Consistent physics underlying ballistic motion prediction

Kevin A Smith (k2smith@ucsd.edu),[1] Peter Battaglia (pbatt@mit.edu),[2] Edward Vul (evul@ucsd.edu)[1]
1. University of California, San Diego, Department of Psychology, La Jolla, CA 92093
2. MIT, Department of Brain and Cognitive Sciences, Cambridge, MA 02139

Abstract

Research into human models of intuitive physics typically falls into one of two camps, either claiming that intuitive physics is biased and not representative of real physics, or claiming that it consists of a collection of veridical physical laws. Here we investigate the causes of this tension, suggesting that prediction is based on real physics, but explanation is susceptible to biases. We gave participants three tasks based on the same physical principles: two prediction tasks and one task that required drawing the future path of motion. We found distinct biases in all three tasks; however, the two prediction tasks could be explained by consistent application of real physical principles under uncertainty, while the drawing task produced many more idiosyncratic biases. This suggests that different tests of intuitive physics are capturing different types of knowledge about the world.

Keywords: intuitive physics; uncertainty; ballistic motion prediction

Introduction

Classic studies have suggested that many people base their physical intuitions on incorrect and inconsistent physical theories (Anzai & Yokoyama, 1984; McCloskey, Caramazza, & Green, 1980). Others have reported that people are biased by surface-level differences between tasks (Kaiser, Jonides, & Alexander, 1986), and that their inferences about simple physical situations rely on shallow heuristics and are frequently mistaken (Proffitt & Gilden, 1989; Todd & Warren, 1982). However over the past few years, a number of researchers have explained human physical predictions using quantitative cognitive models that assume people have an accurate and consistent understanding of the laws of physics that they apply flexibly across tasks (Hamrick, Battaglia, & Tenenbaum, 2011; Sanborn, Mansinghka, & Griffiths, 2013; Smith & Vul, 2013; Téglás et al., 2011).

We suggest that a core difference between the above studies is the task given to participants. Some have asked participants to make a single judgment about the future state of the world, for instance, the direction a tower of blocks will fall (Hamrick, et al., 2011) or where a ball will cross a line (Smith & Vul, 2013). In contrast, classic studies tap into explicit explanations of physics, through verbal problems (Anzai & Yokoyama, 1984) or line drawings of motion (McCloskey, et al., 1980). Here we argue that people can apply correct physical principles consistently to simulate the world forward; however, explicit explanations of how the world will unfold draw upon an idiosyncratic set of background knowledge.

We assessed participants' understanding of the movement of balls after they had fallen off of pendulums in three separate tasks: predicting where a ball would land if cut from a pendulum, deciding when to cut a pendulum string such that the ball would fall into a fixed bucket, and drawing the path of the ball after the string is cut. We picked these tasks because there is evidence that people understand the motion of pendulums (Pittenger, 1985, 1990) and can predict the motion of projectiles under gravity (Saxberg, 1987), both of which must be combined to determine the ultimate trajectory of the balls. But there is also evidence that people show systematic errors when asked to explicitly draw the path of the ball (Caramazza, McCloskey, & Green, 1981), and that these errors are attenuated with kinematic information (Kaiser, Proffitt, Whelan, & Hecht, 1992).

The same physical principles apply to each of these tasks, and so in the present experiment we investigated whether the tasks that require implicit prediction (catching the ball and cutting the string) can be explained by veridical physical principles. We find that subjects' performance on the catching and cutting tasks differs between the tasks, but in the tasks that involved perceptually guided movements the differences can be reconciled by considering a single, valid model of physics that incorporates the different sources of perceptual and motor uncertainty from each task. Conversely, the sketches based on explicit conceptualization were inconsistent and idiosyncratic.

Experiment

Methods

Fifty-seven UC San Diego undergraduates (with normal or corrected vision) participated in this experiment for course credit. All were treated in accordance with UCSD's IRB protocols.

Procedure

Participants viewed a computer monitor from a distance of approximately 60cm, which initially depicted a ball swinging from a string, consistent with pendulum motion. At some point in time the string would be cut and the ball would be released, thus entering ballistic motion. Beneath the pendulum there was always a bucket, and in every trial the participant's goal was to cause the ball to drop into the bucket after being released. How they were allowed to interact with the scene differed between two tasks, which were organized into blocks that were randomized across participants. With the exception of one initial practice trial per task that familiarized participants with the task, the path of the falling ball was occluded in order to prevent participants from learning a simple relationship between the ball's release position and its landing position. At the end of each trial, participants were given binary feedback that indicated whether or not the ball successfully landed in the bucket. After the two tasks on the computer, participants were asked to draw the ball's motion in a diagram task.

Catching task. Participants were instructed to adjust the bucket's horizontal position using the mouse so that the ball would land in the bucket after being released. The release time was pre-determined and varied across trials. To relieve time pressure placed on participants, at the moment the string was cut, all ball and string movement was paused. Once the participant chose a bucket position, they could unpause the motion by clicking the mouse. The center of the bucket was recorded as the participant's judgment about where the ball would land.

Cutting task. The bucket was held fixed at a pre-determined position and participants were instructed to cut the pendulum string by clicking the mouse at a time that would cause the ball to drop into the bucket. The time at which the string was cut was recorded for each trial.

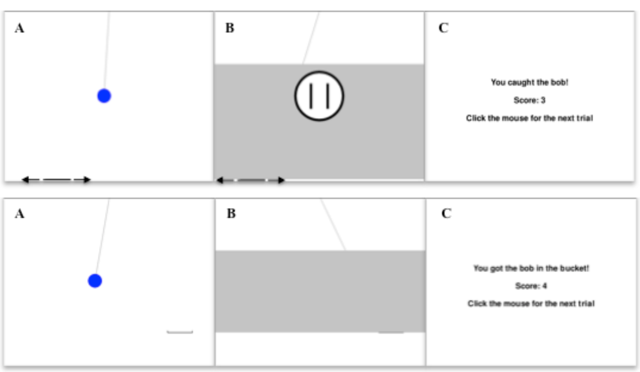

Figure 1: Diagram of the two tasks: catching on top, cutting on the bottom. (A) The pendulum swings freely to start; this ends at a predetermined time (catching) or when the participant clicks the mouse (cutting). (B) An occluder is placed over the string. In the catching task, the action is paused until participants click the mouse, during which time they can move the bucket. In the cutting task, there was no pause, but the falling motion of the ball was occluded. (C) Participants are given feedback on success or failure.

Trials. For each task, participants repeated 48 distinct trials five times each. Trials were matched across tasks such that where the ball landed in a catching trial was the bucket position in the matched cutting trial. In the catching task, there were 16 distinct release times, crossed with three vertical distances between the nadir of the pendulum and position of the bucket – either 20, 35 or 50% of the total screen height. No participant indicated they were aware that the trials were repeated or matched across task in an informal post-experiment survey.[1]

Simulating pendulum motion. Both tasks and all trials used the same pendulum. This pendulum had a length of half of the screen, and reached a maximum angle of 35° from vertical of the nadir. The period of the pendulum was 2.46s. The string was assumed to be massless, and therefore the position of the pendulum at any time could be calculated according to the laws of physics.[2]

Both the pendulum motion and the falling ball obeyed Newtonian mechanics as if the pendulum was positioned at a depth of 6m from the participants. This value was selected through pilot tests to conform to participants' general expectations about the natural period of the pendulum.

Diagram task. After participants completed both tasks, they were given diagrams of pendulums and asked to draw the path of the ball if the string was cut at four positions indicated in those diagrams (a replication of Caramazza, et al., 1981). One participant did not perform this task due to a logistical error.

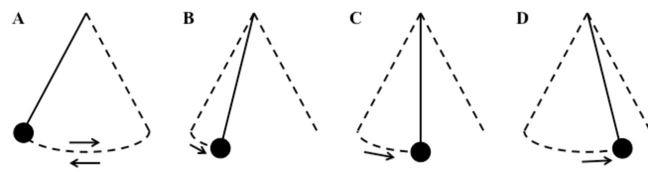

Figure 2: The four problems in the diagram task. Participants were asked to draw the expected path of the ball if the pendulum string were cut at each of the four points.

Results

Accuracy in the catching and cutting tasks was measured as the proportion of trials in which the ball successfully landed in the bucket. Participants' mean accuracies were 30.7% (s.d. 14.1%) on the catching task, and 47.4% (s.d. 15.6%) on the cutting task. Participants' individual accuracies were (Pearson) correlated across tasks, $r = 0.68$. There was no evidence that participants improved over trials on the cutting task ($z=1.23$, $p=0.22$), but they did improve on the catching task ($z=3.04$, $p=0.0024$), from 28.8% accuracy on the first half to 32.8% on the second half.

The remaining analyses quantified participants' performance as the displacement between the ball's landing position and the bucket's position; in the catching task the bucket position was under participants' control and the landing position was under experimental control, and vice versa for the cutting task. We aggregated performance by

[1] One participant noted that they solved trials by "remembering where the ball should go" but it was not clear whether this was memory for the trials or prior knowledge of pendulum motion.

[2] For computational reasons, this was calculated using the small angle approximation to pendulum motion, which should be correct to within 2.4% of actual pendulum timing.

trial across participants in each task to determine how trial factors influenced participants' decisions.

Catching task. Participants' mean bucket positions were correlated with the ball's actual landing positions (*r=0.95, SumSq* = *880**10^3), and were highly consistent with each other (split-half correlation: *r=0.993*). Participants also demonstrated a systematic bias: on average their judgments were slightly shifted away from the actual landing position, toward the center of the pendulum (see Fig. 3). The consistency across participants suggests that the position bias is shared, capturing a commonality in physical models.

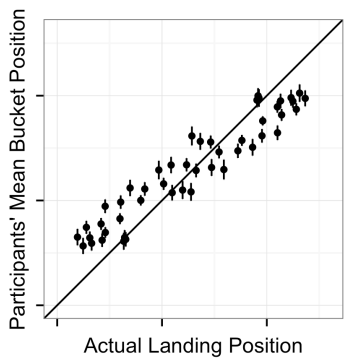

Figure 3: Catching task. Actual landing positions (x-axis) versus participants' mean bucket positions (y-axis) for each trial (individual trials, error bars are 95% CIs).

Cutting task. We calculated the projected landing positions of the ball as a function of each release time chosen by participants, per trial. Participants' mean landing positions were highly correlated with the actual bucket positions (*r=0.98, SumSq* = *187**10^3), and were again highly consistent with each other (split-half correlation: *r=0.998*). Participants also demonstrated a distinct bias, which differed from that in the catching task: when the bucket was near the horizontal position of the pendulum's nadir, participants' mean landing positions were shifted away from it, but when the bucket was far from the nadir, their mean landing positions were shifted toward it (note the sigmoid curvature in Fig. 4). This high inter-participant correlation again suggests a common bias across people.

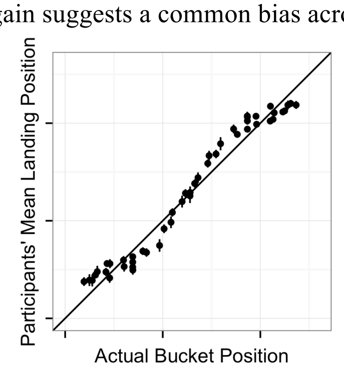

Figure 4: Cutting task. Actual bucket positions (x-axis) versus mean ball landing positions (y-axis) for each trial (individual trials, error bars are 95% CIs).

Comparison. Both tasks required using the same physical principles to determine where the bucket should be placed or when the rope should be cut, yet showed divergent biases. Moreover, the correlation between the mean bucket position and mean landing position for matched trials was high (*r* = *0.93*), but this demonstrates only that participants were in general accurate at this task – the inter-task correlation was less than each task's correlation with the ideal response, suggesting that the sources of deviation from the ideal response are distinct.

Diagram task. Two research assistants naïve to the purpose of this experiment sorted participants' diagram trajectories into one of eight types (see Figure 5). Inter-rater reliability was high (Cohen's kappa = 0.826) – the raters agreed for 47 of the 56 of the participants; where they disagreed, the first author acted as a tie-breaker. Twenty-one percent of the participants' figures were idiosyncratic and could not be categorized. Only 4 (7%) of participants drew the correct path for all diagrams.

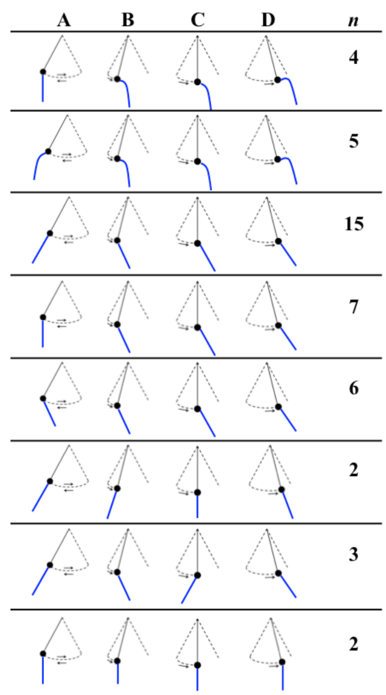

Figure 5: Diagram patterns drawn by more than one participant. Excludes 12 participants who drew idiosyncratic paths. The top pattern represents correct physics.

We reviewed subjects' beliefs about trajectories under gravity: whether they demonstrated that balls would fall in a curved pattern: only 18% of our participants did (less than the 55% reported by Caramazza, et al., 1981). If participants were learning principles about pendulums from the catching or cutting task, we would have expected a higher proportion of curved paths.

Thus participants display high inter-subject reliability on the catching and cutting tasks (despite large differences between the two) but when explicitly drawing pendulum trajectories they show much less agreement and consistency with any kind of physical or non-physical principles. We believe this discrepancy arises because the diagram task taps into idiosyncratic, strategic explanations of physics, but the cutting and catching task behaviors arise from a single consistent application of physical principles under different task demands. We designed a model to test the latter claim.

Physics-based model observer

We designed a model observer that used a single system of physical mechanics rules to predict participants' behavior on both the catching and cutting tasks. These model predictions used real-world physics, just as was used in the experiment to determine the trajectory of the ball both on and off of the pendulum string. The model adapted to each task by adjusting how its physical predictions were applied to the judgment. In the catching task it computed the expected landing position of the ball and selected that as its bucket position, but biased its estimates of the ball's pre-release velocity toward a slower speed based on "misremembering" the velocity through a pause. In the cutting task it computed which release time would cause the ball to land in the bucket and selected that as its judgment, but this timing was subject to errors that reflected realistic constraints on people's timing precision.

Catching task

Description. Because the ball was motionless while participants placed the bucket, participants were required to remember the velocity of the ball and form their judgment based on that memory. This could introduce biases that would cause participants to recall the velocity as slightly different than it had actually been before the pause (Brouwer & Knill, 2009), especially favoring slower speeds (Stocker & Simoncelli, 2006; Weiss, Simoncelli, & Adelson, 2002). This bias was treated as a single parameter (v_{adj}) that determined the proportion of the original velocity the ball would have upon being released. This proportion was constant across all trials.

Based on this (mis)remembered velocity, the model calculated the expected landing position of the ball when it would hit the paddle, and assumed all deviation from that position was Gaussian noise. This placement noise could arise from noise in either the motor system during placement, uncertainty in estimation of the velocity of the ball, or simulation uncertainty that accumulates symmetrically around the position over time (e.g., Smith & Vul, 2013).[3]

Model fit. The model explained participants' average bucket positions well (*r=0.994, SumSq = 41*10*^{3}, see Fig. 6), and accounted for participants' center-shift bias. The model predicted participants' responses as well as participants predicted each other, which suggests that the model captures nearly all of the systematicity in people's underlying judgments.

The best fitting parameters assumed that participants recalled the ball as having 51.7% of its pre-pause velocity magnitude, which caused their judgments of its predicted final horizontal distance to be shifted nearer to the center when it reached the ground. Although this is directionally consistent with our assumption that people remember velocity as slower than it was, the magnitude was larger than expected. Individual errors were predicted to be distributed around that point with a standard deviation equal to 14.5% of the screen width.

Although accuracy increased across trials in the catching task, this had relatively little impact on the model parameters (first half v_{adj}: 47%, second half v_{adj}: 55%). Therefore we do not believe that this pattern of errors was driven by feedback during the task.

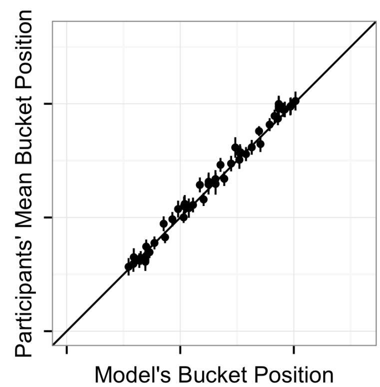

Figure 6: Catching task. Model's bucket positions (x-axis) versus participants' mean bucket positions (y-axis) for each trial (individual points, error bars are 95% CIs).

Uncertainty The model assumed that the error in the catching task arose from Gaussian noise in the bucket position around the expected location. This implies a constant error in paddle position regardless of where the ball lands. Thus error should be constant across trials.

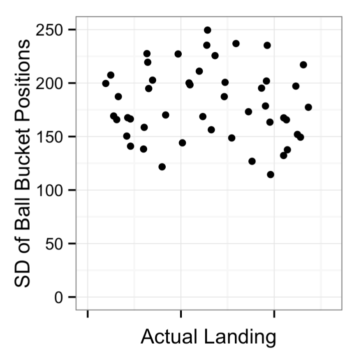

Figure 7: Catching task. Actual landing position (x-axis) versus participants' bucket positions' SD (y-axis) for each trial (individual points).

[3] Simulations indicated that noise in the initial velocity (speed and direction) would give rise to roughly Gaussian error, suggesting that this is a reasonable assumption.

As can be seen in Figure 7, there is no evidence for a linear (*F(1,46)=0.27, p=0.61*) or quadratic (*F(2,45)=1.31, p=0.28*) relationship between the landing position of the ball on each trial and the standard deviation of participants' bucket positions on that trial.[4] This suggests that error does not vary as a function of bucket position, which agrees with our prediction that this is only a combination of motor error and unbiased prediction noise.

Cutting task

Description Participants' release time choices were variable, likely due to imprecise visual estimates of the ball's position and velocity as well as noise inherent to fine motor behaviors. As a result, if the participant intended to release the ball at time t, they may have instead released it at time $t+\varepsilon$. Because the physical dynamics induce a non-linear relationship between ε and the error in landing position, a rational participant should select a time for which the probability of the ball landing in the bucket is highest rather than when it would land closest to the bucket center. If people understand their own timing imprecision (as reported in Hudson, Maloney, & Landy, 2008), then they should marginalize over ε in order to maximize their chance of success. If $R*$ is the intended release time, R is the actual release time, and t_{err} is the variability in timing, the probability of hitting the bucket given $R*$ is:

$$P(hit|R^*, t_{err}) = \int P(hit|R) * P(R|R^*, t_{err})$$

Here $P(hit|R)$ is either 1 or 0, because *hit* depends deterministically on R. The distribution of R given $R*$, $P(R|R*, t_{err})$, was assumed to be Gaussian distribution with mean and SD, $R*$ and t_{err} respectively. The model assumed that people selected $R*$ such that $P(hit|R*)$ was at a local maximum. The cutting task contained an important additional feature: for most trials (58%) there were two time spans in the pendulum period during which the string could be cut to get the ball into the bucket – usually one time while the pendulum is swinging left, and once while swinging right. In these cases, there were two locally maximum modes of $P(hit|R*)$. Puzzlingly, people did not always choose the optimal (higher probability) mode given the model assumptions, but instead often favored the suboptimal mode. This suboptimality may have been due to participants' desire to accumulate more information by waiting for the later time (Battaglia & Schrater, 2007; Faisal & Wolpert, 2009), or minimize trial duration by selecting the earlier time. Since our model did not capture such factors, we simply set the model's choice of modes to match the participants' proportion.

Timing errors were represented by two parameters in this model, describing the bias (t_{bias}) and the noise (t_{err}). These parameters were fit to the observed cut timings, though for consistency, results are presented as the average landing position based on these cuts.

Model fit. The model assumed that people tended to release the ball 38ms after the optimal time, and the variability in responses had a standard deviation of 165ms. This timing variability is similar in magnitude to that reported in another task that required physical prediction (130ms; Faisal & Wolpert, 2009). The correlation between people's mean projected landing position given their choice of release time and that of the model was high ($r=0.993$, $SumSq = 87*10^3$, see Fig. 8).

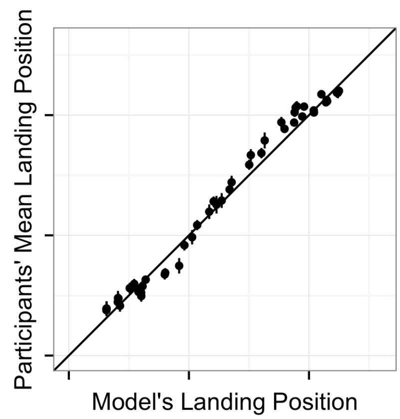

Figure 8: Cutting task. Model's landing position (x-axis) versus participants' mean projected landing positions (y-axis) for each trial (individual points, error bars are 95% CIs).

Uncertainty. The model assumed that the source of error in landing positions was in the cutting time, but a constant error in time does not imply a constant error in landing position: if the ball is released near the apex when moving slowly, a small time error will lead to a small difference in landing position, while if the ball is released at the nadir when moving fastest, the same timing error will lead to a larger difference in landing position.

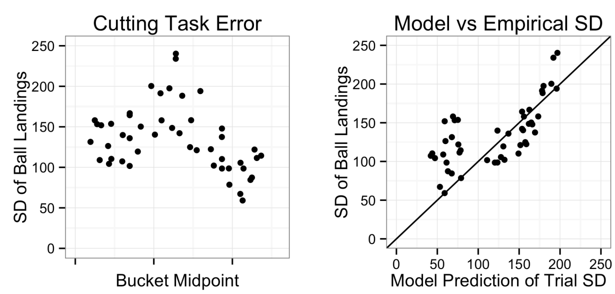

Figure 9: (Left) Variability in empirical ball landings by where the ball will land. (Right) Model predictions of trial variability in the cutting task versus empirical observations. Each point represents a separate trial.

Unlike the catching task, there is a quadratic relationship between the landing position of the bucket on each trial and

[4] We attempted to fit polynomial regressions up to fifth-order to this data but found no significant relationships (all ps > 0.1).

the SD of the ball landing positions on that trial (*F(2,45)=13.8, p<0.001*, see Fig. 9, left). Furthermore, the model predicts this variability. We calculated the SD of landing position that the model expected for each trial and found that it was correlated with participants' projected landing position SD with $r = 0.67$ (see Fig. 9, right), although the model's predicted SD was slightly lower on some trials. This suggests that the physics-based model captures differences in trial variance.

Discussion

In this experiment, we found that people show very different behaviors on three tasks that use the same underlying model of physics: predicting the trajectory of a ball on a pendulum after the string has been cut. Two of the tasks required people to make a judgment about the future state of the world: where the ball will land or when to cut the string to control the ball's landing. While people responded in different ways on each of these two tasks, both sets of responses were consistent with veridical physical principles once task uncertainties were accounted for. On the other hand, participants were much more variable on the diagram task: they often drew trajectories that were physically impossible.

These differences imply that the catching and cutting tasks are tapping a different sort of knowledge than the diagram task. Perhaps people can simulate the world forward in a way consistent with Newtonian physics, but the workings of these simulations are opaque, making description difficult and more reliant on conceptual understandings. This would suggest a need for both types of intuitive physics: research into how people make predictions informs how we use physics to plan our actions or make judgments about the world (e.g., Gerstenberg, Goodman, Lagnado, & Tenenbaum, 2012; Hamrick, et al., 2011), while research into how people describe physical events informs how we form concepts about the workings of the world (e.g., diSessa, 1993).

It has been suggested before that "a person may possess a perceptual appreciation of… natural dynamics… yet be unable to draw upon this knowledge… in a representational context." (Kaiser, Proffitt, & McCloskey, 1985, p. 539). Here we provide evidence that even when people cannot explain how the world will unfold, their predictions and actions are reflective of a veridical physical model of the world.

Acknowledgments

KS and EV were supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of the Interior (DOI) contract D10PC20023. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DOI, or the U.S. Government.

References

Anzai, Y., & Yokoyama, T. (1984). Internal models in physics problem solving. *Cognition and Instruction, 1*(4), 397-450.

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.

Brouwer, A.-M., & Knill, D. C. (2009). Humans use visual and remembered information about object locations to plan pointing movements. *Journal of Vision, 9*(1), 1-19.

Caramazza, A., McCloskey, M., & Green, B. (1981). Naive beliefs in "sophisticated" subjects: Misconceoptions about trajectories of objects. *Cognition, 9*, 117-123.

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction, 10*(2&3), 105-225.

Faisal, A. A., & Wolpert, D. M. (2009). Near optimal combination of sensory and motor uncertainty in time during a naturalistic perception-action task. *Journal of Neurophysiology, 101*, 1901-1912.

Gerstenberg, T., Goodman, N., Lagnado, D. A., & Tenenbaum, J. (2012). *Noisy Newtons: Unifying process and dependency accounts of causal attribution*. Paper presented at the Proceedings of the 34th Annual Meeting of the Cognitive Science Society, Sapporo, Japan.

Hamrick, J., Battaglia, P., & Tenenbaum, J. (2011). *Internal physics models guide probabilistic judgments about object dynamics*. Paper presented at the Proceedings of the 33rd Annual Meeting of the Cognitive Science Society, Boston, MA.

Hudson, T. E., Maloney, L. T., & Landy, M. S. (2008). Optimal compensation for temporal uncertainty in movement planning. *PLoS Computational Biology, 4*(7). doi: doi:10.1371/journal.pcbi.1000130

Kaiser, M. K., Jonides, J., & Alexander, J. (1986). Intuitive reasoning about abstract and familiar physics problems. *Memory & Cognition, 14*(4), 308-312.

Kaiser, M. K., Proffitt, D. R., & McCloskey, M. (1985). The development of beliefs about falling objects. *Attention, Perception, & Psychophysics, 38*(6), 533-539.

Kaiser, M. K., Proffitt, D. R., Whelan, S. M., & Hecht, H. (1992). Influence of animation on dynamical judgments. *Journal of Experimental Psychology: Human Perception and Performance, 18*(3), 669-689.

McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naive beliefs about the motion of objects. *Science, 210*(5), 1139-1141.

Pittenger, J. B. (1985). Estimation of pendulum length from information in motion. *Perception, 14*, 247-256.

Pittenger, J. B. (1990). Detection of violations of the law of pendulum motion: Observers' sensitivity to the relation between period and length. *Ecological Psychology, 2*(1), 55-81.

Proffitt, D. R., & Gilden, D. L. (1989). Understanding natural dynamics. *Journal of Experimental Psychology: Human Perception and Performance, 15*(2), 384-393.

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.

Saxberg, B. V. H. (1987). Projected free fall trajectories II: Human experiments. *Biological Cybernetics, 56*, 177-184.

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

Stocker, A. A., & Simoncelli, E. P. (2006). Noise characteristics and prior expectations in human visual speed perception. *Nature Neuroscience, 9*(4), 578-585.

Téglás, E., Vul, E., Girotto, V., Gonzalez, M., Tenenbaum, J. B., & Bonatti, L. L. (2011). Pure reasoning in 12-month-old infants as probabilistic inference. *Science, 332*, 1054-1059.

Todd, J. T., & Warren, W. H. (1982). Visual perception of relative mass in dynamic events. *Perception, 11*, 325-335.

Weiss, Y., Simoncelli, E. P., & Adelson, E. H. (2002). Motion illusions as optimal percepts. *Nature Neuroscience, 5*(6), 598-604.