenteils unerforscht. In der vorliegenden Verhaltensstudie wurden vierundzwanzig gesunde Teilnehmer gebeten, eine sensomotorische Schätzaufgabe durchzuführen, in der Belohnungsgrößen manipuliert und Belohnungserwartungsfehler abgeleitet wurden. Es wurde vermutet, dass positive Abweichungen in der Belohnungsvorhersage zu erhöhter Gewichtung von sensorischen Informationen durch den Influx phasischen Dopamins führen. Höhere Belohnungsgrößen sollten dagegen aufgrund vermehrter Opportunitätskosten mit beschleunigten Reaktionen verbunden sein. Das Verhalten der Teilnehmer hat gezeigt, dass aktuelle und a priori Informationen größtenteils unabhängig von Belohnungsgrößen und Belohungserwartungsfehlern integriert werden. Dieses Ergebnis deutet darauf hin, dass “Prior” und “Likelihood” unabhängig von belohnungsrelevanten Prozessen repräsentiert werden, welche in Zusammenhang mit der Dopamin-Konzentration in den Basalganglien stehen. Darüber hinaus entsprachen die Resultate lediglich qualitativ der Bayesianischen Entscheidungstheorie und widersprechen somit früheren Berichten von statistisch-optimaler Integration. Da sensorische Informationen über alle Bedingungen hinweg übermäßig hoch gewichtet wurden, legt diese Studie nahe, dass das sensomotorische System zusätzlichen systematischen Urteilsverzerrungen unterliegt.

Schlüsselwörter: Belohnungserwartungsfehler, Belohnungsgrößen, Bayesianische Entscheidungstheorie, sensomotorische Integration, Dopamin

Selbstständigkeitserklärung

Hiermit bestätige ich, dass die vorliegende Masterarbeit mit dem Titel “Reward and Prediction Errors in Bayesian Sensorimotor Control” von mir eigenhändig erstellt wurde. Ich habe alle relevanten Quellen ordnungsgemäß gekennzeichnet und keine zusätzlichen Hilfsmittel verwendet oder verschwiegen.

....

Ort, Datum Unterschrift

Reward and Prediction Errors in Bayesian Sensorimotor Control

To reach metamorphosis, the Japanese oakblue butterfly employs a curious strategy involving sophisticated manipulation of neurochemistry and ants (Hojo, Pierce, & Tsuji, 2015). While in its larvae stage, the oakblue secretes a fluid which, once ingested, causes ants to abandon their colonies and direct their efforts into protecting the caterpillar from nearby predators (Heil, 2015; Hojo et al., 2015). Hojo et al. (2015) found that ants which consumed the caterpillar's secretions tend to neglect alternative food sources, and display altered levels of the neurotransmitter dopamine. In humans, dopamine is implicated in a wide range of areas, including sleep (Monti & Monti, 2007), personality (Depue & Collins, 1999), addiction (Wise, 1996), reward processing (Olds & Milner, 1954), and motor control (Carlsson, Lindqvist, Magnusson, & Waldeck, 1958). Notably, midbrain dopamine neurons are involved in motivation by signaling opportunity costs for action, and in associative learning by encoding an error between a prediction and a reward (Niv, Daw, Joel, & Dayan, 2007; Schultz, Dayan, & Montague, 1997). Within the framework of Bayesian decision theory, midbrain dopamine concentration has recently been linked to alterations in sensorimotor learning (Vilares & Kording, 2017). The present study investigates the impact of reward on sensorimotor integration within the framework of Bayesian statistics. Specifically, it is examined whether reward magnitude affects response vigor, and whether reward prediction errors lead to a differential weighting of prior knowledge and current sensory information.

The Neurobiology of Action and Reward

Located within the basal ganglia, the putamen and caudate form the dorsal striatum, a midbrain structure implicated in motor control, as well as movement planning and learning (Alexander & Crutcher, 1990; Alexander, DeLong, & Strick, 1986; Graybiel, 1990). The dorsal striatum receives dopaminergic input via the nigrostriatal pathway from the A9 dopamine cell cluster of the substantia nigra pars compacta (SNc) (Dahlstroem & Fuxe, 1964; Glimcher, 2011). Atrophy to dopaminergic neurons of the SNc leads to a reduction of dopamine within the striatum, a core feature

of the neurodegenerative disorder Parkinson's Disease (PD), which is marked by motor symptoms, such as tremors and bradykinesia (Damier, Hirsch, Agid, & Graybiel, 1999; Lotharius & Brundin, 2002). Other dopamine emitting neurons in humans are contained in the A8 and A10 groups of the ventral tegmental area (VTA), from which the mesolimbic pathway connects to the ventral striatum, including the nucleus accumbens (Swanson, 1982). Other efferent pathways project to widespread regions of the frontal cortex and the amygdala by virtue of the neurons' extensive axonal branches (Daw & Tobler, 2014; Glimcher, 2011; Matsuda et al., 2009). Crucially, dopamine neurons are simultaneously implicated in reward processing, and interact with motor control and action selection by affecting neural plasticity and neurotransmitter release in the basal ganglia (Daw & Tobler, 2014; Reynolds & Wickens, 2002).

The mechanics of dopamine release for action in the striatum are split into two distinct neural firing modes: tonic and phasic neural activity (Grace, 1991; Maia & Frank, 2011). Tonic dopamine discharge is characterized by a relatively slow and erratic firing rhythm (Dreyer, Herrik, Berg, & Hounsgaard, 2010; Grace & Bunney, 1984b). By releasing dopamine into the extracellular space, tonic dopamine responses determine the baseline reactivity of the midbrain dopamine system (Goto, Otani, & Grace, 2007; Grace, 1991). In contrast, phasic neural firing is marked by a brief but vigorous burst spike firing pattern, elevating intrasynaptic dopamine concentration within the striatum (Grace & Bunney, 1984a). Phasic dopamine neuron activity is triggered by salient stimuli, but its overall amplitude is dependent on the tonic dopamine level (Grace, 1991). Both neural firing modes underlie distinct yet occasionally overlapping purposes; tonic dopamine has been largely associated with motivation, whereas phasic activity is linked to learning stimulus-reward associations, and has become a crucial component in classical conditioning and reinforcement learning models (Cagniard, Balsam, Brunner, & Zhuang, 2006; Hamid et al., 2016; Salamone & Correa, 2012; Schultz et al., 1997; Sutton & Barto, 1998).

Behavioral Dopamine Signatures. Reward-driven behavior offers a window on the motivational effects of tonic dopamine. For example, Guitart-Masip, Beierholm,

Dolan, Duzel, and Dayan (2011) assessed reaction times in an oddball discrimination task, incentivizing detection of deviant stimuli on a trial-by-trial basis with varying reward magnitudes. Immediate rewards were not linked to response vigor, but reaction times were correlated negatively with an average reward rate signal (Guitart-Masip et al., 2011). Tonically active dopamine neurons are hypothesized to carry these average reward rate signals to the nucleus accumbens within the ventral striatum, where dopamine signals opportunity costs for action (Niv et al., 2007). Since subjective reward value is degraded by temporal discounting, higher average reward evokes greater potential costs, thereby galvanizing an organism into responding more vigorously (Niv et al., 2007; Shadmehr, de Xivry, Xu-Wilson, & Shih, 2010). Beierholm et al. (2013) corroborated the involvement of tonic dopamine in response speed by combining the oddball paradigm with a pharmacological intervention. Healthy participants either received citalopram, which blocks the reuptake of serotonin, L-Dopa , which elevates dopamine levels, or a placebo (Beierholm et al., 2013; Schultz, 1998). The researchers found that response speed was predicted by average reward rate, and that this link was more pronounced under L-Dopa (Beierholm et al., 2013).

Classical conditioning illuminates the functions of phasic dopamine activity. Using optogenetic tools in mice, Tsai et al. (2009) compared the reinforcing properties of tonic and phasic neural firing in the conditioned place paradigm. Light pulses were delivered to VTA dopamine neurons at 1 Hz or 50 Hz to induce tonic or phasic firing, respectively, pairing each mode of delivery with a certain chamber. After conditioning, mice were observed to spend more time in the chamber that was associated with 50 Hz stimulation, demonstrating that phasic dopamine activity provides a potent reinforcing signal for associative learning (Tsai et al., 2009). Moreover, a crucial observation is that associative learning depends on the uncertainty of the predictors, a notion signified by the 'blocking' effect (Daw & Tobler, 2014; Kamin, 1969). If a stimulus has come to entirely predict the delivery of a reward, learning will be impaired for any additional stimuli paired with the reward-predictive stimulus (Schultz et al., 1997; Tobler, O'Doherty, Dolan, & Schultz, 2006). These findings suggest that reward-driven learning

is dependent on errors (Daw & Tobler, 2014; Tobler et al., 2006).

The Reward Prediction Error Signal. Electrophysiological recordings indicate that errors are carried by phasic dopamine activity (Schultz et al., 1997). For instance, Hollerman and Schultz (1998) used microelectrodes to record dopamine neurons in monkey SNc and VTA during a picture discrimination task. During the initial stages of learning, administration of a juice reward triggered an excitatory phasic dopamine signal, which progressively waned as the monkeys learned to associate the picture to the reward. In addition, omission of an expected reward drove dopamine activity below baseline, suggesting that midbrain dopamine neurons code for the temporal and absolute divergence between expectation and reward (Hollerman & Schultz, 1998; Schultz et al., 1997). Moreover, in a study by Pessiglione, Seymour, Flandin, Dolan, and Frith (2006), midbrain dopamine neurons were recorded via fMRI during an instrumental learning task under haloperidol or L-Dopa treatment. Prediction errors were associated with increased activation in the ventral striatum and posterior putamen, modulated by overall dopamine availability, and predicted the rate of learning in a computational model (Pessiglione et al., 2006). Thus, the reward prediction error (RPE) has become established as a behavioral 'teaching' signal (Pessiglione et al., 2006; Schultz et al., 1997).

These discoveries about the midbrain dopamine system coincided with previous advances in classical conditioning and reinforcement learning, a computer science discipline concerned with error-dependent learning (Daw & Tobler, 2014; Montague, Dayan, & Sejnowski, 1996; Sutton & Barto, 1998). Specifically, the magnitude of the RPE signal at a given trial (δ_k) was observed to correspond closely to the predictions of the Rescorla – Wagner (1972) learning rule (Daw & Tobler, 2014; Schultz et al., 1997):

$$\delta_k = R_k - V_k(s_k) \tag{1}$$

Here, R_k represents a received reward, and $V_k(s_k)$ expresses a reward prediction associated with a predictive stimulus (Daw & Tobler, 2014; Schultz, 2007a). Moreover, predictions about reward value on upcoming trials (V_{k+1}) are updated by multiplying the RPE with the learning rate (α), and adding the product to the current reward value

estimate (Daw & Tobler, 2014):

$$V_{k+1}(s_k) = V_k(s_k) + \alpha \cdot \delta_k \tag{2}$$

In this iterative equation, the free parameter α determines the manner in which current value estimates are swayed by the reward history. If α approaches one, only the most recent rewards take substantial effect, whereas α values close to zero consider a more extended range of previous rewards (Bayer & Glimcher, 2005). Thus, this algorithm computes a weighted average comprised of all previous rewards received, whereby reward weights follow a pattern of exponential decay (Bayer & Glimcher, 2005; Daw & Tobler, 2014). Advances in reinforcement learning have further shown that midbrain dopamine neuron activity can be captured by considering equations 1 and 2 at specific points in time instead of trials using temporal difference learning algorithms, which helped to account for phenomena such as blocking or secondary conditioning (Schultz et al., 1997; Sutton & Barto, 1998).

Reward Uncertainty Coding. In addition, midbrain dopamine neurons are implicated in propagating reward probability (Schultz, 2007b). In a study by Fiorillo, Tobler, and Schultz (2003), monkeys were trained to associate visual stimuli with varying probabilities of future reward while midbrain dopamine neuron activity was recorded via microelectrodes. A large phasic RPE response followed the delivery of reward when the visual stimulus indicated a reward probability of zero. In line with previous experimental findings, the RPE at the point of reward administration gradually decreased as the reward probability of the cue approached one (Hollerman & Schultz, 1998). Moreover, the researchers observed a novel sustained dopamine response that was maximal for the highest amount of reward uncertainty ($P = .5$), and increased from the point of cue onset to the moment of reward delivery. The amplitude of this uncertainty signal was similar for intermediate reward probabilities ($P = .25$ versus $P = .75$). The authors reasoned that this dopaminergic uncertainty signal is involved in learning and risk-taking behavior, though its behavioral functions are still subject to debate (Fiorillo et al., 2003; Schultz, 2007b). The uncertainty signal also shares commonalities with phasic dopamine responses, and recent theories on dopamine in

decision-making have implicated RPEs in transmitting uncertainty as well (Fiorillo et al., 2003; Friston et al., 2012).

The Computational Basis of Sensorimotor Learning

Uncertainty is a critical component in sensing, perceiving, and acting on the world (Berniker & Kording, 2010; Vilares, Howard, Fernandes, Gottfried, & Kording, 2012). Visual input, for example, is constrained by the environment (e.g., the number of available photons to be absorbed by the eyes), the anatomical layout of the nervous system (e.g., the distribution of photoreceptors across the retina), and neural noise at every stage of the sensory machinery (e.g., random fluctuations in firing rates of neurons in visual cortex) (Faisal, Selen, & Wolpert, 2008; Knill & Pouget, 2004; Tolhurst, Movshon, & Dean, 1983). The motor system similarly faces noise, for example, in motor commands and muscle cells, contributing to the inherent variability of movements (Trommershäuser, Maloney, & Landy, 2003; Wolpert & Ghahramani, 2009). How does the nervous system compute a reliable model of the external world in the presence of uncertainty? Recent studies on sensorimotor control have considered perception and motor behavior as probabilistic processes, and have applied Bayes' theorem to the study of the nervous system (Knill & Pouget, 2004; Kording & Wolpert, 2004).

Consider the basic problem of inferring the properties of a stimulus given some noisy information; this could refer to estimating the velocity of an approaching car, or the position of a light switch at night (Seriès & Seitz, 2013; Wolpert & Ghahramani, 2009). In Bayesian decision theory, these scenarios are expressed as hypotheses about a state of the world A given an observation B. The final estimate $P(A \mid B)$ is referred to as the *posterior* distribution, and is computed by combining current sensory information $P(B \mid A)$ with prior knowledge $P(A)$, and dividing by the normalizing constant $P(B)$ (Bays & Wolpert, 2007; Tenenbaum, Kemp, Griffiths, & Goodman, 2011; Vilares & Kording, 2011):

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)} \propto P(B \mid A) \cdot P(A) \tag{3}$$

Bayes' rule is interpreted as providing the degree to which the *prior* $P(A)$ should be

updated by the *likelihood* $P(B \mid A)$ (Vilares & Kording, 2011). This process depends on the uncertainty associated with each predictor; locating a light switch under low illumination might prompt more reliance on prior knowledge. Crucially, a set of recent studies have indicated significant overlap between behavioral observations and statistically optimal predictions derived from Bayesian decision theory (Knill & Pouget, 2004; Kording & Wolpert, 2004; Vilares et al., 2012).

Behavioral Correlates of Bayesian Inference. Visuomotor integration exemplifies Bayesian inference in the nervous system. For example, Kording and Wolpert (2004) asked their participants to execute a reaching movement toward a visual target while their hands were blocked from view. At the midway point, participants were given feedback about the finger position with a lateral shift that was drawn from a Gaussian distribution (likelihood). After repeated trials, participants learned the average lateral displacement of the finger position (prior). On subsequent trials, end positions of the reaching movements indicated that participants integrated their prior knowledge with the visual feedback (Kording & Wolpert, 2004). Moreover, in a rapid pointing task, Tassinari, Hudson, and Landy (2006) observed that motor behavior echoes the integration of prior and likelihood information as well. In addition, when the researchers increased the uncertainty of sensory information, participants' aim points more closely matched the learned prior distribution. Studies on time and motion perception further substantiate the notion that the nervous system generates predictions about the world that are in agreement with Bayesian statistics (Berniker, Voss, & Kording, 2010; Miyazaki, Nozaki, & Nakajima, 2005; Weiss, Simoncelli, & Adelson, 2002).

Multimodal cue integration provides an additional area of application for Bayes' theorem (Vilares & Kording, 2011). In the McGurk effect, for instance, hearing 'ba' while simultaneously viewing a silent lip recording of 'ga' leads people to perceiving the syllable 'da' (McGurk & MacDonald, 1976). Rather than fully relying on a single modality, the sensory system combines input from both modalities to provide a unified percept (Vilares & Kording, 2011). This pattern has also been observed in the

ventriloquist effect that is argued to originate from close-to-optimal integration of visual and auditory information (Alais & Burr, 2004). Moreover, in a study by Ernst and Banks (2002), participants were provided with visual and proprioceptive feedback about the position of a bar, and had to estimate its height. Contrary to the phenomenon of visual dominance, the authors observed that participants relied more on haptic information as the visual feedback of the bar was rendered increasingly uncertain. This suggests that the nervous system combines cues from multiple modalities by taking into account their uncertainties in a way that is close to the statistical optimum (Alais & Burr, 2004; Ernst & Banks, 2002; Kersten, Mamassian, & Yuille, 2004).

Bayesian Uncertainty Coding. How does the brain represent uncertainty in Bayesian integration? Beierholm, Quartz, and Shams (2009) employed a multisensory spatial localization task to demonstrate that prior uncertainty is encoded independently from current sensory information. Between two experimental sessions, the researchers altered the contrast of visual cues (informative versus uninformative likelihood), and observed that participants' prior representations remained unchanged (Beierholm et al., 2009). Furthermore, Vilares et al. (2012) employed fMRI during a Bayesian decision-making task in which prior and likelihood uncertainty were systematically modified; in different experimental blocks, participants learned the distribution of a target stimulus that was sampled from a wide or narrow Gaussian distribution, while receiving a trial-by-trial likelihood cue that was either high or low in uncertainty. Increased likelihood uncertainty was associated with activation in the visual cortex, whereas prior uncertainty encompassed distributed activation within the limbic system and the basal ganglia, including the amygdala and putamen. Significantly, the authors hypothesized that enhanced activity in the putamen may signal prior uncertainty, thereby initiating an orienting response toward current sensory information (Vilares et al., 2012).

Finally, Vilares and Kording (2017) recently implicated the midbrain dopamine system in Bayesian integration. The authors conducted a similar Bayesian decision-making task as Vilares et al. (2012) with PD patients on and off dopaminergic

medication. Patients on L-Dopa medication were observed to place more weight on current sensory input than prior knowledge. This difference in sensory weights was more pronounced for patients that had recently been diagnosed with PD, which is hypothesized to be related to the capacity of medication to restore dopamine functioning. PD patients were not impaired at learning prior distributions, but rathered differed in the extent to which they sought out sensory information under dopaminergic medication (Vilares & Kording, 2017). Furthermore, Ting, Yu, Maloney, and Wu (2015) assessed participants' integration of prior knowledge and current sensory information in the context of reward probabilities. The researchers found that the integration of prior and likelihood was partly associated with activity in the ventral striatum (Ting et al., 2015). Taken together, this suggests that the midbrain dopamine system might be directly involved in sensorimotor control (Ting et al., 2015; Vilares & Kording, 2017).

The purpose of the present study was to investigate the implications of reward in Bayesian decision theory. Participants completed a Bayesian decision-making task that was adopted from previous studies by Vilares et al. (2012) and Vilares and Kording (2017), and received rewards depending on performance. Reward magnitude was manipulated, and RPEs were computed by taking into account the history of received rewards. First, participants were anticipated to integrate prior and likelihood information in agreement with Bayesian decision theory. In addition, reward magnitude was expected to be related to response speed; higher rewards on average should lead to enhanced response vigor but not accuracy or differences in sensorimotor integration (Niv et al., 2007). Moreover, given the role of prediction errors in learning, it was expected that RPEs mediate the balancing of prior and likelihood integration on a trial-by-trial basis; if the integration of prior and likelihood is affected by dopamine, positive RPEs should lead to higher reliance on the likelihood by virtue of phasic dopamine release. This effect was expected to be stronger when average reward was high since tonic dopamine has been observed to enable phasic firing (Grace, 1991; Pessiglione et al., 2006).

Methods

Participants

Twenty-one female and three male students from the University of Potsdam gave informed consent, and performed the experiment in exchange for monetary compensation or a combination of course credits and a small bonus payment. Based on the effect size for medication type ($\eta_p^2 = .04$) reported by Vilares and Kording (2017), an a priori power analysis via G*Power yielded a minimum sample size of eighteen to detect a similar effect with ninety-five percent power (Faul, Erdfelder, Lang, & Buchner, 2007; Lakens, 2013). The participants' ages ranged from eighteen to forty-one ($M = 26.96, SD = 6.33$), and none indicated impairments to their physical or psychological well-being. All participants had normal or corrected-to-normal vision, except for one participant with an unilateral open-angle glaucoma. After completing several practice trials, this participant did not report any perceptual difficulties related to the task, and was therefore cleared to complete the experiment. Twenty-two participants were right-handed and none indicated having consumed alcohol during the preceding twelve hours. Students that had previously participated in a sensorimotor learning seminar were excluded, so that all participants were naïve to the purpose of the experiment.

Materials

Participants were informed about the course of the experiment via an informed consent sheet that was adapted from the World Health Organization (2011). An additional demographic questionnaire asked the participants about their biological sex, handedness, state of mind, current substance use and vision. After the experiment, the participants were debriefed about the purpose of the experiment and given a form that listed the experimental hypotheses, as well as contact details and further readings. Participants also provided their bank details for payment, and were able to indicate an email address in case they wished to be informed about the results of the study.

Apparatus

The experiment was run on a Dell Precision Tower 3620, and implemented in GNU Octave using the Psychophysics Toolbox Version 3 (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli, 1997). Participants were seated 70 cm in front of a EIZO FORIS FS2735 computer screen (68.5 cm, 2560 x 1440 Pixels, 144 Hz). Responses were recorded via a Dell Optical Mouse MS116.

Task

The main task consisted of moving a vertical bar to the suspected position of a concealed target stimulus (see figure A1). During the experiment, the task was illustrated via a ball game scenario. Participants were asked to guess the landing position of an approaching yellow ball (target), thrown by a fictive player standing at a fixed location behind the computer screen. Five gray balls (cue cloud) preceded the target, and gave the participants an indication about the upcoming position of the yellow ball. After making a guess, the target was shortly displayed so that participants could learn about the player's throwing behavior. The link between the cue cloud and the target was intentionally kept uncertain in order to prevent participants from developing systematic decision-making biases. Between blocks, the participants were informed that the fictive player might be replaced by another player. Optimally estimating the position of the target required integrating information about the fictive players' throwing behavior and the dispersion of the cue cloud. Additionally, the accuracy of participants' estimates determined the amount of the trial-by-trial bonus payments.

Stimuli

On every trial, the horizontal target coordinate was drawn from a Gaussian distribution centered on the screen, varying in standard deviation depending on prior condition and experimental block ($\mu_P = .5, \sigma_P = .085 \cap \sigma_P = .025$). Vertical target locations were chosen from a uniform distribution covering the central twenty percent of

the screen ($U\left(.4, .6\right)$). The cue coordinates were drawn from Gaussian distributions centered on the target's horizontal and vertical locations. The horizontal dispersion of the cue cloud was varied within blocks according to one of two likelihood conditions ($\sigma_L = .15 \cap \sigma_L = .06$), whereas the vertical expanse was set to a fixed value ($\sigma_V = .05$). The parameter values of the Gaussian distributions were adapted from previous studies by Vilares et al. (2012), as well as Vilares and Kording (2017). During the training session, target and cue coordinates were sampled from uniform distributions to forestall potential carry-over effects ($U_X\left(.05, .95\right), U_Y\left(.4, .6\right)$). The target appeared in yellow, and the cues in dark gray, both covering one percent of the screen width in diameter. The vertical bar was set to dark gray and one pixel in width.

In addition, trial-by-trial bonus payments were determined via an interval that was centered on the horizontal target coordinate, covering eight percent of the screen width. Reward amount decreased bilaterally from the midpoint in a linear fashion. Depending on the experimental block, participants could earn a maximum of 2 or .5 cents per trial. A similar previous study that was conducted at the faculty gave an indication about participants' accuracy in such tasks ($M = .044, SD = 0.037$), and informed the definition of the linear equation.

Procedure

After completing the informed consent and demographic questionnaire forms, participants were seated individually in a dimly lit cabin and instructed to visualize the ball game scenario. After completing ten practice trials, the experimenter consulted with the participants to clarify any remaining questions. Participants were able to trigger each trial by moving the mouse cursor over a central fixation cross. After 250 ms, the fixation cross vanished and the cue cloud was continuously displayed on the screen. There was no time constraint on participants' responses, and the vertical bar always appeared at the middle of the screen at the start of each trial. Once participants confirmed their placement of the vertical bar, the target was shown for 250 ms. Subsequently, the amount of the bonus payment was presented for 1000 ms.

Bonus payments were cumulative and added to a fixed amount of seven euros or course credit. The total bonus payment earned was displayed on the screen after completion of the fourth block ($M = 4.225, SD = .326$). It took participants approximately sixty minutes to complete the experiment, while each response lasted 1.65 seconds on average ($SD = 1.31$). At the end of the experiment, all participants were debriefed.

Data Analysis

The purpose of the experiment was to analyze participants' responses in a sensorimotor learning task under reward within the framework of Bayesian inference. For each trial, optimal estimates for the integration of prior knowledge and sensory information were given by Bayes' rule (Kording & Wolpert, 2004; Vilares et al., 2012):

$$X_{opt} = \frac{\sigma_L^2}{\sigma_L^2 + \sigma_P^2}\mu_P + \frac{\sigma_P^2}{\sigma_L^2 + \sigma_P^2}\mu_L \tag{4}$$

Since both fictive players aimed at the center of the screen, the mean of the prior distribution (μ_P) was a constant at .5. Variation in the players' throwing behavior was linked to the standard deviation of the prior (σ_P), which was set to .085 or .025. The dispersion of the cue cloud referred to the variance of the likelihood (σ_L), which assumed one of two factor levels (.15 or .06). Finally, the average horizontal position of the cue cloud determined the mean of the likelihood distribution (μ_L) (Vilares et al., 2012).

Moreover, this study was concerned with analyzing participants' reliance on sensory information between reward conditions. Equation 4 was restated to indicate the optimal dependence on likelihood information (sensory weight) on each trial (Vilares & Kording, 2017):

$$sw = \frac{X_{opt} - \mu_P}{\mu_L - \mu_P} \tag{5}$$

Optimal sensory weights were compared to participants' behavior by substituting X_{opt} in equation 5 with participants' target estimates. A logistic transformation was applied to the obtained sensory weights ($1/(1 + e^{-SW_k})$) to limit their range to zero to one (Vilares & Kording, 2017). Sensory weights equal to one indicated exclusive reliance on

likelihood information while sensory weights of zero indicated complete reliance on prior knowledge (Vilares & Kording, 2017).

Furthermore, participants' prior means were inferred to augment the sensory weight analysis. To this end, target estimates were analyzed as a function of the centroid of the cue cloud (equation 4), and the associated intercept was computed by rearranging the formula to give $\beta_0 = \sigma_L^2/(\sigma_L^2 + \sigma_P^2)\mu_P$ (Vilares & Kording, 2017). Prior means were then inferred for each combination of prior and likelihood, assuming that participants behaved in accord with Bayesian decision theory (Vilares & Kording, 2017):

$$\hat{\mu}_P = \frac{\beta_0}{1 - sw} \tag{6}$$

Finally, trial-by-trial variations in sensory weight (equation 5) were investigated as a function of RPEs. On every trial, RPEs were calculated by subtracting the expected reward value from the actual reward received (see equation 1). Expected values were updated according to equation 2. For subsequent analyses, the reward prediction error at $k - 1$ was used to predict sensory weights in trial k. The reinforcement model was initiated separately with varying learning rate parameters (.2 versus .5 versus .8). At the start of each block, expected values were reset to zero.

Statistical Analysis and Design

The experiment comprised a 2 x 2 x 2 balanced repeated measures design with one time-varying covariate ($N = 600$). The independent variables were prior (narrow versus wide), likelihood (narrow versus wide), reward (high versus low), and RPE ($[-1, 1]$). Participants' horizontal estimate coordinates were transformed according to equation 5, and served as dependent variable. Reaction times and absolute target deviations were recorded for additional analyses. The experiment was divided into four blocks of 150 trials each, and included 10 practice trials. Within blocks, prior and reward were held constant so that participants were exposed to each level twice, yielding a total of twenty-four possible permutations. Levels of the likelihood factor were randomly interspersed and evenly distributed within each block. RPEs were inferred in retrospect

via equations 1 and 2. Moreover, responses faster than 100 ms or slower than 10 seconds were excluded from the analysis (Baayen & Milin, 2010; Whelan, 2008).

Statistical analyses were carried out in *R* using the packages *tidyverse*, *lme4*, *lmerTest*, *psycho*, *car*, and *optimx* (Bates, Mächler, Bolker, & Walker, 2015; Fox & Weisberg, 2011; Kuznetsova, Brockhoff, & Christensen, 2017; Makowski, 2018; Nash, 2014; Nash & Varadhan, 2011; R Core Team, 2017; Wickham, 2017). Linear mixed-effects models fitted by maximum likelihood were specified to analyze participants' reaction times, target deviations, and sensory weights. While these models comprised different main effects and interactions dependent on the experimental hypotheses, they shared a similar effects structure. Each independent variable was treated as a fixed effect, allowed to have a random slope, and a corresponding random intercept was assumed for each participant. Interactions were entered as fixed effects only. In addition, hypothesis tests were carried out with t-tests using Satterthwaite's method (Kuznetsova et al., 2017). Moreover, a cube root transformation was applied to the target deviations, and reaction times were adjusted via a logarithmic transformation. Subsequent plots of all dependent variables' distributions were visually inspected to confirm normality. An optimizer was employed if a model failed to converge initially.

Reaction times were analyzed with a single mixed-model, whereas target deviations and sensory weights were analyzed with separate mixed-models for RPEs that were computed with different learning rates (.2 versus .5 versus .8). In the exploratory analysis, target deviations and sensory weights were also tested with models that included the most extreme RPE values ($\text{RPE} > .5 \cap \text{RPE} < -.5$) and RPE as a two-level factor (positive versus negative), and an additional trial-based analysis inspected sensory weights during different intervals of the experiment. Additional one- or two-sample t-tests were employed to test participants' sensory weights against a Bayes-optimal and a senses-only model, and to examine differences between participants' inferred prior means and the imposed prior mean. For all graphical representations, confidence intervals were computed using the Cousineau-Morey method

for within-subject designs (Morey, 2008; Pohlitzer-Ahles, 2018). Finally, application of a family-wise error rate correction was rejected since hypotheses during the main analysis and post-hoc testing were clearly defined (Armstrong, 2014; Streiner & Norman, 2011).

Results

This study was designed to test the integration of prior knowledge and current information under reward in a sensorimotor estimation task. Two different prior and likelihood distributions were presented during four experimental blocks, and a monetary reward was administered on each trial based on participants' performance and two varying reward maxima. RPEs were computed in retrospect based on each participant's reward history. This experimental design allowed to investigate the effects of reward magnitude and prediction errors on the integration of prior and likelihood.

Main Analysis

Mean estimates for reaction times, target deviations, and sensory weights per experimental condition are displayed in table B1. Participants exhibited speeded responses when prior and likelihood distributions were both narrow instead of wide. When only one of the distributions was narrow, participants reacted quicker to likelihood information. Higher rewards elicited faster responses when the prior distribution was wide, but this trend reversed when the prior distribution was narrow. Furthermore, participants' accuracy increased from wide to narrow prior and likelihood distributions. Narrow likelihood distributions paired with a wide prior outperformed wide likelihood distributions together with a narrow prior. Finally, sensory weights were higher for narrow than for wide likelihoods conditional on prior distributions. Reward only had a marginal effect on target deviations and sensory weights.

Table C1 depicts the results of the linear mixed-model analysis for reaction time. There was a significant main effect of likelihood ($t\,(24.02) = 6.196, p < .001$), suggesting that participants reacted faster when presented with narrow likelihoods. The main effects of prior ($t\,(24) = 1.579, p = .128$) and reward ($t\,(24) = .545, p = .591$) were not significant, but a significant interaction between prior and reward ($t\,(14150) = -4.851$,

$p < .001$) was found. This implies that the effect of reward was different for the two prior conditions depending on reward; Participants reacted quicker to high rewards when prior uncertainty was high, but slower when prior uncertainty was low. In addition, there was an interaction between prior and likelihood reward ($t\,(14150) = 2.105, p = .035$), suggesting that the effect of narrow likelihoods was stronger under low prior uncertainty (see figure D1). These findings are in contrast to the hypothesis that reward enhances response vigor.

Next, participants' task accuracy was assessed with a linear mixed-model (see table C1). Results are reported for a learning rate of .5 since RPE effects were qualitatively similar for varying learning rates. The main effects of prior ($t\,(23.98) = -7.476, p < .001$) and likelihood ($t\,(24.03) = -19.108, p < .001$) were statistically significant. This finding indicates that participants were more accurate when prior and likelihood uncertainty decreased. There was also a significant interaction between prior and likelihood ($t\,(14160) = 4.838, p < .001$), and visual inspection suggested that there was less divergence in accuracy between prior conditions for narrow compared to wide likelihoods. The main effects of reward ($t\,(26.03) = 1.275, p = .214$) and RPE ($t\,(27.52) = .54, p = .593$) were not significant. Figure E1(A) depicts mean target deviations for each combination of prior and likelihood averaged over reward.

Furthermore, results for participants' sensory weights are shown in table C1. Results are reported for a learning rate of .5. There were significant main effects of prior ($t\,(24.14) = -7.873, p < .001$) and likelihood ($t\,(41.27) = 6.33, p < .001$); participants increased or decreased their sensory weights when presented with a narrow likelihood or prior distribution, respectively. The main effects of reward ($t\,(47.94) = -.889, p = .378$) and RPE ($t\,(40.94) = .211, p = .834$), and the interaction between reward and RPE ($t\,(14150) = -.795, p = .427$) did not reach statistical significance, indicating that sensory weights were not sensitive to changes in reward. These findings are not in line with the expectation that prediction errors modulate reliance on likelihood information, and that the effects of RPE are stronger under high reward. However, there was a significant interaction between prior and reward

($t\,(14170) = 2.451, p = .014$). Compared to low reward, sensory weights were elevated for narrow priors under high reward, and decreased for wide priors. Figure F1 depicts how participants adjusted their sensory weights when presented with varying degrees of uncertainty and different reward magnitudes.

Several one-sample t-tests were conducted to test whether participants' sensory weights were in agreement with Bayesian statistics. These tests were statistically significant for each combination of prior and likelihood (pl: $t\,(23) = 16.52, p < .001$, pL: $t\,(23) = 47.237, p < .001$, Pl: $t\,(23) = -38.574, p < .001$, PL: $t\,(23) = 3.934, p = .001$), indicating that participants' sensory weights were not Bayes-optimal. Moreover, when comparing sensory weights to a senses-only model, t-tests for each experimental condition were significant (pl: $t\,(23) = -38.385, p < .001$, pL: $t\,(23) = -38.866, p < .001$, Pl: $t\,(23) = -53.542, p < .001$, PL: $t\,(23) = -44.261, p < .001$). This finding suggests that participants' sensory weights were different from a hypothetical observer not taking into account prior knowledge. Differences in sensory weights between participants, the Bayesian optimum, and the senses-only model are captured in figure G1A.

In addition, inferred prior means are depicted in figure H1. The differences between participants' prior estimates and the actual prior mean were statistically significant for each combination of prior and likelihood (pl: $t\,(23) = -2.1, p = .047$; pL: $t\,(23) = 7.139, p < .001$; Pl: $t\,(23) = -205.22, p < .001$; PL: $t\,(23) = -13.523, p < .001$). These tests show a discrepancy between participants' prior estimates and the imposed mean across conditions. This pattern held true when the first twenty trials of each block were excluded to account for learning effects (Vilares & Kording, 2017).

Exploratory Analysis

To contrast positive and negative RPEs, additional linear mixed-models for target deviation and sensory weight were analyzed with RPE as a two-level factor. There was no main effect of RPE ($t\,(23.82) = .276, p = .785$) on target deviation. Plus, RPE ($t\,(65.84) = .07, p = .945$) failed to reach significance in the sensory weight model. No interactions involving RPE were statistically significant. This suggests that there were

no performance differences between positive and negative RPEs (see figures E1B and G1B). In addition, when only taking into account the most extreme RPE values to predict sensory weights, the main effect of RPE ($t\,(34.03) = -.04, p = .968$) as well as relevant interactions were not statistically significant. Furthermore, excluding trials on which reaction times were two standard deviations from the mean did not change the effects structure reported in the main analysis. There was, however, an additional interaction between likelihood and RPE ($t\,(13570) = -2.416, p = .016$) for target deviation, which indicated that positive RPEs boosted performance for narrow likelihoods, but were associated with decreasing accuracy for wide likelihoods.

Additional linear mixed-models were analyzed to investigate sensory weights every fifty trials (see figure I1). While the main effects of prior and likelihood were significant across intervals, the beta coefficients for prior gradually increased with an increasing number of trials (-4.26 versus -5.225 versus -8.338), and likelihood had its smallest effect in the last fifty trials (3.813 versus 4.484 versus 3.336). This signifies that the discrepancy between narrow and wide priors was most accentuated toward the end of the experiment. Moreover, the interaction between prior and reward became significant only toward the end of the experiment ($t\,(4723) = 3.46, p < .001$). The interaction indicates that the effect of reward was different for the levels of prior; sensory weights were lower for low rewards when the prior was narrow, but higher for low rewards when the prior was uncertain. Note that the sensory weights did not match the statistically optimal Bayesian model across intervals.

Discussion

The objective of this study was to determine how reward affects sensorimotor control within the framework of Bayesian decision theory. Participants were expected to integrate prior and likelihood distributions according to Bayes' rule. Indeed, both sources of uncertainty were reflected in participants' estimates; when the likelihood distribution was informative, participants placed less weight on prior knowledge and vice versa. This trade-off was suboptimal but in qualitative agreement with Bayesian

predictions. Moreover, reward magnitude was hypothesized to influence response speed. The present results did not indicate that response vigor is enhanced by higher average reward, nor did reward magnitude consistently affect the weighting of prior and likelihood information. Furthermore, it was anticipated that positive RPEs increase reliance on current sensory information. However, participants' sensory weights varied independently from RPE fluctuations. An exploratory analysis indicated that this finding held true throughout the experiment, and that participants' reliance on sensory information remained unchanged when considered as a function of RPE sign and the most extreme RPE values.

Sensorimotor Control Reflected Suboptimal Bayesian Inference

The finding that prior and likelihood information were integrated in qualitative but not quantitative agreement with Bayesian statistics corresponds to previous results by Vilares and Kording (2017). Specifically, the present study equally observed that participants overemphasized current sensory information across conditions (Vilares & Kording, 2017). The tendency to neglect prior knowledge in favor of case-relevant information has previously been demonstrated in high-level cognitive tasks, and has spurred debate on the conditions under which humans are rational Bayesian decision-makers (Bar-Hillel, 1980; Gigerenzer & Hoffrage, 1995; Ting et al., 2015; Tversky & Kahneman, 1974). The results of the present study thus suggest that sensorimotor control follows Bayesian inference, but might be subject to additional heuristics such as the base-rate fallacy (Tversky & Kahneman, 1974). This conclusion stands in contrast to another study by Vilares et al. (2012), in which near-optimal integration of prior and likelihood was reported for a similar sensorimotor learning task; participants' sensory weights in the Vilares et al. (2012) study were indistinguishable from Bayesian predictions when overall uncertainty was either high or low. In addition, the present findings oppose a cluster of research indicating statistically optimal integration in diverse areas such as motor control (Kording & Wolpert, 2004), reward processing (Ting et al., 2015), and multimodal cue integration (Alais & Burr, 2004;

Ernst & Banks, 2002).

This mismatch in prior and likelihood integration might be explained by differences in the task design. In contrast to Vilares et al. (2012) and Vilares and Kording (2017), participants in the present study received no information about the prior mean and no semantic link between likelihood and target. Accordingly, inferred prior means exhibited deviations from the imposed mean across conditions (see figure H1). This is an unexpected finding since the mean of the prior was demonstrated to be readily inferred within the first ten trials (Berniker et al., 2010). Notably, participants in the Vilares and Kording (2017) study displayed a larger range of sensory weights between conditions, indicating that prior and likelihood uncertainty were assessed more effectively. In addition, participants in the present study weighted prior knowledge more strongly over time, but this effect was restricted to minimal changes in sensory weight compared to Vilares and Kording (2017) (see figure I1). Together, these findings suggest that participants' suboptimality in sensorimotor integration was related to an impairment in inferring the relative prior and likelihood uncertainties.

Difficulties in sensorimotor learning might be attributable to the administration of reward. In economic decision-making, aversion to ambiguity has been found to represent a deviation from rational choice (Camerer & Weber, 1992; Inukai & Takahashi, 2009; Osmont, Cassotti, Agogué, Houdé, & Moutier, 2015). For instance, the Ellsberg paradox indicates that people tend to prefer choices with certain probabilities of gaining reward over choices with uncertain outcomes (Ellsberg, 1961; Glimcher & Rustichini, 2004). Hsu, Bhatt, Adolphs, Tranel, and Camerer (2005) investigated this tendency using fMRI and found that choice uncertainty paralleled activation in the limbic system and prefrontal cortex while expected value correlated positively with striatal activity. The researchers observed a lagged decline in striatum activity following increased amygdala activation under high uncertainty, suggesting that uncertainty modulates choice selection by signalling lower associated reward (Hsu et al., 2005). In the present study, likelihood uncertainty was readily available while prior uncertainty was maximal at beginning. In this context, choosing to rely on likelihood

information might have entailed less ambiguity about expected reward.

Bayesian Integration Proved Robust to Changes in Reward

High reward did not enhance response vigor, contradicting observations that reward is subject to temporal discounting (Shadmehr et al., 2010), and that the average reward rate signals opportunity costs for action (Beierholm et al., 2013; Guitart-Masip et al., 2011; Niv et al., 2007). Instead, response speed was driven by likelihood and prior uncertainty. This finding corresponds to a study by Hansen, Hillenbrand, and Ungerleider (2012), who observed speeded reaction times with informative priors, and argued that prior knowledge boosts sensory processing in visual cortex under perceptual uncertainty (Hansen et al., 2012). In addition, the present results are also consistent with the notion that decreased uncertainty in sensory information accelerates action preparation (Hyman, 1953). In this respect, Bestmann et al. (2008) showed that corticospinal excitability was maximal when the uncertainty of a movement instructive stimulus was low. Moreover, in the present study, reward was implicated in maladaptive sensorimotor learning since high reward slowed responses under decreasing uncertainty, which was accompanied by less Bayes-optimal integration. Previous research on reward processing has gathered support for the opposite pattern; Manohar et al. (2015), for example, demonstrated that saccades are executed more swiftly and accurately under high reward by suppressing neural noise. In addition, Chong et al. (2015) observed that PD patients on medication showed more inclination to exert effort in a grip force task than patients off medication.

The difference between reward levels might not have been large enough to elicit motivational effects. In the studies by Beierholm et al. (2013) and Guitart-Masip et al. (2011), participants were able to win up to a hundred pence per trial. Likewise, Manohar et al. (2015) rewarded participants for accurate performance with either ten or fifty pence. Therefore, it is possible that motivation and learning in the current study have deteriorated due to low reward (Hamid et al., 2016). Previous research has, for example, observed that low reward leads participants to exert less effort than when

receiving no reward (Gneezy & Rustichini, 2000; Heyman & Ariely, 2004). Notably, experimental sessions in the Beierholm et al. (2013), Guitart-Masip et al. (2011), and Chong et al. (2015) studies were also shorter in duration compared to the present study. This suggests that participants' task performance might have been additionally impacted by fatigue (Beierholm et al., 2013).

Furthermore, the observation that neither positive RPEs nor high reward enhanced reliance on current sensory information disagrees with the results by Vilares and Kording (2017) and Vilares et al. (2012). Specifically, these findings suggest that Bayesian sensorimotor control is mostly insensitive to reward alterations, and might therefore be largely independent of midbrain dopamine concentration. In this respect, Ting et al. (2015) also found that the degree to which decisions are based on likelihood information is partly regulated by brain areas outside the basal ganglia. The researchers proposed that the medial prefrontal cortex might be a candidate region for the representation and integration of prior knowledge and sensory information. In another study by d'Acremont, Schultz, and Bossaerts (2013), likelihood information correlated with activity in the medial prefrontal cortex as well, while prior representation and integration of both sources of information were ascribed to the inferior frontal gyrus. Moreover, the observation that positive RPEs enhance task performance solely under low sensory uncertainty opposes a large body of work that demonstrated the involvement of RPEs in associative learning by providing a behavioral teaching signal (Hollerman & Schultz, 1998; Lak, Nomoto, Keramati, Sakagami, & Kepecs, 2017; Pessiglione et al., 2006; Schultz et al., 1997).

These discrepancies might be due to the temporal characteristics of the RPE signal. Specifically, RPE responses initially ensue at reward delivery, but are transferred to the reward-predictive stimulus after repeated pairing and decreased uncertainty (Glimcher, 2011; Schultz, 2002). For instance, after monkeys have learned which lever triggers a juice reward, the RPE signal is observed at the time of the lever press and not at reward delivery (Schultz et al., 1997). Therefore, the RPE signal might have shifted to the presentation of sensory information in the current study, especially when

likelihood uncertainty was minimal. A recent study by Lak et al. (2017) examined the RPE response in perceptual decision-making under uncertainty, and found that prediction errors at cue presentation predicted choice accuracy in a computational model. The authors also accounted for cue uncertainty, suggesting that RPEs may carry a weight of sensory uncertainty that is related to the expected value, similar to the mechanism outlined by Hsu et al. (2005).

Methodological Limitations

The Rescorla – Wagner (1972) learning rule and the update algorithm that was employed in the present study do not account for these temporal changes (Daw & Tobler, 2014; Sutton & Barto, 1998). Plus, the potential implications of the blocking phenomenon remain uncertain by analyzing RPEs at the trial level (Sutton & Barto, 1998). For instance, Tobler et al. (2006) presented their participants with varying stimulus pairs of which one stimulus fully predicted the occurrence of reward. When reward outcome matched the cue's prediction, there was no RPE and no further learning about the second stimulus (Tobler et al., 2006). In the present study, there was an accuracy advantage for likelihood information, which implies that prediction errors were lowest under decreasing likelihood uncertainty (Daw & Tobler, 2014). In this context, RPEs might have blocked learning about the prior distribution. To counteract these limitations, future studies could employ temporal difference learning algorithms (Schultz et al., 1997).

In addition, the majority of RPEs exhibited minimal deviations from the predicted values under the current model (see figure I1). This aspect might have impacted sensorimotor integration since dopamine levels vacillate strongest under large prediction errors (Schultz et al., 1997). Even though this limitation was accounted for in the exploratory analysis, future studies might include conditions in which expected reward is suddenly omitted or multiplied to induce more extreme RPEs (Hollerman & Schultz, 1998). Moreover, the methodological design cannot exclude the possibility that participants failed to generate expected values and corresponding prediction errors.

Follow-up studies should aim to extract reliable RPE estimates by incorporating electrophysiological evidence, such as fMRI recordings (Fiorillo et al., 2003; Schultz et al., 1997).

Furthermore, the trial-by-trial rewards might have been too small to observe motivational effects of reward magnitude on task performance. This limitation could be resolved by adopting a lottery-based payout system; Beierholm et al. (2013) and Guitart-Masip et al. (2011), for example, paid participants the earnings of a randomly selected ten percent of all experimental trials. In addition, in the study by Chong et al. (2015), a point-based system was employed instead of administering reward directly. Moreover, this study compared discrete reward levels whereas previous studies modeled the average reward rate signal directly (Beierholm et al., 2013; Guitart-Masip et al., 2011). Although participants performed equally well between reward conditions, and therefore earned more reward in the high reward condition on average, this discrepancy might have compromised the scope of the present findings. A possibility for future research is to further isolate the contribution of reward magnitude, and vary potential trial-based rewards within blocks to compare fluctuating reward rates (Beierholm et al., 2013; Manohar et al., 2015).

Conclusion

Midbrain dopamine neurons assume key roles in reward processing and motor control (Daw & Tobler, 2014; Glimcher, 2011; Schultz et al., 1997). Within the framework of Bayesian sensorimotor learning, dopamine has recently been linked to alterations in the integration of prior knowledge and current sensory information (Vilares & Kording, 2017). The present study investigated changes in sensorimotor control by varying reward magnitude and inferring RPEs, thereby assessing the potential contributions of tonic and phasic dopamine, respectively (Niv et al., 2007; Schultz et al., 1997). Participants' behavior was in qualitative agreement with Bayesian statistics, but deficits in learning prior distributions and in combining prior and likelihood information were present across conditions and irrespective of reward. This

result suggests that the sensorimotor system exhibits systematic biases in estimating prior and likelihood uncertainties, and may not be considered statistically optimal at the behavioral level (Griffiths, Chater, Norris, & Pouget, 2012; Rahnev & Denison, 2018). Moreover, response vigor was not modulated by reward magnitude, and changes in RPEs did not correlate with fluctuations in sensory weight. These findings indicate that Bayesian integration might be largely independent of reward variables that have previously been related to changes in midbrain dopamine concentration (Beierholm et al., 2013; Pessiglione et al., 2006). Additional electrophysiological evidence is needed to investigate whether the process of integrating prior and likelihood information is represented in brain areas outside the reward-related pathways of the basal ganglia.

Acknowledgements

I would like to thank my thesis advisors, Professor Ralf Engbert and Dr. André Krügel, for valuable encouragement and supervision. I also wish to express my gratitude to Dr. Sarah Risse and Dr. Daniel J. Schad for insightful discussions on linear mixed-models. Lastly, I thank Petra Schienmann for assistance in recruiting participants.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262. https://doi.org/10.1016/j.cub.2004.01.029.

Alexander, G. E., & Crutcher, M. D. (1990). Preparation for movement: Neural representations of intended direction in three motor areas of the monkey. *Journal of Neurophysiology*, *64*(1), 133–150. https://doi.org/10.1152/jn.1990.64.1.133.

Alexander, G. E., DeLong, M. R., & Strick, P. L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual Review of Neuroscience*, *9*(1), 357–381. https://doi.org/10.1146/annurev.ne.09.030186.002041.

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics*, *34*(5), 502–508. https://doi.org/10.1111/opo.12131.

Baayen, R. H., & Milin, P. (2010). Analyzing reaction times. *International Journal of Psychological Research*, *3*(2), 12–28. http://dx.doi.org/10.21500/20112084.807.

Bar-Hillel, M. (1980). The base-rate fallacy in probability judgments. *Acta Psychologica*, *44*(3), 211–233. https://doi.org/10.1016/0001-6918(80)90046-3.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Bayer, H. M., & Glimcher, P. W. (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron*, *47*(1), 129–141. https://doi.org/10.1016/j.neuron.2005.05.020.

Bays, P. M., & Wolpert, D. M. (2007). Computational principles of sensorimotor control that minimize uncertainty and variability. *The Journal of Physiology*, *578*(2), 387–396. https://doi.org/10.1113/jphysiol.2006.120121.

Beierholm, U. R., Guitart-Masip, M., Economides, M., Chowdhury, R., Düzel, E., Dolan, R., & Dayan, P. (2013). Dopamine modulates reward-related vigor.

Neuropsychopharmacology, *38*(8), 1495–1503. https://doi.org/10.1038/npp.2013.48.

Beierholm, U. R., Quartz, S. R., & Shams, L. (2009). Bayesian priors are encoded independently from likelihoods in human multisensory perception. *Journal of Vision*, *9*(5), 23–23. https://doi.org/10.1167/9.5.23.

Berniker, M., & Kording, K. P. (2010). Bayesian approaches to sensory integration for motor control. *Wiley Interdisciplinary Reviews: Cognitive Science*, *2*(4), 419–428. https://doi.org/10.1002/wcs.125.

Berniker, M., Voss, M., & Kording, K. P. (2010). Learning priors for Bayesian computations in the nervous system. *PLOS ONE*, *5*(9). https://doi.org/10.1371/journal.pone.0012686.

Bestmann, S., Harrison, L. M., Blankenburg, F., Mars, R. B., Haggard, P., Friston, K. J., & Rothwell, J. C. (2008). Influence of uncertainty and surprise on human corticospinal excitability during preparation for action. *Current Biology*, *18*(10), 775–780. https://doi.org/10.1016/j.cub.2008.04.051.

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433–436. https://doi.org/10.1163/156856897X00357.

Cagniard, B., Balsam, P. D., Brunner, D., & Zhuang, X. (2006). Mice with chronically elevated dopamine exhibit enhanced motivation, but not learning, for a food reward. *Neuropsychopharmacology*, *31*(7), 1362–1370. http://dx.doi.org/10.1038/sj.npp.1300966.

Camerer, C., & Weber, M. (1992). Recent developments in modeling preferences: Uncertainty and ambiguity. *Journal of Risk and Uncertainty*, *5*(4), 325–370. https://doi.org/10.1007/BF00122575.

Carlsson, A., Lindqvist, M., Magnusson, T., & Waldeck, B. (1958). On the presence of 3-hydroxytyramine in brain. *Science*, *127*, 471. https://doi.org/10.1126/science.127.3296.471.

Chong, T. T.-J., Bonnelle, V., Manohar, S., Veromann, K.-R., Muhammed, K., Tofaris, G. K., ... Husain, M. (2015). Dopamine enhances willingness to exert

effort for reward in Parkinson's disease. *Cortex*, *69*, 40–46. https://doi.org/10.1016/j.cortex.2015.04.003.

d'Acremont, M., Schultz, W., & Bossaerts, P. (2013). The human brain encodes event frequencies while forming subjective beliefs. *Journal of Neuroscience*, *33*(26), 10887–10897. https://doi.org/10.1523/JNEUROSCI.5829-12.2013.

Dahlstroem, A., & Fuxe, K. (1964). Evidence for the existence of monoamine-containing neurons in the central nervous system. I. Demonstration of monoamines in the cell bodies of brain stem neurons. *Acta Physiologica Scandinavica*, *62*, 1–55.

Damier, P., Hirsch, E., Agid, Y., & Graybiel, A. (1999). The substantia nigra of the human brain: II. Patterns of loss of dopamine-containing neurons in Parkinson's disease. *Brain*, *122*(8), 1437–1448. https://doi.org/10.1093/brain/122.8.1437.

Daw, N. D., & Tobler, P. N. (2014). Value learning through reinforcement: The basics of dopamine and reinforcement learning. In P. W. Glimcher & E. Fehr (Eds.), *Neuroeconomics* (2nd ed., pp. 283–298). Cambridge, MA: Academic Press.

Depue, R. A., & Collins, P. F. (1999). Neurobiology of the structure of personality: Dopamine, facilitation of incentive motivation, and extraversion. *Behavioral and Brain Sciences*, *22*(3), 491–517. http://dx.doi.org/10.1017/S0140525X99002046.

Dreyer, J. K., Herrik, K. F., Berg, R. W., & Hounsgaard, J. D. (2010). Influence of phasic and tonic dopamine release on receptor activation. *Journal of Neuroscience*, *30*(42), 14273–14283. https://doi.org/10.1523/JNEUROSCI.1894-10.2010.

Ellsberg, D. (1961). Risk, ambiguity, and the Savage axioms. *The Quarterly Journal of Economics*, 643–669. https://doi.org/10.2307/1884324.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a.

Faisal, A. A., Selen, L. P. J., & Wolpert, D. M. (2008). Noise in the nervous system. *Nature Reviews Neuroscience*, *9*(4), 292–303. http://dx.doi.org/10.1038/nrn2258.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical

sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146.

Fiorillo, C. D., Tobler, P. N., & Schultz, W. (2003). Discrete coding of reward probability and uncertainty by dopamine neurons. *Science*, *299*(5614), 1898–1902. https://doi.org/10.1126/science.1077349.

Fox, J., & Weisberg, S. (2011). *An R companion to applied regression* (2nd ed.). Thousand Oaks, CA: Sage.

Friston, K. J., Shiner, T., FitzGerald, T., Galea, J. M., Adams, R., Brown, H., ... Bestmann, S. (2012). Dopamine, affordance and active inference. *PLOS Computational Biology*, *8*(1). https://doi.org/10.1371/journal.pcbi.1002327.

Gigerenzer, G., & Hoffrage, U. (1995). How to improve Bayesian reasoning without instruction: Frequency formats. *Psychological Review*, *102*(4), 684. https://doi.org/10.1037/0033-295X.102.4.684.

Glimcher, P. W. (2011). Understanding dopamine and reinforcement learning: The dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences*, *108*(42), 15647–15654. https://doi.org/10.1073/pnas.1014269108.

Glimcher, P. W., & Rustichini, A. (2004). Neuroeconomics: The consilience of brain and decision. *Science*, *306*(5695), 447–452. https://doi.org/10.1126/science.1102566.

Gneezy, U., & Rustichini, A. (2000). Pay enough or don't pay at all. *The Quarterly Journal of Economics*, *115*(3), 791–810. https://doi.org/10.1162/003355300554917.

Goto, Y., Otani, S., & Grace, A. A. (2007). The yin and yang of dopamine release: A new perspective. *Neuropharmacology*, *53*(5), 583–587. https://doi.org/10.1016/j.neuropharm.2007.07.007.

Grace, A. A. (1991). Phasic versus tonic dopamine release and the modulation of dopamine system responsivity: A hypothesis for the etiology of schizophrenia. *Neuroscience*, *41*(1), 1–24. https://doi.org/10.1016/0306-4522(91)90196-U.

Grace, A. A., & Bunney, B. S. (1984a). The control of firing pattern in nigral dopamine neurons: Burst firing. *Journal of Neuroscience*, *4*(11), 2877–2890. https://doi.org/10.1523/JNEUROSCI.04-11-02877.1984.

Grace, A. A., & Bunney, B. S. (1984b). The control of firing pattern in nigral dopamine neurons: Single spike firing. *Journal of Neuroscience*, *4*(11), 2866–2876. https://doi.org/10.1523/JNEUROSCI.04-11-02866.1984.

Graybiel, A. M. (1990). Neurotransmitters and neuromodulators in the basal ganglia. *Trends in Neurosciences*, *13*(7), 244–254. https://doi.org/10.1016/0166-2236(90)90104-I.

Griffiths, T. L., Chater, N., Norris, D., & Pouget, A. (2012). How the Bayesians got their beliefs (and what those beliefs actually are): Comment on Bowers and Davis (2012). *Psychological Bulletin*, *138*(3), 415–422. http://dx.doi.org/10.1037/a0026884.

Guitart-Masip, M., Beierholm, U. R., Dolan, R., Duzel, E., & Dayan, P. (2011). Vigor in the face of fluctuating rates of reward: An experimental examination. *Journal of Cognitive Neuroscience*, *23*(12), 3933–3938. https://doi.org/10.1162/jocn_a_00090.

Hamid, A. A., Pettibone, J. R., Mabrouk, O. S., Hetrick, V. L., Schmidt, R., Vander Weele, C. M., ... Berke, J. D. (2016). Mesolimbic dopamine signals the value of work. *Nature Neuroscience*, *19*(1), 117–126. https://doi.org/10.1038/nn.4173.

Hansen, K., Hillenbrand, S., & Ungerleider, L. (2012). Effects of prior knowledge on decisions made under perceptual vs. categorical uncertainty. *Frontiers in Neuroscience*, *6*, 163. https://doi.org/10.3389/fnins.2012.00163.

Heil, M. (2015). Manipulators live better, but are they always parasites? *Trends in Plant Science*, *20*(9), 538–540. https://doi.org/10.1016/j.tplants.2015.08.001.

Heyman, J., & Ariely, D. (2004). Effort for payment: A tale of two markets. *Psychological Science*, *15*(11), 787–793. https://doi.org/10.1111/j.0956-7976.2004.00757.x.

Hojo, M. K., Pierce, N. E., & Tsuji, K. (2015). Lycaenid caterpillar secretions manipulate attendant ant behavior. *Current Biology*, *25*(17), 2260–2264. https://doi.org/10.1016/j.cub.2015.07.016.

Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience*, *1*(4), 304–309. https://doi.org/10.1038/1124.

Hsu, M., Bhatt, M., Adolphs, R., Tranel, D., & Camerer, C. F. (2005). Neural systems responding to degrees of uncertainty in human decision-making. *Science*, *310*(5754), 1680–1683. https://doi.org/10.1126/science.1115327.

Hyman, R. (1953). Stimulus information as a determinant of reaction time. *Journal of Experimental Psychology*, *45*(3), 188–196. http://dx.doi.org/10.1037/h0056940.

Inukai, K., & Takahashi, T. (2009). Decision under ambiguity: Effects of sign and magnitude. *International Journal of Neuroscience*, *119*(8), 1170–1178. https://doi.org/10.1080/00207450802174472.

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Campbell & R. M. Church (Eds.), *Punishment and Aversive Behavior* (pp. 279–296). New York, NY: Appleton-Century-Crofts.

Kersten, D., Mamassian, P., & Yuille, A. (2004). Object perception as Bayesian inference. *Annual Review of Psychology*, *55*, 271–304. https://doi.org/10.1146/annurev.psych.55.090902.142005.

Kleiner, M., Brainard, D., & Pelli, D. (2007). *What's new in Psychtoolbox-3.* Tutorial session presented at the 30th European Conference on Visual Perception, Arezzo, Italy.

Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719. https://doi.org/10.1016/j.tins.2004.10.007.

Kording, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*(6971), 244. https://doi.org/10.1038/nature02169.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Lak, A., Nomoto, K., Keramati, M., Sakagami, M., & Kepecs, A. (2017). Midbrain dopamine neurons signal belief in choice accuracy during a perceptual decision. *Current Biology*, *27*(6), 821–832. https://doi.org/10.1016/j.cub.2017.02.026.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, *4*, 863. https://doi.org/10.3389/fpsyg.2013.00863.

Lotharius, J., & Brundin, P. (2002). Pathogenesis of Parkinson's disease: Dopamine, vesicles and α-synuclein. *Nature Reviews Neuroscience*, *3*(12), 932. https://doi.org/10.1038/nrn983.

Maia, T. V., & Frank, M. J. (2011). From reinforcement learning models to psychiatric and neurological disorders. *Nature Neuroscience*, *14*(2), 154–162. https://doi.org/10.1038/nn.2723.

Makowski, D. (2018). The psycho package: An efficient and publishing-oriented workflow for psychological science. *Journal of Open Source Software*, *3*(22), 470. https://doi.org/10.21105/joss.00470.

Manohar, S. G., Chong, T. T.-J., Apps, M. A., Batla, A., Stamelou, M., Jarman, P. R., ... Husain, M. (2015). Reward pays the cost of noise reduction in motor and cognitive control. *Current Biology*, *25*(13), 1707–1716. https://doi.org/10.1016/j.cub.2015.05.038.

Matsuda, W., Furuta, T., Nakamura, K. C., Hioki, H., Fujiyama, F., Arai, R., & Kaneko, T. (2009). Single nigrostriatal dopaminergic neurons form widely spread and highly dense axonal arborizations in the neostriatum. *Journal of Neuroscience*, *29*(2), 444–453. https://doi.org/10.1523/JNEUROSCI.4029-08.2009.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746–748. https://doi.org/10.1038/264746a0.

Miyazaki, M., Nozaki, D., & Nakajima, Y. (2005). Testing Bayesian models of human coincidence timing. *Journal of Neurophysiology*, *94*(1), 395–399. https://doi.org/10.1152/jn.01168.2004.

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *Journal of Neuroscience*, *16*(5), 1936–1947. https://doi.org/10.1523/JNEUROSCI.16-05-01936.1996.

Monti, J. M., & Monti, D. (2007). The involvement of dopamine in the modulation of sleep and waking. *Sleep Medicine Reviews*, *11*(2), 113–133. https://doi.org/10.1016/j.smrv.2006.08.003.

Morey, R. D. (2008). Confidence intervals from normalized data: A correction to Cousineau (2005). *Tutorials in Quantitative Methods for Psychology*, *4*(2), 61–64. https://doi.org/10.20982/tqmp.04.2.p061.

Nash, J. C. (2014). On best practice optimization methods in R. *Journal of Statistical Software*, *60*(2), 1–14. https://doi.org/10.18637/jss.v060.i02.

Nash, J. C., & Varadhan, R. (2011). Unifying optimization algorithms to aid software system users: Optimx for R. *Journal of Statistical Software*, *43*(9), 1–14. https://doi.org/10.18637/jss.v043.i09.

Niv, Y., Daw, N. D., Joel, D., & Dayan, P. (2007). Tonic dopamine: Opportunity costs and the control of response vigor. *Psychopharmacology*, *191*(3), 507–520. https://doi.org/10.1007/s00213-006-0502-4.

Olds, J., & Milner, P. (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *Journal of Comparative and Physiological Psychology*, *47*(6), 419–427. http://dx.doi.org/10.1037/h0058775.

Osmont, A., Cassotti, M., Agogué, M., Houdé, O., & Moutier, S. (2015). Does ambiguity aversion influence the framing effect during decision making? *Psychonomic Bulletin & Review*, *22*(2), 572–577. https://doi.org/10.3758/s13423-014-0688-0.

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442. http://dx.doi.org/10.1163/156856897X00366.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, *442*(7106), 1042–1045. https://doi.org/10.1038/nature05051.

Pohlitzer-Ahles, S. (2018). An introduction to within-subject confidence intervals. [Blog post]. Retrieved from http://www.mypolyuweb.hk/~sjpolit/confidenceintervals.html

R Core Team. (2017). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/

Rahnev, D., & Denison, R. N. (2018). Suboptimality in Perceptual Decision Making. *Behavioral and Brain Sciences*. Advance online publication. https://doi.org/10.1017/S0140525X18000936.

Reynolds, J. N., & Wickens, J. R. (2002). Dopamine-dependent plasticity of corticostriatal synapses. *Neural Networks*, *15*(4-6), 507–521. https://doi.org/10.1016/S0893-6080(02)00045-X.

Salamone, J. D., & Correa, M. (2012). The mysterious motivational functions of mesolimbic dopamine. *Neuron*, *76*(3), 470–485. https://doi.org/10.1016/j.neuron.2012.10.021.

Schultz, W. (1998). Predictive reward signal of dopamine neurons. *Journal of Neurophysiology*, *80*(1), 1–27. https://doi.org/10.1152/jn.1998.80.1.1.

Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron*, *36*(2), 241–263. https://doi.org/10.1016/S0896-6273(02)00967-4.

Schultz, W. (2007a). Behavioral dopamine signals. *Trends in Neurosciences*, *30*(5), 203–210. https://doi.org/10.1016/j.tins.2007.03.007.

Schultz, W. (2007b). Multiple dopamine functions at different time courses. *Annual Review of Neuroscience*, *30*, 259–288. https://doi.org/10.1146/annurev.neuro.28.061604.135722.

Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science*, *275*(5306), 1593–1599. https://doi.org/10.1126/science.275.5306.1593.

Seriès, P., & Seitz, A. (2013). Learning what to expect (in visual perception). *Frontiers in Human Neuroscience*, *7*, 668. https://doi.org/10.3389/fnhum.2013.00668.

Shadmehr, R., de Xivry, J. J. O., Xu-Wilson, M., & Shih, T.-Y. (2010). Temporal discounting of reward and the cost of time in motor control. *Journal of Neuroscience*, *30*(31), 10507–10516. https://doi.org/10.1523/JNEUROSCI.1343-10.2010.

Streiner, D. L., & Norman, G. R. (2011). Correction for multiple testing: Is there a resolution? *Chest*, *140*(1), 16–18. https://doi.org/10.1378/chest.11-0523.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.

Swanson, L. (1982). The projections of the ventral tegmental area and adjacent regions: A combined fluorescent retrograde tracer and immunofluorescence study in the rat. *Brain Research Bulletin*, *9*(1-6), 321–353. https://doi.org/10.1016/0361-9230(82)90145-9.

Tassinari, H., Hudson, T. E., & Landy, M. S. (2006). Combining priors and noisy visual cues in a rapid pointing task. *Journal of Neuroscience*, *26*(40), 10154–10163. https://doi.org/10.1523/JNEUROSCI.2779-06.2006.

Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, *331*(6022), 1279–1285. https://doi.org/10.1126/science.1192788.

Ting, C.-C., Yu, C.-C., Maloney, L. T., & Wu, S.-W. (2015). Neural mechanisms for integrating prior knowledge and likelihood in value-based probabilistic inference. *Journal of Neuroscience*, *35*(4), 1792–1805. https://doi.org/10.1523/JNEUROSCI.3161-14.2015.

Tobler, P. N., O'Doherty, J. P., Dolan, R. J., & Schultz, W. (2006). Human neural learning depends on reward prediction errors in the blocking paradigm. *Journal of Neurophysiology*, *95*(1), 301–310. https://doi.org/10.1152/jn.00762.2005.

Tolhurst, D. J., Movshon, J. A., & Dean, A. F. (1983). The statistical reliability of signals in single neurons in cat and monkey visual cortex. *Vision Research*, *23*(8), 775–785. https://doi.org/10.1016/0042-6989(83)90200-6.

Trommershäuser, J., Maloney, L. T., & Landy, M. S. (2003). Statistical decision theory and trade-offs in the control of motor response. *Spatial Vision*, *16*(3), 255–275. https://10.1163/156856803322467527.

Tsai, H.-C., Zhang, F., Adamantidis, A., Stuber, G. D., Bonci, A., De Lecea, L., & Deisseroth, K. (2009). Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. *Science*, *324*(5930), 1080–1084. https://doi.org/10.1126/science.1168878.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124–1131. https://doi.org/10.1126/science.185.4157.1124.

Vilares, I., Howard, J. D., Fernandes, H. L., Gottfried, J. A., & Kording, K. P. (2012). Differential representations of prior and likelihood uncertainty in the human brain. *Current Biology*, *22*(18), 1641–1648. https://doi.org/10.1016/j.cub.2012.07.010.

Vilares, I., & Kording, K. P. (2011). Bayesian models: The structure of the world, uncertainty, behavior, and the brain. *Annals of the New York Academy of Sciences*, *1224*(1), 22–39. https://doi.org/10.1111/j.1749-6632.2011.05965.x.

Vilares, I., & Kording, K. P. (2017). Dopaminergic medication increases reliance on current information in Parkinson's disease. *Nature Human Behaviour*, *1*(8). https://doi.org/10.1038/s41562-017-0129.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience*, *5*(6), 598. https://doi.org/10.1038/nn0602-858.

Whelan, R. (2008). Effective analysis of reaction time data. *The Psychological Record*, *58*(3), 475–482. https://doi.org/10.1007/BF03395630.

Wickham, H. (2017). *Tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1. Retrieved from https://CRAN.R-project.org/package=tidyverse

Wise, R. A. (1996). Neurobiology of addiction. *Current Opinion in Neurobiology*, *6*(2), 243–251. https://doi.org/10.1016/S0959-4388(96)80079-1.

Wolpert, D. M., & Ghahramani, Z. (2009). Bayes rule in perception, action and cognition. In T. Bayne, A. Cleeremans, & P. Wilken (Eds.), *The Oxford Companion to Consciousness*. Oxford, United Kingdom: Oxford University Press.

World Health Organization. (2011). Informed consent form templates. Retrieved June 05, 2018, from http://www.who.int/rpc/research_ethics/informed_consent/en/.

Appendix A

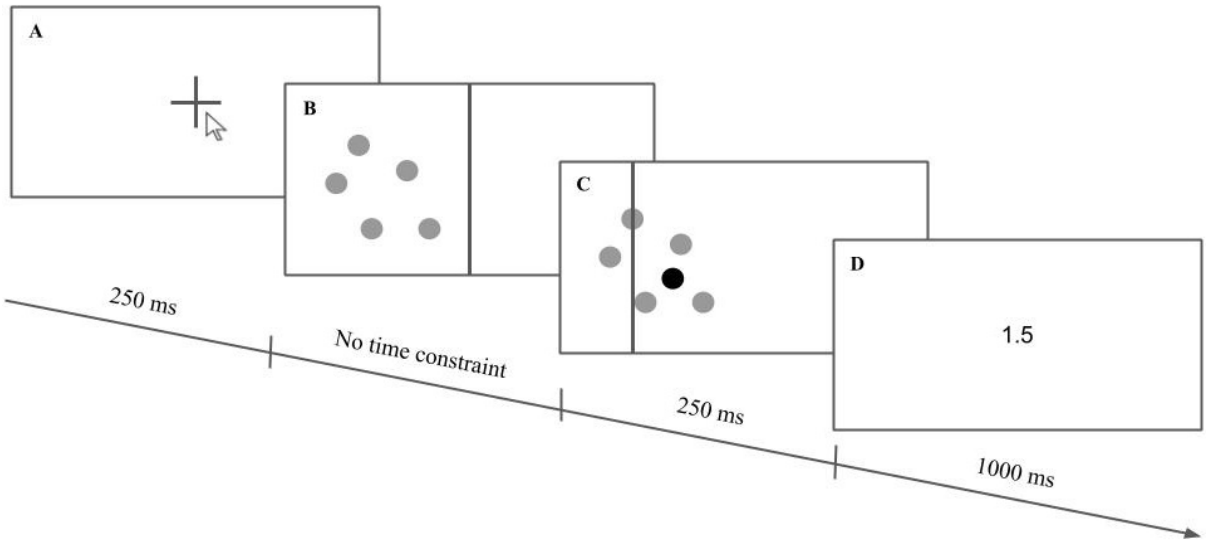


Figure A1. Illustration of the experimental task. A) A central fixation cross is displayed and disappears 250 ms after the mouse cursor is moved to the center of the screen B) The likelihood is displayed as a random cloud on a trial-by-trial basis, sampled from a Gaussian distribution centered at the horizontal target position. Participants move the horizontal bar to the estimated target position. C) The actual target location (here a dot in black) is displayed for 250 ms after participants confirm their estimates via a mouse button press. D) The trial reward is shown for 1000 ms.

Appendix B

Table B1

Mean target deviation, reaction time, and sensory weight with corresponding standard deviations for each experimental condition.

Predictor			Reaction time	Target deviation	Sensory weight
Prior	Likelihood	Reward	$M(SD)$	$M(SD)$	$M(SD)$
Wide	Wide	High	1.715(1.364)	.06(.047)	.641(.206)
Wide	Wide	Low	1.832(1.379)	.06(.048)	.654(.208)
Wide	Narrow	High	1.603(1.183)	.029(.024)	.671(.159)
Wide	Narrow	Low	1.715(1.265)	.029(.023)	.681(.16)
Narrow	Wide	High	1.634(1.211)	.046(.042)	.607(.23)
Narrow	Wide	Low	1.59(1.06)	.045(.042)	.6(.221)
Narrow	Narrow	High	1.492(1.076)	.023(.018)	.627(.217)
Narrow	Narrow	Low	1.47(.996)	.021(.017)	.625(.209)

Appendix C

Table C1

Main analysis for reaction time, target deviation, and sensory weight.

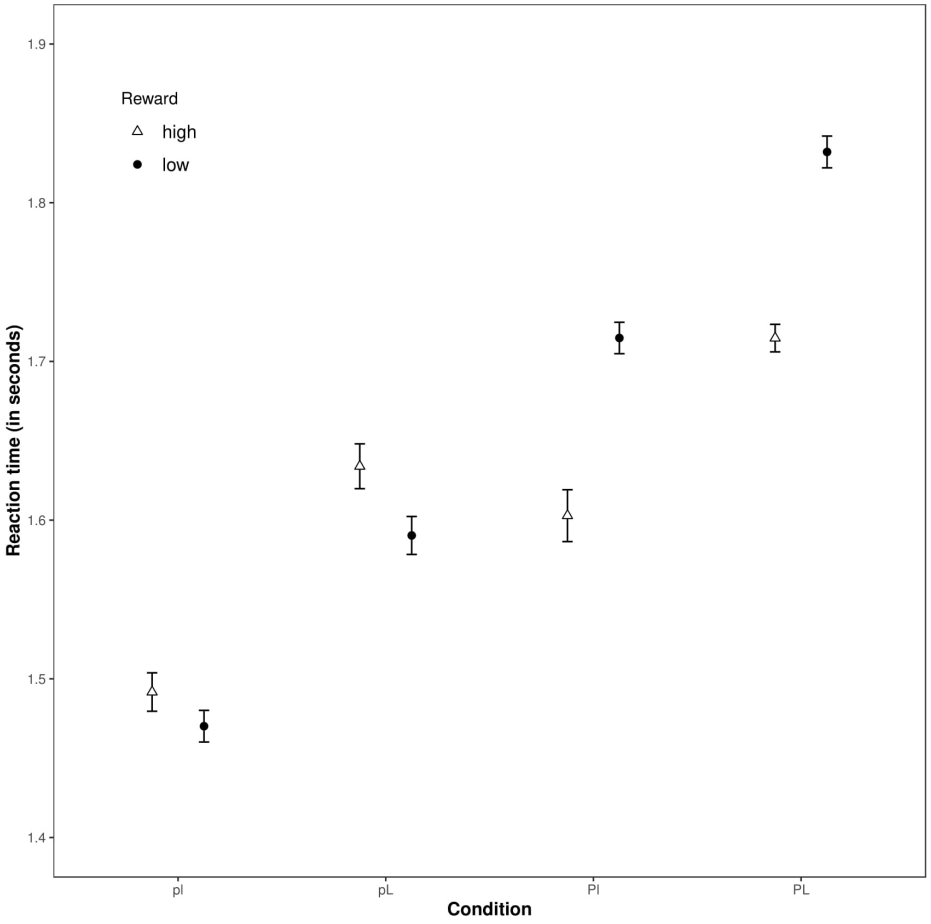
Response	Predictor	β	SE	df	t	p
Reaction	(Intercept)	-.289	.086	24	-3.363	.003*
time	Prior	.093	.059	24	1.579	.128
	Likelihood	.063	.01	24.02	6.196	$< .001$*
	Reward	.032	.059	24	.545	.591
	Prior x Likelihood	.029	.014	14150	2.105	.035*
	Prior x Reward	-.066	.014	14150	-4.851	$< .001$*
	Likelihood x Reward	-.001	.014	14150	-.073	.942
	Prior x Likelihood x Reward	.018	.027	14150	.65	.516
Target	(Intercept)	.305	.003	24.01	109.307	$< .001$*
deviation	Prior	-.03	.004	23.98	-7.476	$< .001$*
	Likelihood	-.072	.004	24.03	-19.108	$< .001$*
	Reward	.002	.002	26.03	1.275	.214
	RPE	.001	.003	27.52	.54	.593
	Prior x Likelihood	.016	.003	14160	4.838	$< .001$*
	Prior x Reward	.005	.003	14150	1.591	.112
	Prior x RPE	.005	.005	13610	1.072	.284
	Likelihood x Reward	.001	.003	14160	.213	.832
	Likelihood x RPE	-.009	.005	14180	-1.878	.06
	Reward x RPE	.003	.005	14110	.74	.459
	Prior x Likelihood x Reward	.006	.006	14160	.924	.356
	Prior x Likelihood	-.006	.009	14190	-.626	.531

Table continued on next page

Response	Predictor	β	SE	df	t	p
Target	x RPE					
deviation	Prior x Reward	-.003	.009	14150	-.307	.759
	x RPE					
	Likelihood x Reward	-.011	.009	14190	-1.229	.219
	x RPE					
Sensory	(Intercept)	.064	.008	23.98	83.907	$< .001$*
weight	Prior	-.047	.006	24.14	-7.873	$< .001$*
	Likelihood	.025	.004	41.27	6.33	$< .001$*
	Reward	-.003	.004	47.94	-.889	.378
	RPE	.001	.005	40.94	.211	.834
	Prior x Likelihood	-.006	.007	14180	-.924	.355
	Prior x Reward	.016	.007	14170	2.451	.014*
	Prior x RPE	.011	.009	13810	1.124	.261
	Likelihood x Reward	$< .001$	.007	14180	.044	.965
	Likelihood x RPE	$< .001$	.009	14210	.011	.992
	Reward x RPE	-.007	.009	14150	-.795	.427
	Prior x Likelihood	-.007	.013	14180	-.523	.601
	x Reward					
	Prior x Likelihood	-.004	.019	14210	-.237	.812
	x RPE					
	Prior x Reward	.001	.019	14170	.047	.963
	x RPE					
	Likelihood x Reward	-.002	.019	14210	-.087	.931
	x RPE					

Note. *Statistically significant at $alpha = .05$.

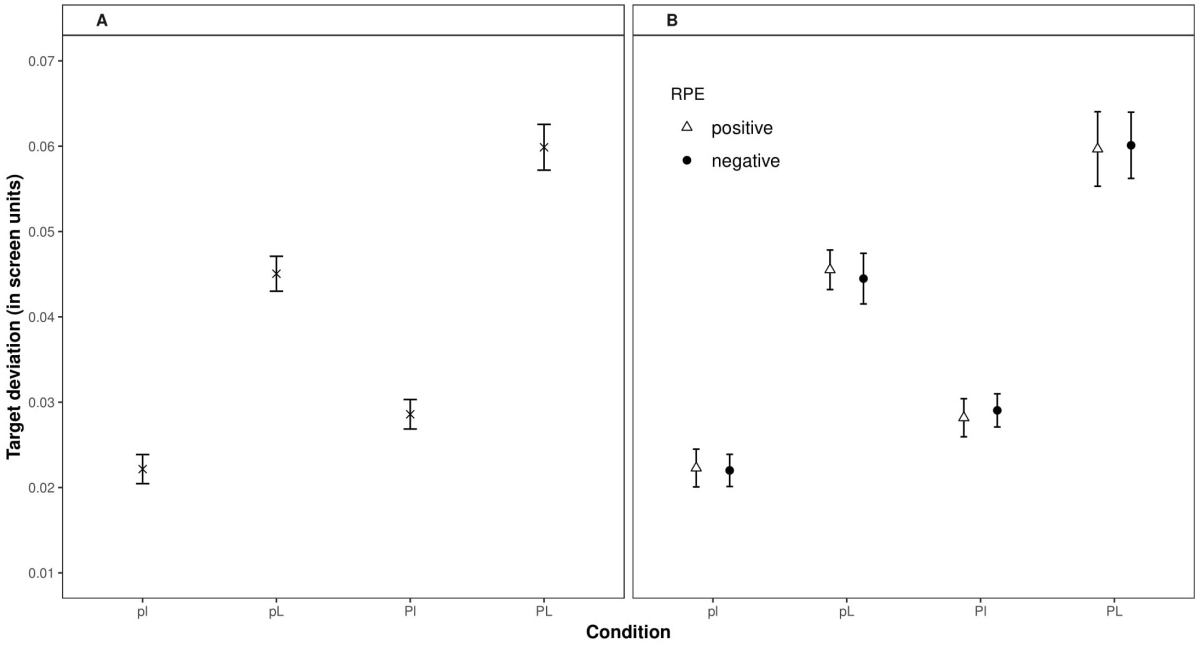
Appendix D



Note. Factor level abbreviations: prior (p = narrow; P = wide) and likelihood (l = narrow; L = wide).

Figure D1. Reaction times for each experimental condition. The main effect of likelihood ($t\,(24.02) = 6.196, p < .001$) was significant. An additional interaction between prior and likelihood ($t\,(14150) = 2.105, p = .035$) indicated that the effect of narrow likelihoods was enhanced under low prior uncertainty. Additionally, there was a significant interaction between prior and reward ($t\,(14150) = -4.851, p < .001$); high reward was associated with a speed advantage when the prior was wide, but low reward elicited faster responses when the prior was narrow. The error bars depict the 95 percent confidence interval for each condition.

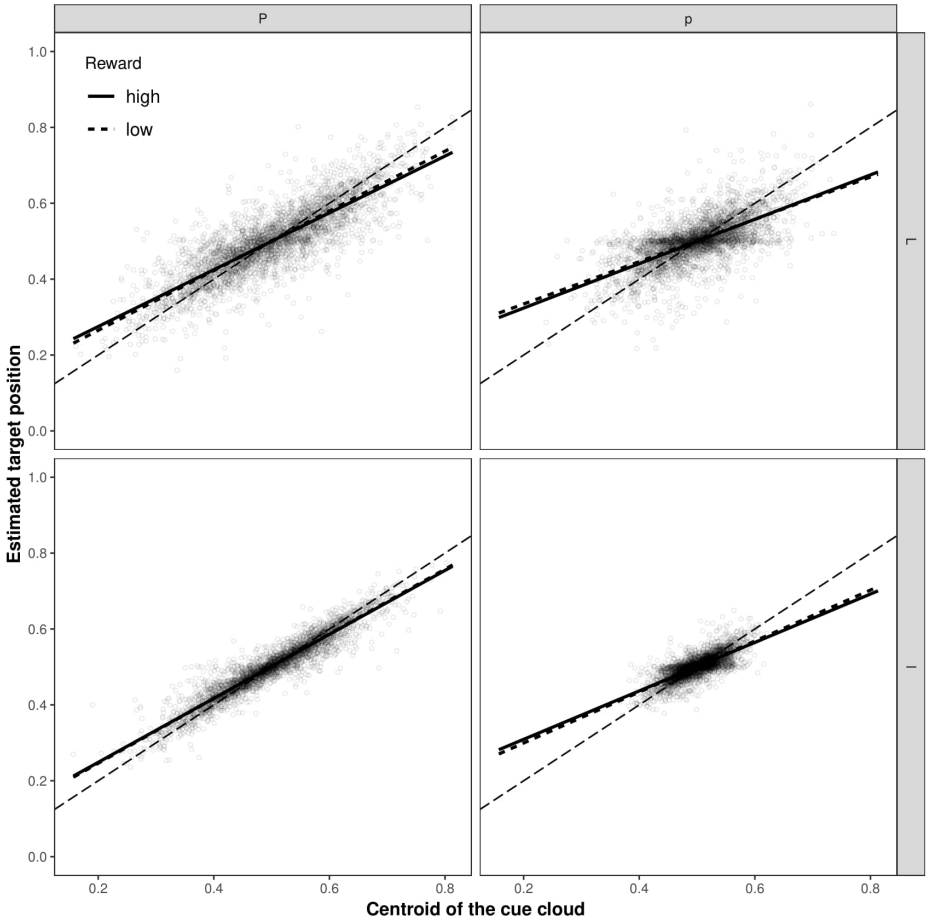
Appendix E



Note. Factor level abbreviations: prior (p = narrow; P = wide) and likelihood (l = narrow; L = wide).

Figure E1. (A) Target deviation for each combination of prior and likelihood averaged over reward. The main effects of prior ($t\,(23.98) = -7.476, p < .001$) and likelihood ($t\,(24.03) = -19.108, p < .001$), and the interaction between prior and likelihood (($t\,(14160) = 4.838, p < .001$)) were significant. (B) Target deviations as a function of RPE sign for each combination of prior and likelihood. The main effect of RPE ($t\,(23.82) = .276, p = .785$) was not statistically significant. Error bars denote 95 percent confidence intervals of the mean in (A) and (B).

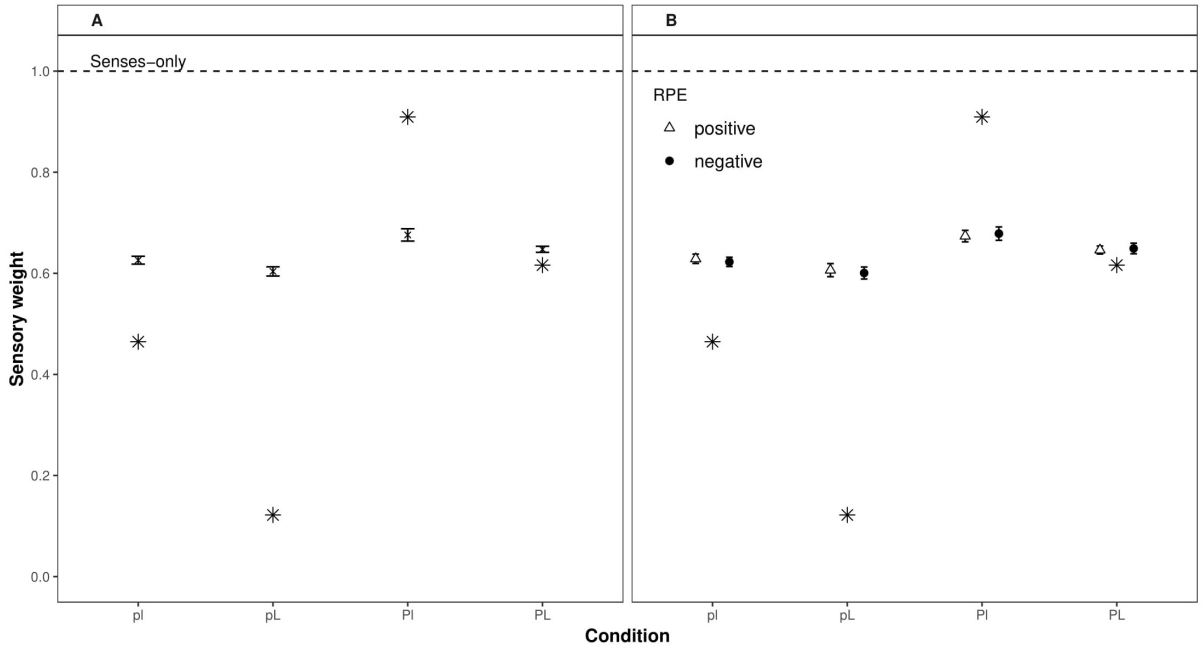
Appendix F



Note. Factor level abbreviations: prior (p = narrow; P = wide) and likelihood (l = narrow; L = wide).

Figure F1. Estimated target position as a function of the centroid of the cue cloud. The diagonal dashed line illustrates complete reliance on the likelihood ($sw = 1$), whereas a flat slope would be expected if participants exclusively employed prior knowledge ($sw = 0$). The main effects of prior ($t\,(24.14) = -7.873, p < .001$) and likelihood ($t\,(41.27) = 6.33, p < .001$) were significant, whereas the main effect of reward ($t\,(47.94) = -.889, p = .378$) was not significant. There was a crossed interaction between prior and reward ($t\,(14170) = 2.451, p = .014$), which signified that sensory weights were higher under high reward when prior uncertainty was low and vice versa.

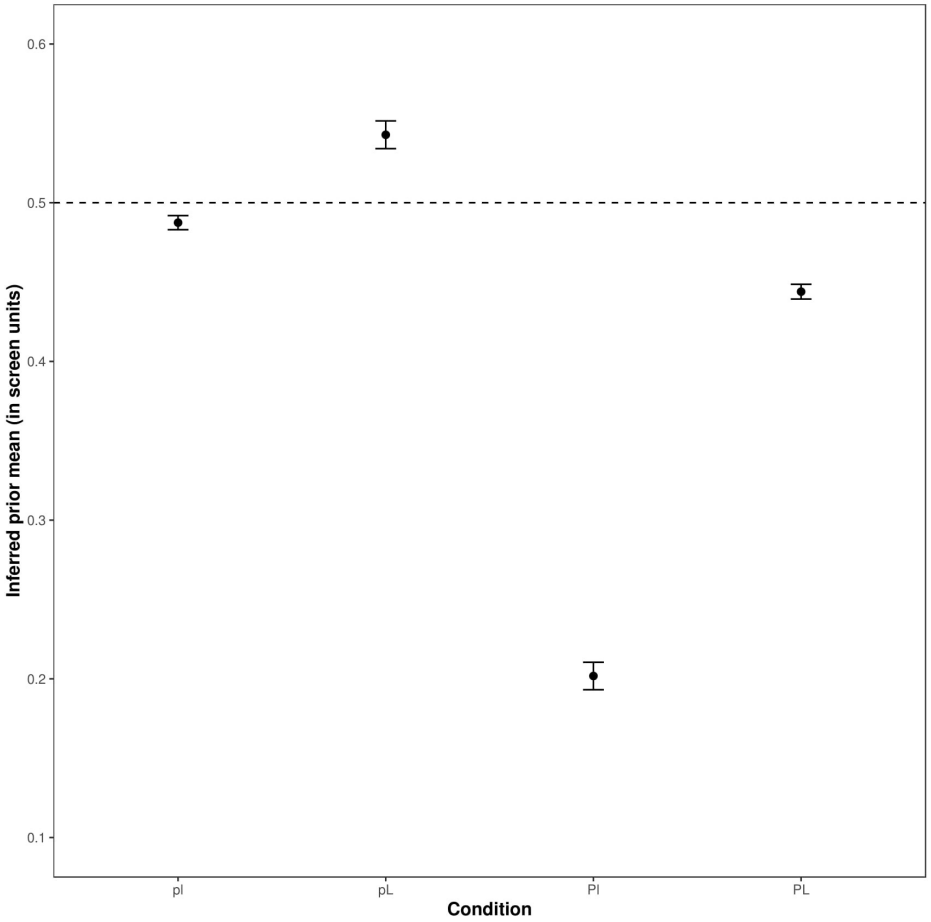
Appendix G



Note. Factor level abbreviations: prior (p = narrow; P = wide) and likelihood (l = narrow; L = wide).

Figure G1. (A) Sensory weights for each combination of prior and likelihood averaged over reward. Participants' sensory weights displayed deviations from the Bayes-optimal weights (e.g., pL: $t\,(23) = 47.237, p < .001$) and the senses-only model (e.g., pL: $t\,(23) = -38.866, p < .001$) across conditions. (B) Sensory weights as a function of RPE for each experimental condition. The main effect of RPE ($t\,(65.84) = .07, p = .945$) was not significant. Error bars depict 95 percent confidence intervals of the mean in (A) and (B).

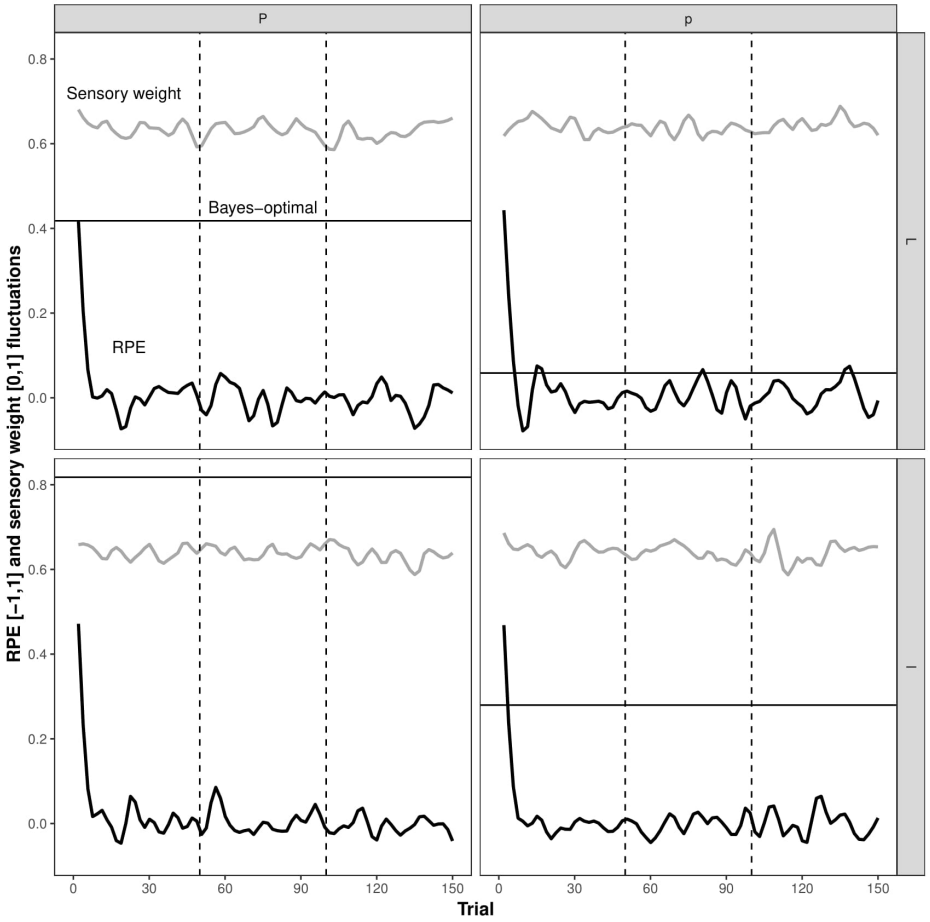
Appendix H



Note. Factor level abbreviations: prior (p = narrow; P = wide) and likelihood (l = narrow; L = wide).

Figure H1. Inferred prior means for each combination of prior and likelihood averaged over reward. The dashed line indicates the true mean of the prior. Error bars illustrate 95 percent confidence intervals of the mean. Participants' average prior estimates were dissimilar to the actual prior mean across conditions (e.g., pl: $t\,(23) = -2.1, p = .047$).

Appendix I



Note. Factor level abbreviations: prior (p = narrow; P = wide) and likelihood (l = narrow; L = wide).

Figure I1. Trial-based analysis for sensory weights. The main effects of prior and likelihood were present throughout the experiment. During the last fifty trials, the interaction between prior and reward reached statistical significance ($t\,(4723) = 3.46$, $p < .001$). Note that during the last interval the main effect of prior was strongest while the likelihood had its weakest effect. Participants' sensory weights failed to converge onto the Bayes-optimal weights across conditions.