Prospective uncertainty: The range of possible futures in physical predictions

Kevin A Smith (k2smith@ucsd.edu) and Edward Vul (evul@ucsd.edu)
Department of Psychology
University of California, San Diego
9500 Gilman Drive
La Jolla, CA 92093 USA

Abstract

Recent research has suggested that people make physical predictions based on extrapolation from a noisy representation of the world, which gives rise to a probabilistic distribution over possible future worlds. But can people use the uncertainty of their predictions to inform their decisions, or can people access only a single possible future? Here we demonstrate that confidence-sensitive decisions about the future track the amount of uncertainty expected from probabilistic forward extrapolations. Participants were asked to make predictions about where a ball would go and indicate an expected range around that prediction. This range was well correlated with two measures of uncertainty: variability in predictions across participants and the amount of uncertainty expected by a model of physical prediction. This suggests that people form a probabilistic distribution over possible futures in the course of physical prediction and base their decisions about the future on this range.

Keywords: prediction; uncertainty; physical reasoning; noisy Newton physics

Introduction

Imagine you are playing pool, and the game is coming to the end. You have a straight shot to win the game: the most likely outcome of the shot you are planning is for the cue ball to hit the 8-ball and knock it into the pocket, winning the game. But what if you are wrong? If you accidentally hit the cue ball differently (e.g., too hard or with too little English), it may also go into the pocket and you will instead lose. Thus if you consider alternative possibilities, you might think that your planned shot is too risky. For many real-world situations, it is imperative to not only consider the most likely outcome of a physical event, but also less likely possibilities (Kording, 2007). But do we actually take the range of possible futures into account when we make such predictions?

Recent research has suggested that people accomplish physical prediction tasks by extrapolating the future state of the world using accurate laws of Newtonian mechanics, but also incorporate uncertainty about the current state of the world – termed the "noisy Newton" theory of physical reasoning (Battaglia, Hamrick, & Tenenbaum, 2013; Sanborn, Mansinghka, & Griffiths, 2013; Smith & Vul, 2013). In this framework, it is assumed that people start with a model of the current state of the world that includes uncertainty about the locations, attributes, and motions of objects due to unobservable properties (e.g., friction coefficients) and noisy perceptual estimates. People then use an "intuitive physics engine" to propagate the world forward given this uncertainty and obtain probabilistic estimates about the objects' future locations and motions, drawing on these probabilistic predictions to make relevant decisions (Battaglia et al., 2013; Smith, Dechter, Tenenbaum, & Vul, 2013; Smith & Vul, 2013).

However, while prior work has assumed that that people develop a probability distribution over possible future physical states, this theory has not been directly tested. Although prior physical models explain behavior across people and scenarios, it is possible that aggregate probabilistic behavior can arise from individuals who each only consider a single future outcome (Daw & Courville, 2008) and who would thus be incapable of considering the dispersion of possible outcomes. In this case, prior models would not describe how individuals behave, but would only constrain behavior averaged over large numbers of people or scenarios. Therefore, we aim to investigate the richness of individuals' representations about the future: do we consider a distribution of possible physical outcomes, or just a single sampled outcome?

Prior studies of decisions under uncertainty in sensory-motor tasks have shown that people behave as if they have access to a full probability distribution, or at least the first few moments (e.g., mean and variance). For instance, we can integrate haptic and visual information in a way that is sensitive to the variance of the noise in each of the modalities (Ernst & Banks, 2002), we behave as if we are inferring the expected value over a probabilistic distribution of outcomes in visual (Whiteley & Sahani, 2008) and motor (Trommershäuser, Landy, & Maloney, 2006) gambles, and we can make rational tradeoffs between our visual and motor uncertainty (Battaglia & Schrater, 2007). Similarly, we are sensitive to probability distributions over more abstract outcomes; our conditional predictions about quantities like the total baking time of a cake reflect an appreciation of not only the mean and variance, but also the shape of the distribution of cake baking times (Griffiths & Tenenbaum, 2006). Yet most of these tasks require reasoning about a probability distribution over the current state of the world, or internal noise that might be learned from experience (e.g., motor uncertainty). Thus it is not clear whether this knowledge of uncertainty would extend to tasks in which the uncertainty arises from stochastic extrapolation of the physical world.

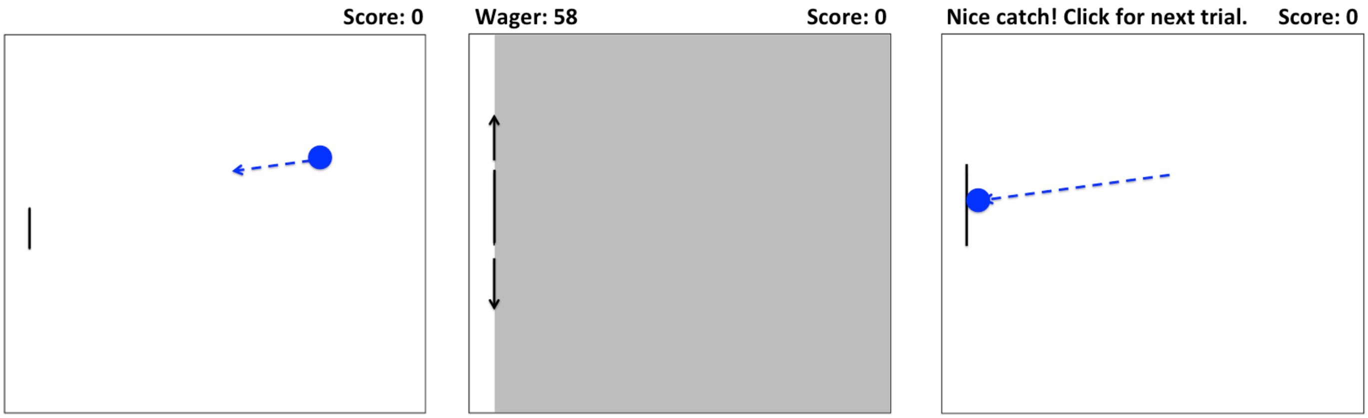

Figure 1: Diagram of a trial. *Left:* Participants observe a ball in motion for 500ms (dotted line is not shown). *Center:* The path of the ball becomes occluded and participants must move the paddle and change its size to catch the ball when it would cross the plane of the paddle. A ‘Wager’ is shown that grows larger with smaller paddles, and vice versa. *Right:* After a response is registered, the continued path of the ball is shown and participants earn points if the ball is caught.

Here we first present an experiment in which participants made predictions about a range of outcomes where an object would end up after moving under occlusion, and find that participants’ predicted range widths are correlated with the amount of variability across participants, suggesting people know when their predictions will be more variable. We then show the breadths of these predictions are consistent with the uncertainty arising from forward predictions in a prior noisy Newton model of human physical extrapolation (Smith & Vul, 2013). Together, these results suggest that people can, and do, use the prospective uncertainty of their physical predictions to inform their judgments.

Experiment

To measure subjective uncertainty in human physical predictions, we showed participants a ball bouncing around a rectangular table on a computer screen and asked them to predict where the ball would end up after a period of occlusion by adjusting the vertical position and size of a paddle.

Methods

Fifty UCSD undergraduate students participated in this experiment for course credit. All participants had normal or corrected to normal vision. Of these participants, seven were excluded from analysis because we did not capture a full set of data (either due to participants leaving early or data recording errors), leaving 43 participants’ data for review.

Participants viewed a 40cm x 32cm computer monitor from a distance of 60cm. The screen depicted a “table” (31.25cm x 28.125cm) that a computerized ball would move around, bouncing off of the table walls according to idealized Newtonian physics, implemented in the Chipmunk 2D physics engine (Lembcke, 2011). The ball always moved at a constant velocity of 15.625cm/s. Participants controlled a paddle at one end of the table: they could move the paddle vertically (but not horizontally) with the mouse, and could change the size of the paddle with the mouse scroll wheel. Paddle size was constrained such that the smallest size was 0.78cm, and the largest was 7.8cm. The paddle size always started at 3.1cm, but participants were required to adjust the size at least once to discourage them from simply using the default setting.

Participants observed the motion of the ball for 500ms (Figure 1, *left*), and then a portion of the screen would be occluded. Participants were then asked to position the paddle so that it would catch the ball if the ball continued on its trajectory (Figure 1, *center*). Time was not limited, so participants could spend as long as they liked positioning and resizing the paddle before clicking the mouse to register their response. Participants earned points if they caught the ball. Feedback was provided by showing the path of the ball, starting from the point where it was occluded and traveling until it reached the plane of the paddle. Finally, a notification appeared if the participants had caught the ball and earned points (Figure 1, *right*).

To motivate participants to use paddle sizes other than the largest, the number of points earned for a catch was inversely proportional to the size of the paddle. Thus we expected that when participants were more certain, they would confidently choose a smaller paddle size to earn more points (and vice versa when they were less certain). Points were used as an intrinsic motivation but did not otherwise affect credit or compensation for participants.

Each participant was provided with the same 450 trials (differing in the observed motion of the ball and horizontal location of the paddle) in a randomized order. There were 50 trials in each of nine conditions that varied how the ball would move while occluded: the distance the ball travelled (short: 18.75cm, medium: 25cm, long: 31.25cm) crossed with the number of times the ball bounced off of the sides of the table (0, 1, or 2) before it reached the plane of the paddle.[1] To ensure the ball would travel a fixed distance, the

[1] Two of the trials from the short, 2 bounce condition were excluded from analysis due to an error that led to different observations across participants.

horizontal position of the paddle varied across trials.

For each trial, we obtained two measurements from each participant: the center point of the paddle to indicate their best guess about where the ball would go, and the size of the paddle to indicate their confidence in that prediction. On each trial, we also determined where the ball actually crossed the plane of the paddle according to computerized Newtonian mechanics, and we will call this location "ground truth."

Participants were given five practice trials with no occlusion to ensure that they understood the ball would move with Newtonian mechanics and then eight further practice trials with occluded movement to ensure they were comfortable with the task and controls.

Results

Because idealized Newtonian mechanics is deterministic, the ground truth outcomes from the computer physics engine have no variability, and thus we cannot compare participants' internal measures of uncertainty against "ground truth variability" in the trial outcomes based on computer physics. Instead, we measured whether people have an internal measure of variability that accurately reflects the variability in behavior across people, which can be used as a proxy for the amount of variability in any individual's predictions. We aggregated across people within each trial to calculate two measures of uncertainty: (1) *across-subject variability* in behavior was measured as the standard deviation of participants' predictions for each trial, and (2) *within-subject uncertainty* was measured as the average paddle length used for each trial. Because these measures were determined from separate response dimensions, any relationships between the two measures must be modulated solely by participants' own conception of their prediction uncertainty.

Uncertainty across conditions We initially tested whether our measures of across-subject variability and within-subject uncertainty followed our qualitative predictions across conditions: as the difficulty increased with greater distance or more bounces, would both variability and uncertainty increase? Prior work suggests that variability across participants should increase as both distance and number of bounces increase (Smith & Vul, 2013), so if people have an accurate representation of their own uncertainty, they should make the paddle larger in the more difficult conditions as well.

Consistent with prior work, we find that the standard deviation of predictions across participants increases with distance ($F(2,439)=51$, $p<0.001$) and number of bounces ($F(2,439)=77$, $p<0.001$), and is somewhat modulated by the interaction of the two ($F(4,439)=2.7$, $p=0.030$; see Figure 2, top), which suggests that the difference between distances becomes smaller with more bounces. Post-hoc analyses using a Tukey Range Test suggest that both the short and medium no-bounce conditions are significantly less variable than all other groups, and the long and medium one and two

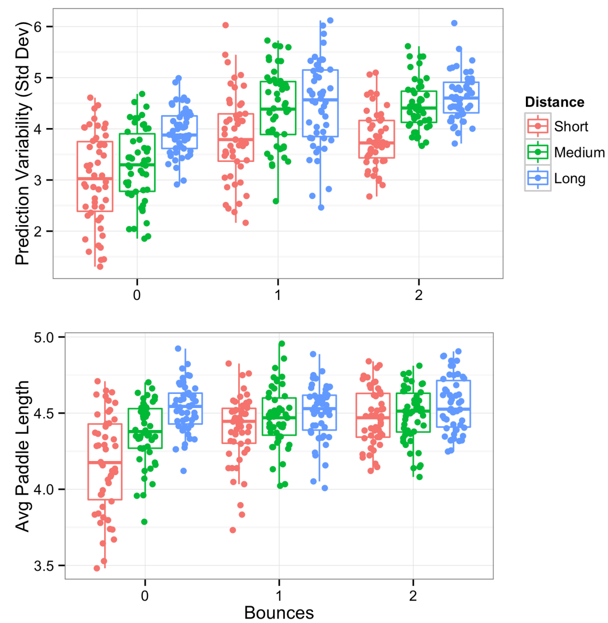

Figure 2: *Top:* Across-subject variation (standard deviation of paddle center across participants [cm]). *Bottom:* Within-subject uncertainty (average paddle size across participants [cm]). Each point represents a single trial, grouped by condition, with a boxplot representing the interquartile range across trials for that condition. Uncertainty, represented by greater variability and paddle lengths, increased with more bounces and greater distance.

bounce groups are significantly more variable than all other groups.

A similar pattern is seen in the average length of the paddles: participants use larger paddles as the distance ($F(2,439)=30$, $p<0.001$) and number of bounces ($F(2,439)=19$, $p<0.001$) increase, though again there is an interaction ($F(4,439)=7.0$, $p<0.001$; see Figure 2, bottom), again due to smaller differences between distance conditions with 1 or 2 bounces. Post-hoc analysis suggests that this was more driven by the short, no bounce condition, for which people used significantly shorter paddles than all other conditions, and the medium, no-bounce condition, which had significantly smaller paddles than any of the long conditions. All analysis was performed on aggregate trial data, so each trial contributed a single data point to the analysis.

While it is encouraging that both across-subject variability and within-subject uncertainty varied systematically across conditions, there is considerable variation in both measures across trials within each condition; for instance, some trials in the easiest (short, no bounce) condition yielded greater uncertainty than some trials in the hardest (long, two bounce) condition. Furthermore, participants cannot know the distance the ball will travel or the number of bounces it will take without first extrapolating the motion of the ball, so the trial condition is only meaningful to subjects insofar as it maps onto their own extrapolation. We therefore next ask whether variability and uncertainty differ systematically across trials regardless of condition.

Meta-knowledge of uncertainty by trial In order to test whether people have accurate access to prediction uncertainty beyond gross trial categories, we can ask whether the measures of within-subject uncertainty (paddle sizing) correlate with across-subject variability across trials.

The standard deviation of predictions across participants correlated significantly with the average paddle length for each trial (*r=0.45, 95% CI: [0.37, 0.51]*, see Figure 3). This correlation remains significant even when variability due to trial condition (number of bounces and distance) is factored out ($r_{partial}$*=0.23, 95% CI: [0.14, 0.31]*), indicating that people discern their uncertainty for specific trials based on subtle differences in observations.

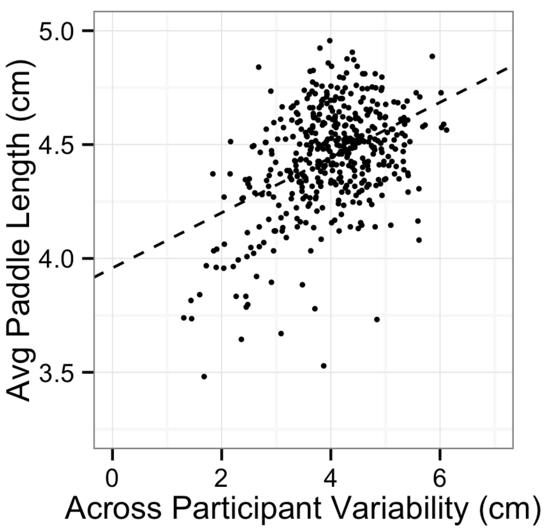

Figure 3: Within-subject uncertainty (average paddle length) is well explained by across-subject variability (standard deviation of paddle center; *r=0.45*). Each point represents a separate trial.

Individual differences in certainty While we demonstrated that there is an aggregate relationship between within-subject uncertainty as measured by mean paddle size and across-subject variability of paddle placement, there was also a large amount of individual variability in the risk participants were willing to take – some generally used smaller paddles, and some were more risk averse and almost always used as large a paddle as possible. The average paddle size used varied across participants from 1.9cm (25% of the maximum paddle size) to 7.6cm (98% of the maximum size).

Furthermore, participants were remarkably consistent in their relative paddle sizing. We calculated how each participant's paddle size differed from the average for each trial. The split-half correlation of this difference (averaged across two sets of trials, within participants) was extremely high: (*r=0.998, 95% CI: [0.996, 0.999]*), suggesting that some participants in general are simply more or less risk averse. However, this risk aversion was not driven by some participants knowing they are more accurate – there was no appreciable correlation between average paddle sizing and individual average prediction errors (*r=0.04, 95% CI: [-0.26, 0.34]*).

Despite the range in riskiness, we can ask whether there is evidence that individual participants had access to internal measures of uncertainty. Though the standard deviation of predictions for a given trial is necessarily an across-participant measure, we treat this as an approximate measure of each individual participant's internal uncertainty. We therefore tested whether an individual's paddle sizing correlated with the across-participant variability across all trials.

Because there is more noise in individual paddle sizes than average paddle sizes, the individual correlations are lower and extremely variable (*mean r: 0.10, sd: 0.11,* see Figure 4). Nonetheless, there is evidence that participants on average modulate their paddle sizes by how variable across-participant predictions are (*t(42)=6.1, p<0.001*), and we found positive correlations for 74% of participants (32 of 43). This suggests that most individuals informed their paddle sizes by an assessment of prospective uncertainty.

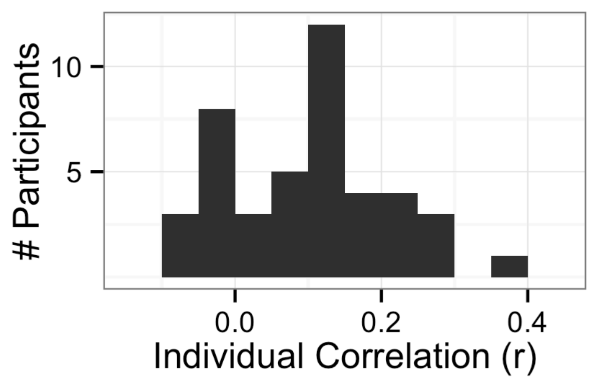

Figure 4: Histogram of correlations between standard deviation of predictions across participants and individual paddle sizes across trials. Most participants had a positive correlation, suggesting individual access to probabilistic distributions over potential future states of the world.

Learning over time Participants received feedback after each trial, which could allow for learning non-physical contingencies between the initial ball motion and where the ball crossed the plane of the paddle. If participants were learning in this way, it is possible that the relationship between variability and paddle size might have arisen from mechanisms other than physical extrapolation.

However, there was evidence of only small improvement in prediction over time. Controlling for subject and trial effects, we find that prediction error (average paddle distance from the ground truth) only decreased by $6.4*10^{-4}$ cm/trial (*95% CI: [$3.1*10^{-4}$, $9.8*10^{-4}$]*), which would suggest an improvement of 0.29 cm over the experiment, or about 6.8% of the average error.

In addition, there was no evidence that participants changed their paddle sizing strategy over the experiment: controlling for subject and trial effects, there was only a size increase of $8.0*10^{-5}$ cm/trial (*95% CI: [$-4.7*10^{-5}$, $20.6*10^{-5}$]*), which would be equivalent to an increase of 0.04 cm, or 0.8% of the average paddle length over the entire experiment. Nor was there evidence that the average participant's scores improved over the experiment: there was only an estimated score improvement of 0.0036 points per trial (*95% CI: [-0.0006, 0.0078]*), which would be an average improvement of only 1.6 points from the initial to final trials.

Because there was little change in participants' performance on the task over the course of the experiment, we can be confident that we are not studying strategies learned during the experiment, but rather physical intuitions that were formed previously.

Forming a representation of uncertainty

Although an individual's uncertainty tracks with the amount of variability across people, we might wonder how that representation of uncertainty is formed. For this we turn to a prior model of physical prediction from Smith and Vul (2013). We show that the amount of uncertainty expected by this physical prediction process explains participants' individual uncertainty, even above and beyond the expectations from across-participant variability in predictions.

Physical prediction model

Smith and Vul (2013) used a task similar to the current experiment to elicit physical predictions.[2] In this study we found that predictions of ball motions were well captured by assuming that people used a noisy Newton model with two general sources of uncertainty: *perceptual uncertainty* about the initial location and trajectory of the ball at the moment its motion was occluded, and *dynamic uncertainty* accounting for accumulated noise as the ball traveled along its path (e.g., it might not travel in a perfectly straight line due to a rough floor or imperfections in the ball). We also found that people had a prior belief that the ball would travel towards the center of the screen, and that their predictions were a combination of motion extrapolation with this prior belief.

The uncertainty and prior parameters for this model were fit based on aggregate participants' predictions, as described in Smith and Vul (2013). Crucially, no information about paddle size was used to parameterize the model, which allows for a clean comparison between the uncertainty in predictions expected by the model and the individual uncertainty expressed by participants via paddle sizing.

[2] The main differences in that study were that participants used a paddle with a fixed length and had a limited time to "catch" the ball.

Accuracy of physical model

As observed in similar work (Hamrick, Smith, Griffiths, & Vul, 2015; Smith, Dechter, Tenenbaum, & Vul, 2014; Smith & Vul, 2013), participants' predictions are well explained by the physical model. Participants' average predictions for each trial were correlated with the ground truth position where the ball crossed the plane of the paddle (*r=0.87, 95% CI: [0.85, 0.89]*), but participants were systematically biased to predict the ball would cross closer to the middle of the screen. Nonetheless, the physical prediction model could account for these systematic biases; this model made similar errors (deviations from ground truth) to participants on each trial (*r=0.80, 95% CI: [0.77,0.83]*). This suggests that we are capturing the process participants are using to make physical predictions.

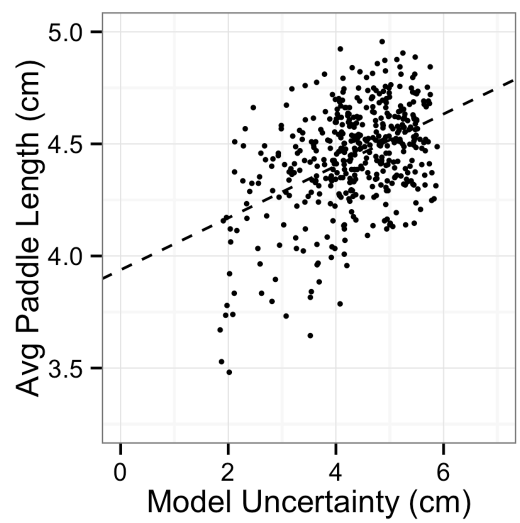

Figure 3: Within-participant average paddle length is well explained by the uncertainty in the noisy Newton model (*r=0.45*). Each point represents a separate trial.

Measures of uncertainty

Because this physical model produces probabilistic predictions for each trial, we can treat the variability in these predictions as a representation of the uncertainty that participants' would be expected to hold for each trial. We can therefore investigate whether participants adjust paddle size based on this prospective uncertainty about possible futures.

The average paddle size for each trial was well correlated with the uncertainty expected under the noisy Newton model (*r=0.45, 95% CI: [0.37, 0.52]*, see Figure 5), which provides evidence that participants were drawing on information from a probability distribution over possible outcomes to decide how confident they are.

However, the uncertainty expected by the noisy Newton model also correlated well with the standard deviation of participants' predictions (*r=0.51, 95% CI = [0.43, 0.57]*). Thus we must ask whether the uncertainty in the model is capturing any internal psychological mechanisms, or simply correlating with paddle sizing because it is explaining the variability in participants' predictions. To test this we can ask whether the model uncertainty has any explanatory power beyond the measure of across-participant standard deviation. We find that with across-participant variability partialed out, there is still a relationship between noisy Newton uncertainty and paddle length ($r_{partial}=0.28$, *95% CI: [0.20,0.37]*). Conversely, we find that across-participant standard deviation can explain the uncertainty produced with paddle sizing, even above and beyond the uncertainty from the noisy Newton model ($r_{partial}=0.29$, *95% CI: [0.20,0.37]*). This pattern of partial correlations indicates that both across-subject variability and noisy forward physical simulations are capturing partially independent aspects of the uncertainty that individuals use when determining their confidence to adjust the size of a paddle.

Discussion

In this paper we show that people have access to an internal representation of uncertainty over potential future world states, which can be used in decisions of how much risk to take on in a particular situation. Across trials, participants shrank their paddle to catch the ball when there was less variability in possible futures and used larger paddles when there was more variability.

Uncertainty in the future was measured in two ways: as the variability of point predictions across participants and as the uncertainty estimated with a probabilistic model of physical reasoning. Both estimates of uncertainty were related, but each separately explained how people use uncertainty to set the size of the paddle. This may be because both are noisy estimates of the same mechanism: across-subject variability in paddle placement will reflect more than the average participant uncertainty, while the average participant uncertainty may not be perfectly captured by our simple physical model. Nonetheless, each measure seems to capture part of the internal representation of uncertainty that people hold.

These results provide further insight into the mechanisms underlying physical prediction, demonstrating that people do hold a probabilistic representation of future physical outcomes, as suggested by the noisy Newton theory. This leads to the question of how this sort of uncertainty arises in the brain. Determining the posterior predictive distribution of where the ball will cross the plane of the paddle is a computationally taxing task, and similar distributions can only be achieved in noisy Newton models through sampling a number of possible futures. Concurrent research suggests that this method might be shared by the mind – we appear to make judgments about physical events based on a limited number of samples of future world states (Hamrick et al., 2015), since a limited number of samples may be enough to make efficient predictions (Vul, Goodman, Griffiths, & Tenenbaum, 2014). While it is possible to get some sense of the spread of predictions based on just a few samples, it remains an open question how people exploit a sampling process to estimate their prospective uncertainty and inform their judgments.

Acknowledgments

Kevin A Smith was supported by a UCSD Chancellor's Interdisciplinary Collaboratories grant. Edward Vul was supported by NSF CPS grant: 1239323.

References

Battaglia, P., Hamrick, J., & Tenenbaum, J. (2013). Simulation as an engine of physical scene understanding. *Proceedings of the National Academy of Sciences, 110*(45), 18327-18332.

Battaglia, P., & Schrater, P. R. (2007). Humans trade off viewing time and movement duration to improve visuomotor accuracy in a fast reaching task. *The Journal of Neuroscience, 27*(26), 6984-6994.

Daw, N., & Courville, A. (2008). The pigeon as particle filter. *Advances in Neural Information Processing Systems, 20*.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*, 429-433.

Griffiths, T. L., & Tenenbaum, J. B. (2006). Optimal prediction in everyday cognition. *Psychological Science, 17*(9), 767-773.

Hamrick, J., Smith, K. A., Griffiths, T. L., & Vul, E. (2015). *Think again? Optimal mental simulation tracks problem difficulty*. Paper presented at the 37th Annual Conference of the Cognitive Science Society, Pasadena, CA.

Kording, K. (2007). Decision theory: What "should" the nervous system do? *Science, 318*, 606-610.

Lembcke, S. (2011). Chipmunk 2D (Version 6.2.0): Howling Moon Software. Retrieved from http://chipmunk-physics.net/

Sanborn, A. N., Mansinghka, V. K., & Griffiths, T. L. (2013). Reconciling intuitive physics and Newtonian mechanics for colliding objects. *Psychological Review, 120*(2), 411-437.

Smith, K. A., Dechter, E., Tenenbaum, J. B., & Vul, E. (2013). *Physical predictions over time.* Paper presented at the 35th Annual Conference of the Cognitive Science Society, Berlin, Germany.

Smith, K. A., Dechter, E., Tenenbaum, J. B., & Vul, E. (2014). Tracking hidden objects with efficient physical prediction. *Journal of Vision, 14*(10), 846.

Smith, K. A., & Vul, E. (2013). Sources of uncertainty in intuitive physics. *Topics in Cognitive Science, 5*(1), 185-199.

Trommershäuser, J., Landy, M. S., & Maloney, L. T. (2006). Humans rapidly estimate expected gain in movement planning. *Psychological Science, 17*(11), 981-988.

Vul, E., Goodman, N., Griffiths, T. L., & Tenenbaum, J. (2014). One and done? Optimal decisions from very few samples. *Cognitive Science, 38*(4), 599-637.

Whiteley, L., & Sahani, M. (2008). Implicit knowledge of visual uncertainty guides decisions with asymmetric outcomes. *Journal of Vision, 8*(3). doi: 10.1167/8.3.2