WARWICK
THE UNIVERSITY OF WARWICK

Original citation:
Sanborn, Adam N., Mansinghka, Vikash and Griffiths , Thomas L. (2009) A Bayesian framework for modeling intuitive dynamics. In: CogSci 2009: 31st Annual Meeting of the Cognitive Science Society, Amsterdam, Netherlands, 29 Jul - 1 Aug 2009.

Permanent WRAP URL:
http://wrap.warwick.ac.uk/36011

A note on versions:
The version presented here is a working paper or pre-print that may be later published elsewhere. If a published version is known of, the above WRAP URL will contain details on finding it.

For more information, please contact the WRAP Team at: wrap@warwick.ac.uk

warwick.ac.uk/lib-publications

A Bayesian Framework for Modeling Intuitive Dynamics

Adam N. Sanborn (asanborn@gatsby.ucl.ac.uk)
Gatsby Computational Neuroscience Unit, University College London
London, UK

Vikash K. Mansinghka (vkm@mit.edu)
Brain and Cognitive Sciences Department, Massachusetts Institute of Technology
Boston, MA USA

Thomas L. Griffiths (tom_griffiths@berkeley.edu)
Department of Psychology, University of California, Berkeley
Berkeley, CA USA

Abstract

People have strong intuitions about the masses of objects and the causal forces that they exert upon one another. These intuitions have been explored through a variety of tasks, in particular judging the relative masses of objects involved in collisions and evaluating whether one object caused another to move. We present a single framework for explaining two types of judgments that people make about the dynamics of objects, based on Bayesian inference. In this framework, we define a particular model of dynamics – essentially Newtonian physics plus Gaussian noise – which makes predictions about the trajectories of objects following collisions based on their masses. By applying Bayesian inference, it becomes possible to reason from trajectories back to masses, and to reason about whether one object caused another to move. We use this framework to predict human causality judgments using data collected from a mass judgment task.

Keywords: collisions; Bayesian modeling; perception; causality; mass judgments

Following the ground-breaking work of Michotte (1963), the perception of collisions has been studied as a way of determining how people infer unobservable variables from observable variables. In visual perception of collisions, the observer can see the movements of the objects, but needs to infer the hidden properties of the objects (Chaput & Cohen, 2001; Cohen & Ross, in press; Gilden & Proffitt, 1989; Runeson, 1977; Runeson, Juslin, & Olsson, 2000; Schlottmann & Anderson, 1993; Todd & Warren, 1982). There are many possible questions that could be asked of observers. Two of the most commonly used tasks are judging which object is heavier and judging whether a collision occurred.

Judgments of the ratio of masses of colliding objects have motivated arguments about the perceptual invariants. In a mass judgment task, subjects are presented with two objects colliding on a screen and are asked to choose which object has greater mass. People make characteristic patterns of errors, which have led researchers to propose that human mass ratio judgments are based on heuristics (e.g., Cohen & Ross, in press; Gilden & Proffitt, 1989; Todd & Warren, 1982), though other researchers argue that the correct mass ratios are computed for experienced observers (Runeson, 1977; Runeson et al., 2000). Michotte (1963) has directly motivated a second area of research. In this similar task, subjects had to determine whether two objects had collided or whether the objects were moving independently. Studies in this area have collected data on how changes in physical variables relate to the judgments of causality (e.g., Chaput & Cohen, 2001; Spelke, 1994; Schlottmann & Anderson, 1993).

Despite the similar tasks used in mass judgments and judgments of causality, these two lines of research have proceeded almost always independently (but see Kaiser & Proffitt, 1984). Our goal is to present a unifying explanation for these two types of judgments. We use a simple probabilistic model of physical dynamics that makes the same assumptions about the probability with which different physical properties of objects are likely to be observed, but differs in the hypotheses to be evaluated and the data that are used. Unlike previous models of mass judgments, we assume that people are not using heuristics, but instead are making their decisions based on the underlying physical laws. Using these laws, we assume subjects are making the optimal decision based on noisy observations of the velocities of the objects.

The plan of this paper is as follows. First we describe the mass and causality judgment tasks, along with the physical laws that dictate the optimal decisions in these tasks. Next we generalize the optimal decisions to a more realistic situation in which the observed velocities are perceived with noise. Our generative probabilistic model applies to both mass and causality tasks. Next we replicate two venerable experiments: one for mass judgments and one for causality judgments. We fit our model to the data from the mass judgment task, and show that we can accurately mimic human performance in this task. In addition, using the data from the mass judgment experiment, we successfully predict how subjects will respond in the causality judgment task. Finally, we discuss how our model relates to other work in these two areas.

Task Descriptions and Physical Laws

The mass judgment and causality judgment tasks are similar experimental paradigms, with a few key differences. In this section we describe the variables in the two tasks and what a classical model of mechanics would predict.
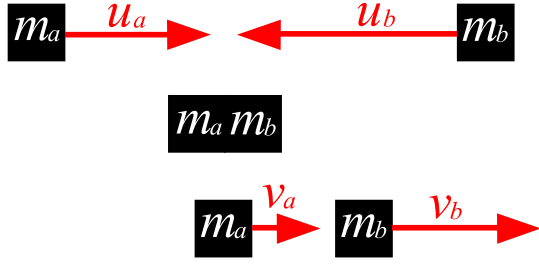
Figure 1: Stages of a mass judgment movie. In the first row, two objects have initial velocities u_a and u_b and masses m_a and m_b. The second row shows the collision and the elasticity e. The final velocities v_a and v_b are shown in the last row.

Mass Judgments

The first task we will examine is the mass judgment task (e.g., Cohen & Ross, in press; Gilden & Proffitt, 1989; Runeson, 1977; Runeson et al., 2000; Todd & Warren, 1982). In this task, as shown in Figure 1, two objects with masses m_a and m_b move with initial velocities u_a and u_b. They participate in a collision with coefficient of elasticity e, and then move apart with final velocities v_a and v_b. The subject's task is to infer which mass m_a or m_b is greater.

Using Newtonian physics, the observable initial and final velocities of the objects, and the elasticity of the collision (i.e., coefficient of restitution), the final velocities can be determined,

$$v_a = \frac{m_a u_a + m_b(u_b + e(u_b - u_a))}{m_a + m_b} \qquad (1)$$

$$v_b = \frac{m_b u_b + m_a(u_a + e(u_a - u_b))}{m_a + m_b} \qquad (2)$$

Elasticity is the ratio of velocities before and after impact which can be calculated from the initial and final velocities,

$$e = \frac{v_a - v_b}{u_b - u_a} \qquad (3)$$

The mass ratio, which is the question that we pose to participants, can also be calculated from the initial and final velocities,

$$\frac{m_a}{m_b} = \frac{u_b - v_b}{v_a - u_a} \qquad (4)$$

This result suggests a way for participants to evaluate a mass ratio from velocities. However, it has been shown that human mass ratio judgments are affected by elasticity (Runeson, 1977; Todd & Warren, 1982), while Equation 4 shows that the judgment based on Newtonian physics is independent of elasticity.

Causality Judgments

The second class of question is a judgment of whether a one object causes another to move. In this task, as shown in Figure 2, one object moves towards another with equal mass at initial velocity u_a. It stops a distance of g from the other
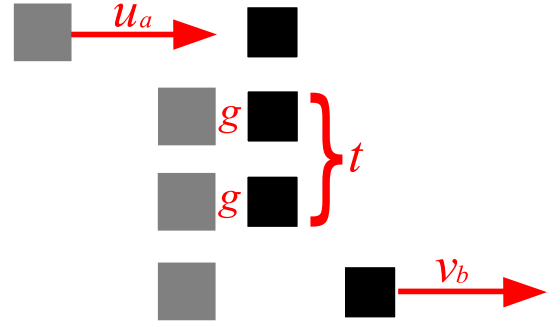

Figure 2: Stages of a causality judgment movie. In the first row, the gray square is moving with initial velocity u_a and the black square is stationary. The second and third rows show the gap g between squares after the first square stops and the time delay t between the first square stopping and the second square starting to move. The final row shows the gray square stationary and the black square moving with final velocity v_b

object, and after a delay of t, the second object moves with final velocity v_b. The task is to decide whether the first object caused the movement of the second, or whether the second object moved on its own.

This kind of task was famously studied by Michotte (1963), and later by other researchers (Chaput & Cohen, 2001; Schlottmann & Anderson, 1993). The masses of the two objects are usually taken to be equal (ie. a mass ratio of 1), but the paradigm introduces two new variables relevant to assessing causality: the gap between objects at the collision g and the interval after the first object stops before the objects begin to move t.

A Bayesian Framework

Bayesian inference can provide a unifying framework for understanding human judgments in both of these tasks if we view them as inferences to hypotheses from perceptual data (Yuille & Kersten, 2006). Specifically, if we can define a probabilistic model that specifies the relationships among these variables, the two tasks can be viewed as involving inferences about one variable based on another. We use the Bayesian network in Figure 3. The final velocities v_a and v_b follow distributions with modes given by Newtonian dynamics (Equations 1 and 2), meaning that they depend on u_a, u_b, m_a, m_b, and e. We chose to use Gaussian distributions to reflect perceptual variability with a mean parameter equal to the Newtonian dynamics and fixed variance parameter σ^2. Elasticity can vary between 0 and 1, and we chose a uniform prior distribution over elasticity for simplicity. The masses m_a and m_b are bounded below by zero and we chose an exponential with decay parameter λ to reflect the intuition that we rarely see extremely massive objects colliding. As with the final velocities, the initial velocities are distributed as a Gaussian distributions, with mean parameter set to zero and the variance parameter was set to $\frac{c}{m}$, meaning that more massive objects are expected to move more slowly, scaled by parameters c.
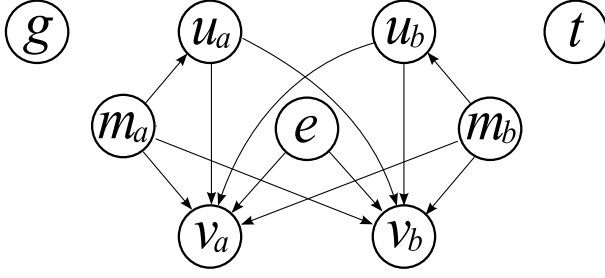
Figure 3: Graphical model for both mass and causality judgments. The initial velocities u_a and u_b and final velocities v_a and v_b depend on the object masses m_a and m_b. The final velocities also depend on the initial velocities and the elasticity e. The gap g and time delay t are independent of the other variables.

The mass and causality judgment tasks require different inferences from our statistical model. For the mass judgment, there is no gap and no delay before the objects move after the collision. These variables can be removed from the full model. The data are u_a, u_b, v_a, and v_b, and the hypotheses are that $m_a > m_b$ or $m_b > m_a$. Conditioning on the observed variables, we can calculate $P(m_a > m_b | u_a, u_b, v_a, v_b)$ by summing out the possible values of e, m_a, and m_b.

For the causality judgment, our Bayesian framework is applied differently. Our data are u_a, v_b, g, and t, and our hypotheses are that the event was or was not a Newtonian collision. These two hypotheses correspond to different probabilistic models, which we term the causality model and coincidence model. For the causality model, v_b depends on u_a and g and t are close to zero. For the coincidence model, g, t, and u_a have the same distribution as before, and as v_b is now independent of u_a, it is distributed identically to u_a. The probabilities of the observed variables under the two model are calculated and one of the models is selected as the response.

The unifying framework is able to predict data in both the mass judgment and causality judgment experiments. Studies of these two phenomena have collected data in different labs using different displays and different dependent measures. In our experiments, we gather data from our model using a similar display for each task in which we collect all-or-none judgments for relative mass (Experiment 1) and causality (Experiment 2).

Experiment 1: Judgments of Relative Mass

Our first experiment replicates Experiment 1 of Todd and Warren (1982) and collects judgments of relative mass under a variety of conditions, manipulating the mass ratio and the elasticities of the collision displays.

Methods

Twenty four participants were recruited from a university community for this study. Four participants were discarded due to a computer error. Each participant was paid $4 for less

than one hour of participation. The eyes of the participants were situated approximately 44 cm away from the display.

Participants were presented with movies of two white squares colliding with each other along one dimension. They were told that these squares were blocks sliding along an invisible smooth surface and colliding with one another. Participants were instructed to press a key corresponding to whichever block they thought was heavier. The two white squares with 1 cm sides started outside the visible area of the screen and moved toward each other at their initial velocities. The two objects moved toward each other at their initial velocities until the edges of the two squares touched at the center of the screen. Following the collision, the squares immediately moved away from each other at their final velocities. The trial ended automatically as soon as the faster object reached the edge of the visible display, but participants could end the trial at any point by responding. No feedback was given to participants during the experiment.

A total of 252 trials were presented to each participant. There were twelve combinations of mass ratios and elasticities. One example of each combination was used presented to participants (order randomized for each participant) at the beginning of the experiment to become acclimated to the display. The data from these practice trials were not included in the analysis. The test trials consisted of 20 replications of each combination of mass ratio and elasticity with the order of presentation was randomized for each participant. The mass ratios were set to be 1.25, 1.5, 2.0, or 3.0. On each trial, the heavier object was set to be the right or left object with equal probability. The elasticities used were 0.9, 0.5, and 0.1. The initial velocities of the left square ranged from 1.91 to 4.45 cm/s in steps of 0.13 cm/s. The initial velocity of the right square was determined by the initial velocity of the left square using the formula, $u_b = u_a - 6.35$ cm/s. The initial velocity of each trial was drawn uniformly from the set of initial velocities. Given these variables and Equations 4 and 3, the final velocities of the two objects are uniquely determined.

Results and Discussion

The average accuracy over participants for each of the combinations of mass ratio and elasticity is shown in Figure 4. The pattern of results show better accuracy with larger mass ratios and greater elasticity. An ANOVA showed a main effect of mass ratio, $F(3, 19) = 155, p < 0.001$, and a main effect of elasticity, $F(2, 19) = 16.0, p < 0.001$. However, the interaction between mass ratio and elasticity was only marginally significant, $F(6, 19) = 1.89, p < 0.1$.

The model presented above for mass judgments was fit to the data from this experiment. The best-fitting parameters were $\lambda = 0.647, c = 4.999$, and $\sigma^2 = 0.238$. These parameters make the predictions shown in Figure 4, which has a high correlation with the human data ($r = 0.98$).

Experiment 2: Judgments of Causality

Our second experiment tackles a different inference, causality, while manipulating a similar set of physical variables.
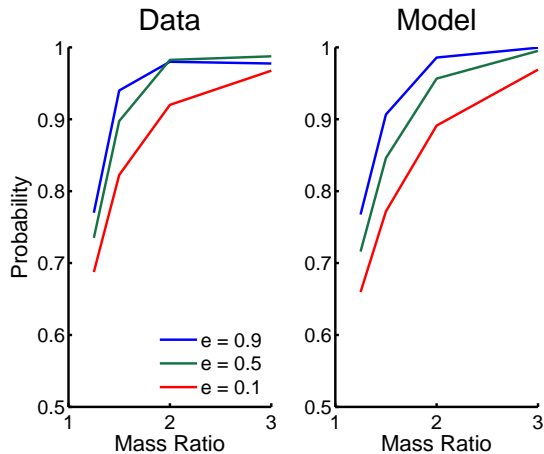
Figure 4: Data and modeling results for the mass ratio judgment task. The horizontal axis is the mass ratios of the collisions shown to subjects and the vertical axis is subject accuracy. The separate lines correspond to different elasticities in the collisions.

This experiment is similar to that of Schlottmann and Anderson (1993) as it manipulates the same variables and asks for causality judgments. The key differences are that we collected all-or-none judgments instead of ratings, and presented displays in which we could examine the effects of assumptions of elasticity on the judgments.

Methods

Fifteen participants were run in each of two instruction conditions: *real or random* and *causality*. The same equipment and viewing distance were used in this experiment as in the previous experiment.

In addition to being told that the blocks were sliding along an invisible smooth surface, participants were instructed that each of the blocks were made of the same material and had the same mass. Following these generic instructions, instructions specific to each condition were given to participants. In the *real or random* condition the additional instructions were,

> Your task is to decide whether each movie came from a real collision of the blocks or a random combination of the variables. A real collision looks like the blocks actually collide. A random collision looks a little like a real collision, except that the velocities of the blocks, gap between the blocks, and the time delay before the second block starts moving are all selected randomly. Remember, both blocks always have the same mass.

Following these instructions, *real or random* participants were shown the boundaries of each of the variables. These instructions were meant to convince participants to use a uniform distribution over each of these variables as their alternative distribution. After each movie, participants responded by keypress as to whether the trial was a real collision or was drawn from the random distribution.

In the *causality* condition, participants were only instructed to decide whether the gray block caused the white block to move or whether the white block moved by itself. After each movie, participants in this condition were asked, "Did it look like the white box moved because the gray box hit it? Was the white box's movement produced by the gray box? Or did the white box take off on its own?"

Three hundred trials were presented to each participant, with half drawn from a near-Newtonian collision distribution and half drawn from a distributions with heavier tails. The trials were drawn in this way to make the instructions seem more plausible. The velocities ranged from 6 cm/s to 15 cm/s. The gap ranged from 0.1 mm to 4 mm, and the time delay ranged from 0 to 250 ms. All samples that fell outside the bounds of the variables were resampled. Each movie began with the fading in of a central white block and a gray block positioned 6.75 cm left of center. The sequence of events in the trials in this experiment are described above.

Results and Discussion

The correlation between responses in the *real or random* condition and the *causality* condition was computed after first binning the trials. The ranges of all four manipulated variables (gap, time delay, initial velocity, and final velocity) were divided into two equal-sized bins. The bins were crossed between variables and the percentage of trials judged to be a collision was computed for each bin, yielding sixteen values for each condition. The correlation between the *real or random* condition and the *causality* condition was $r = 0.94$, showing that there was good agreement between the conditions. The data from these two conditions are aggregated in the remaining analyses.

Best fits were computed for the gap and time delay variables. These variables were fit separately because they are independent from the velocities in both the collision and coincidence models. The collision model represents these two variables as independent zero-mean Gaussian distributions. The coincidence model used a uniform distribution over the range of gaps and time delays shown in this experiment. For each variable, the standard deviations of the Gaussian collision distribution were fit to the data. The standard deviation of the gap was 0.0014 m. The standard deviation of the time delay was 0.0825 s.

The prior assumptions about velocities identified by estimating the parameters of the model can be tested using the collision judgments made in trials with different initial and final velocities in this experiment. We choose to analyze final minus initial velocity, because judgments of causality are very dependent on this difference. Subject trials were generated from a non-uniform distribution to enhance the believability of the instructions, so the data were first binned to remove this influence. Twenty-five equal-sized bins were created that spanned the range of final minus initial velocity. The percentage of collision responses in each bin was computed and is displayed in Figure 5. The probability of choosing collision over coincidence peaks around zero and decreases as the difference between velocities grows. Interestingly, the probability of choosing a collision decreases asymmetrically.

When the initial velocity is larger than the final velocity, differences produce a higher probability of collision choice than when final velocity is larger than initial velocity. This effect was not found in Schlottmann and Anderson (1993), because no trials were tested in which the final velocity was greater than the initial velocity and they did not instruct subjects that the objects had equal mass. Our model does predict this effect, for reasons described below.

We made predictions on this task using the best-fitting parameters from Experiment 1. Carrying over the parameters from the first experiment, we predicted collision probabilities from both the causal model and the coincidence model of velocities that were specified above. Collision responses were assumed to match the posterior probability that the observed event was generated by the collision model, taking the two models to have equal prior probability.

Predictions from our model are shown in Figure 5. Our model predicts the same peak height, the same width at the peak, and a very similar peak location. Interestingly, our model matches human data in predicting an asymmetry for for positive and negative values of initial minus final velocity. This pattern was consistent for both the data and model across a variety of bin sizes. The asymmetries in the model are due to the model using a range of elasticities. Using a value of $e = 1$, the final velocity is equal to the initial velocity, but when $e < 1$, the final velocity is less than the initial velocity. Stated in terms of final minus initial velocity, $e = 1$ produces no difference, while every other value produces a negative difference. As the prior has mass over all elasticity values, the model predicts a negative skew in the choice function. The match between the model predictions and human data suggests that subjects also are considering a range of elasticities.

Where the model and the data differ is when there is a sizeable difference between final and initial velocity. The data show that subjects are much more willing to make a collision response for these velocity differences than the model is. The extra spread that the data show compared to the model may be a result of the difference in velocity ranges used in the experiments. Experiment 1 used a velocity range of 1.91 to 4.45 cm/s, while Experiment 2 used a velocity range of 6 to 15 cm/s. As velocity increases, the noise in the perceived velocity increases as well (Snowden & Braddick, 1991). It is reasonable to expect then, that the tails of the model would be wider if we more accurately accounted for noise in velocity perception.

General Discussion

Mass-ratio and causality judgments have been studied mostly independently, despite using similar stimulus displays. Our goal was to model these two types of judgments under a single framework and we were able to predict several interesting features of collision judgments using the data collected in a separate mass judgment experiment. Our model assumed that the observed velocities in the display were perceived noisily,

and calculated the optimal decision using noisy perception and classical mechanics. Now we turn to comparisons of our model versus existing models for each of these two types of judgments.

Todd and Warren (1982) proposed that people used a perceptual heuristic to make mass judgments. The heuristic they used was, $m_a > m_b \iff v_b > u_a$, meaning that if the velocity after collision of Object B was greater than the velocity after collision of Object A, than Object A had greater mass. Gilden and Proffitt (1989) extended this model to two dimensions and introduced an additional heuristic that subjects used to make judgments of mass ratios, that balls that ricochet more are lighter. Subjects were assumed to switch between heuristics based on the salience of the information in each. These heuristics explain mass judgments well, but are unable to make predictions for collision judgments. Using heuristics for a particular task provides a discriminative model contrast to our generative model of collisions. As these heuristics only model one particular task, and this task only approximately, then they are unable to generalize to new types of tasks.

Other models of causality judgments are very different from our model. These models were generally applied to explain whether the perception of causality is innate or inferred from other variables (e.g., Chaput & Cohen, 2001; Schlottmann & Anderson, 1993). Though we infer the causality of the collision, our model does not bear on this discussion, as even innate perception of causality is assumed to be limited by perceptual noise Michotte (1963) . The information integration approach was successful at describing ratings of causality by computing a weighted average of the cues. For two cues a and b, the model would be,

$$r = \frac{w_{a,x}\psi_{a,x} + w_{b,y}\psi_{b,y} + w_0\psi_0}{w_{a,x} + w_{b,y} + w_0} \qquad (5)$$

where w_C is the weight for cue C, $\psi_{C,\ell}$ is the value of cue C at level ℓ. The configural cue is represented by 0. Though the information integration approach was only applied to causality judgments by Schlottmann and Anderson (1993), it could be easily extended to fit our mass ratio judgment data by using the mass ratio m_a/m_b and elasticity e as cues. The main difference between our model and the information integration approach is that we give an explanation of the how the physical variables work together, while the information integration approach relies on a general-purpose method of estimating the contribution of different cues to decisions which does not result in generalization to other tasks. So for the mass ratio judgment data, it would require 14 parameters to fit the data, but these parameters would not allow it to make any predictions in the causality judgment task.

While our model makes accurate predictions using the correct model of classical mechanics in our tasks, it seems unlikely that this approach could be expanded to any physical situation. More complex situations, such as the two-dimensional collisions of Gilden and Proffitt (1989) and the common misconception that heavier objects should fall faster
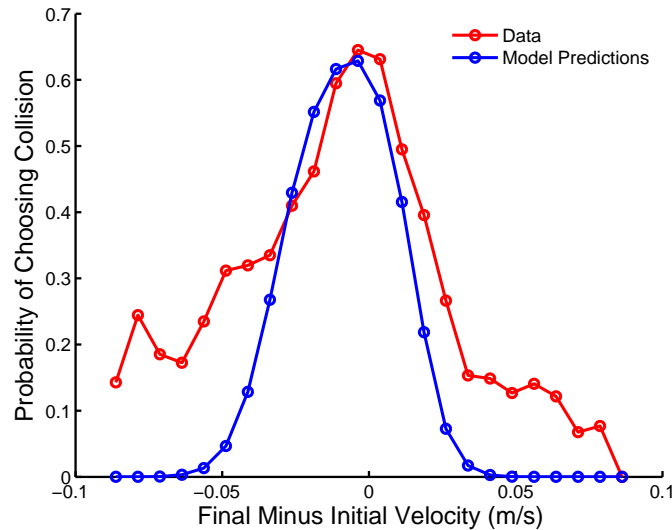
Figure 5: Human data and model predictions in the causality experiment. The final velocity of Object B minus the initial velocity of Object A lies along the horizontal axis. The vertical axis is the probability of choosing a collision over the alternative with data aggregated over subjects and conditions. The red points are the binned data, the blue points are the parameter-free model predictions.

argue against an accurate physical model in all cases. In these situations, it may be that we make optimal judgments in the presence of noise while using the wrong physical model. Understanding exactly when and why our cognitive judgements are inconsistent with physical law and rational inference remains an important challenge for future work.

Conclusion

We modeled human perceptual judgments in a one-dimensional collision task using a model of classical mechanics that was generalized to include noise in the observations. We collected data in two tasks, one for judgments of mass and the other for judgments of causality. Previously, these tasks have been modeled in very different ways and by assuming that people are using heuristics. Our model makes predictions across tasks in an optimal way given the noise in observations. We accurately predicted human data in the mass judgment task, and used our fits to these data to make accurate, parameter-free predictions in the causality task.

Acknowledgments

References

Chaput, H., & Cohen, L. (2001). A model of infant causal perception and its development. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the 23rd Annual Conference of the Cognitive Science Society* (p. 182-187). Mahwah, NJ: Erlbaum.

Cohen, A., & Ross, M. (in press). Exploring mass perception with Markov chain Monte Carlo. *Journal of Experimental Psychology: Human Perception and Performance*.

Gilden, D., & Proffitt, D. (1989). Understanding collision dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 372-383.

Kaiser, M., & Proffitt, D. (1984). The development of sensitivity to dynamically relevant causal information. *Child Development*, *55*, 1614-1624.

Michotte, A. (1963). *The perception of causality*. New York: Basic Books.

Runeson, S. (1977). *On visual perception of dynamic events*. Unpublished doctoral dissertation, University of Uppsala, Uppsala, Sweden.

Runeson, S., Juslin, P., & Olsson, H. (2000). Visual perception of dynamic properties: cue heuristics versus direct-perceptual competence. *Psychological Review*, *107*, 525-555.

Schlottmann, A., & Anderson, N. (1993). An information integration approach to phenomenal causality. *Memory and Cognition*, *21*, 785-801.

Snowden, R., & Braddick, O. (1991). The temporal integration and resolution of velocity signals. *Vision Research*, *31*, 907-914.

Spelke, E. (1994). Initial knowledge: six suggestions. *Cognition*, *50*, 431-445.

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

Yuille, A., & Kersten, D. (2006). Vision as Bayesian inference: analysis by synthesis? *TRENDS in Cognitive Sciences*, *10*, 301-308.