Running head: EXPLORING MASS PERCEPTION

Exploring Mass Perception with Markov Chain Monte Carlo

Andrew L. Cohen & Michael G. Ross

University of Massachusetts, Amherst

Draft 10/13/08 – please do not quote

Address correspondence to:

Andrew L. Cohen

Department of Psychology

University of Massachusetts

Amherst, MA 01002

413-545-4714

FAX 413-545-0996

acohen@psych.umass.edu

Copyright 2008 by the American Psychological Association, Inc.
This article may not exactly replicate the final version published in the APA journal.
It is not the copy of record.The final version will be published in the Journal of Experimental Psychology: Human Perception and Performance (http://www.apa.org/journals/xhp/).

Abstract

A number of previous studies have examined the ability to judge the relative mass of objects in idealized collisions. Using a newly developed technique of psychological Markov chain Monte Carlo sampling (Sanborn & Griffiths, 2008), this work explores participants' perceptions of different collision mass ratios. The results reveal inter-participant differences and a qualitative distinction between the perception of 1:1 and 1:2 ratios. The results strongly suggest that participants' perceptions of 1:1 collisions are described by simple heuristics. The evidence for 1:2 collisions favors heuristic perception models that are sensitive to the sign but not the magnitude of perceived mass differences.

Exploring Mass Perception with Markov Chain Monte Carlo

Observers have some ability to make relative mass judgments after viewing two-ball collisions. On a typical trial of a colliding balls experiment, two (computer-simulated) balls roll across a flat surface, collide, and then roll away from each other. A typical observer's task is either to determine which of the two balls is heavier or to make a quantitative estimate of mass ratio. Observers can become quite adept at determining the relative mass of the two balls (Cohen, 2006; Jacobs, Michaels, & Runeson, 2000; Jacobs, Runeson, & Michaels, 2001; Runeson, Juslin, & Olsson, 2000). Unlike the velocities and angles of the balls, the masses of the balls cannot be observed visually. The central research issue is how people use the visually presented information to determine the relative mass of the balls.

There have been two main classes of relative mass judgment models. The first type of model assumes that observers have access to a correct model for determining relative mass. The direct perception approach, for example, is based on the work of Gibson (1966) and rests on the assumption that the visual field contains potentially complex information that reliably specifies properties of the environment that are central to the guidance of action. With some experience, organisms are assumed to be able to detect and use this complex information directly, without the need for cognitive enhancement (Runeson, 1977). Because the relative mass of colliding balls can be calculated from the velocities of the two balls, and because knowledge of object masses is fundamental to the ability to interact with the environment, researchers who have taken the direct perception approach assume that observers can directly detect this value. (Also see Sanborn, Mansinghka, & Griffiths, 2006, for another model that takes this approach.)

The second type of mass perception model explains observer's mass judgments as the

result of high-level reasoning. The heuristic theorists (Gilden & Proffitt, 1989, 1994; Proffitt & Gilden, 1989; Todd & Warren, 1982) claim that the perception of properties such as relative mass is primarily based on inference. They claim that observers can use only simple motion cues that must be supplemented with high-level cognitive processes that realize their beliefs about the physical world. These processes are usually assumed to act as a decision-making system using heuristics. For example, in the case of the colliding balls task, Gilden and Proffit (1989) suggested that observers base their mass judgments on the speeds and angles of the balls. They suggested that observers judged a ball as lighter if it was either moving faster after the collision or reflected at a greater angle.

Because it addresses both what information can be perceived and what decision processes act upon this information, the debate between these two camps is an important one. Unfortunately, despite two decades of debate, these theories have proven difficult to distinguish empirically (e.g., Hecht, 1996) with some research suggesting that people have good models of mass perception (e.g., Runeson, et al, 2000) and other research concluding the opposite (e.g., Gilden & Proffitt, 1989). The primary goal of this paper is to use a recently developed empirical technique, psychological Markov chain Monte Carlo (Sanborn & Griffiths, 2008), to compare specific, testable implementations of the direct perception and heuristic theories. A secondary goal is to use this new technique to establish a broad range of data that any successful mass perception model needs to explain.

As a motivation for the use of psychological Markov chain Monte Carlo (discussed in the next section), note that any method for judging collision mass ratios can be described by a probability distribution indicating the frequency with which a particular collision will evoke a particular response. For example, the response of a human observer to the question “which ball is

heavier?" (e.g., Runeson, et al., 2000) in a collision experiment can be described by

$$P(\text{Perceived mass ratio} > 1 \mid \text{Collision}), \tag{1}$$

where

$$\text{Perceived mass ratio} = \frac{\text{Perceived mass of Ball 1}}{\text{Perceived mass of Ball 2}}. \tag{2}$$

Note that this distribution, like all others in this paper, describes the probability of a human observer having a particular, not necessarily veridical, perception.[1] The value of (1) indicates the probability of responding "Ball 1 is heavier" for any given collision. The probability of responding "Ball 2 is heavier" is $1 - P(\text{Perceived mass ratio} > 1 \mid \text{Collision})$.

A similar approach is possible for related tasks. For example, the observer's behavior when asked to choose a mass ratio from among a set of mass ratios (e.g., Jacobs, et al, 2001) could be described by

$$P(\text{Perceived mass ratio} = k \mid \text{Collision}), \tag{3}$$

where k takes on the values of the mass ratio choices.

Both (1) and (3), and any similar distribution of perceived mass ratio, can be computed from the more general

$$P(\text{Perceived mass ratio} \mid \text{Collision}) \tag{4}$$

probability density function. It is important to note that (4) is atheoretical; it is not assumed that a human observer uses (4) to calculate the probability of a mass ratio given a collision. Rather, regardless of the processes used, (4) provides a useful description of the output of these processes; accurate knowledge of this function would allow perfect simulation of the observer's behavior. For this reason, it would be very useful to conduct an experiment that enabled us to estimate this distribution for human subjects.

Both because the space of collisions is large and because an observer's conscious access

to the probabilities of (4) is questionable, it may not be possible to directly estimate this distribution. Basic probability theory, however, indicates a related function that is both highly informative and can be practically estimated from experimental data. According to Bayes' rule,

$$P(\text{Perceived mass ratio} \mid \text{Collision}) \propto P(\text{Collision} \mid \text{Perceived mass ratio}) P(\text{Perceived mass ratio}). \tag{5}$$

It can be seen on the right side of (5) that if the prior probability that an observer will perceive a particular mass ratio,

$$P(\text{Perceived mass ratio}), \tag{6}$$

were known, knowledge of the likelihood function of the mass ratio,

$$P(\text{Collision} \mid \text{Perceived mass ratio}) \tag{7}$$

would allow prediction of an observer's perceived mass ratio for any collision, i.e., (4). The likelihood function (7) is especially informative. Given a particular mass ratio, this equation indicates the collisions an observer expects to see. Although it is difficult to recover the distribution described by (4), surprisingly, it is possible to accurately and efficiently estimate the probability density in (7). The remainder of the paper focuses on estimating this quantity.[2]

If (7) were mapped out, it would reveal much about how observers make relative mass judgments. Knowing (7) would, for example, reveal if an observer's judgments were in line with the veridical mass ratios. An accurate observer would place most probability mass on collisions that could be physically generated for a given mass ratio, while an inaccurate observer would either place substantial probability mass elsewhere or would place little mass in some appropriate regions. By comparing the distributions produced by (7) for different mass ratios, it would also be possible to determine if a consistent set of processes drove the observer's perceptions or if they varied with mass ratio.

As discussed previously, it is not always practical to determine (7) directly. If it were

possible to draw enough samples from this distribution, however, most quantities of interest could be estimated: means, variances, and other moments could be computed; the distribution could be approximated and plotted; and regions of high and low probability could be determined. The challenge is to draw samples from the distribution without requiring the observer to directly provide them. The next section describes psychological Markov chain Monte Carlo sampling, a technique that can sample from an observer's collision probability distribution.

Markov Chain Monte Carlo Sampling

The goal of Markov chain Monte Carlo (MCMC) sampling (Hastings, 1970; Metropolis, Rosenbluth, Rosenbluth, Teller, & Teller, 1953) is to efficiently produce samples from a probability distribution from which it is difficult to sample directly. In physics and artificial intelligence, the most common MCMC application is drawing samples from a high-dimensional distribution that cannot be practically normalized. In the current context, MCMC will be used to draw samples from, and in doing so, estimate, the distribution defined by (7).[3]

An MCMC algorithm produces samples from a *target distribution* $P(x)$ by specifying a *proposal distribution* $Q(y \mid x_t)$ and an *acceptance rule* $A(y, x_t)$. Q is selected to be an easily sampled distribution, and A is a function that returns one of its two arguments. The basic MCMC algorithm begins with an initial, randomly chosen sample x_0 and with $t = 0$, and then repeats the following steps:

1. Draw a new sample y from distribution $Q(y \mid x_t)$.
2. Set $x_{t+1} = A(y, x_t)$, i.e., use function A to choose between the current sample and the last accepted sample.

3. Set $t = t + 1$ and repeat Step 1.

If Q and A have been constructed according to the principles of MCMC, the distribution of x_t will approach $P(x_t)$ as $t \to \infty$.

Implementing MCMC requires the selection of an appropriate A and Q. These functions interact to guarantee that the distribution of the samples approach P. Barker's method (Hastings, 1970) provides one convenient set of constraints for specifying A and Q. The requirements of Barker's method are satisfied if $Q(y \mid x_t) = Q(x_t \mid y)$; $P(y) > 0$ implies that $Q(y \mid x_t) > 0$; and A accepts a proposed sample y with probability $\frac{P(y)}{P(y) + P(x_t)}$.

Sanborn and Griffiths (2008) realized that Barker's acceptance rule corresponds closely to the Luce choice rule (Luce, 1959), a widely used cognitive model of human choice. Furthermore, they noted that the target distribution P only appeared in the acceptance rule A and imposed very weak requirements on the proposal distribution Q. They demonstrated that if a human observer accepts and rejects proposed samples according to a Luce choice rule, the accepted samples can be treated as samples from the observer's distribution. In other words, their key insight is that a description (P) of an observer's category model can be recovered by allowing the human observer to accept or reject samples within the MCMC algorithm. Here we use the term "model" to refer to an observer's category representation and the decision processes that act upon that representation. Sanborn & Griffiths (2008) have demonstrated that this psychological Markov chain Monte Carlo (pMCMC) technique can discover probability distributions that represent humans' judgments of category membership for simple animal drawings and stick figures. In the experiments described below, pMCMC is used to estimate the target distribution defined by (7).

From a psychological point of view, pMCMC offers many advantages. First, pMCMC

does not rely on the participant providing explicit ratings to indicate the importance of different perceptual factors to category membership. The resulting distribution is the one implied by the observer's choices, and choice between alternatives is arguably a more natural task than providing numerical ratings. Second, pMCMC is a very efficient approach for collecting data. In particular, pMCMC produces a useful sample for almost every iteration of the algorithm. Third, the only assumption made is that observers are employing a Luce choice rule and that enough samples are gathered to ensure asymptotic behavior. There is no assumption about the functional form of P. Therefore, the samples can be used to directly test different hypotheses about observers' models, which are difficult to distinguish using indirect behavioral experiments.

Because MCMC is an asymptotic algorithm, it can be difficult to determine when enough iterations have run to guarantee that the next sample truly comes from P. It is common practice to run several independent MCMC *chains*, each one a separate instance of MCMC starting from its own random initialization point, and observe if they have properly *mixed* after a reasonable number of iterations. Chains are considered mixed if they produce samples from the same regions of the sample space. Mixed chains are a good indication that P, rather than the different initializations of the chains, is determining the samples (Gelman, Carlin, Stern, & Rubin, 2004). We follow the lead of Sanborn and Griffiths (2008) and run several chains in parallel to verify that mixing is occurring, and then keep all the samples after an initial "burn-in period", discarding their ordering information (Sanborn, 2008). A burn-in period is required because samples early in the chain are more indicative of the proposal distribution and the initial sample x_0 than of the participant's beliefs about collisions.

Experiment 1

The goal of this experiment is to use pMCMC to determine a participant's collision probability distribution under different mass ratio hypotheses. The experiment has two conditions. Condition 1 draws samples from a participant's collision probability distribution assuming a 1:1 mass ratio between the two balls. Condition 2 draws samples conditioned on a 1:2 mass ratio. A mass ratio of 1:1 means that Ball 1 and Ball 2 are of equal mass and a mass ratio of 1:2 means that Ball 2 has twice the mass of Ball 1. The distribution of the samples will indicate which types of collisions the participant considers probable and improbable in each case. Because they measure the participant's beliefs, the probability distributions may deviate from those that would be produced by a participant employing the laws of physics.

The design of the experiment mimics the flow of the MCMC method. Recall that there are two basic steps in MCMC. First, a sample is drawn from the proposal distribution, Q. Second, the acceptance rule, A, is used to choose between the newly proposed sample and a previously accepted sample. Analogously, on each trial of the experiment, the participant is presented with two collisions, one previously accepted collision and one new collision from a suitable proposal distribution, and is asked to select between them. In Condition 1, for example, participants were asked which of two collisions looked closer to having a 1:1 mass ratio. As discussed previously, if the participant is utilizing the Luce choice rule to compare the two collisions, the decision process is a proper MCMC acceptance rule in Barker's algorithm and will produce samples from the target distribution, $P(\text{Collision} \mid \text{Perceived } 1:1)$.

It is crucial to select a proposal distribution that balances coverage and efficiency. A proposal distribution must produce samples from the entire sample space with enough frequency so that modes of the target distribution will be discovered and even low-density regions will

produce some samples in a reasonable length of time. Because the outcome of a frictionless, perfectly elastic collision is solely determined by the initial velocities of the two balls and their mass ratio, the proposal distribution can generate sample collisions by selecting values for each of these three parameters. In order to ensure good coverage of the space, we assumed that the psychological effects of pre-collision velocity would be symmetric and fixed the speed of Ball 1 across all trials.[4] Furthermore, the balls only had horizontal velocity, had identical shapes and sizes, and had the same initial vertical positions, ensuring that all motion was horizontal.[5] Given these constraints, only the mass ratio and Ball 2's initial horizontal velocity were free to be chosen randomly for each trial. After extensive pilot work, it was decided that selecting the natural logarithm of the mass ratio and Ball 2's initial horizontal velocity from uniform distributions best balanced the competing goals of coverage and efficiency. The uniform distribution satisfies the Barker criteria for a proposal distribution, provides excellent coverage of the space, and produces well-mixed chains and consistent results. The logarithmic representation of mass ratio was used to ensure that the samples would represent positive mass ratios and avoid biasing the distribution in favor of large ratios.

To ensure that the data provided good coverage of the space and to minimize the chance that participants memorized and compared collisions from previous trials, participants were presented with three alternating MCMC chains. To start the experiment, the participant was presented with the first trial from Chain 1 and accepted one of two samples from the proposal distribution. This step was then repeated with Chain 2 and then Chain 3. The participant was next asked to decide between the previously accepted sample from Chain 1 and a new proposed sample. The participant was then given the previously accepted sample from Chain 2 and a new proposed sample and likewise for Chain 3. This rotation continued until the end of the

experiment. The use of multiple chains also helps disguise the repetition in samples.

The results of the experiment will be used to evaluate whether two probability models, based on the direct perception and heuristic approaches to the perception of relative mass, account for the pattern of data produced by pMCMC. Recall that the direct perception approach assumes that participants can draw on an accurate model for judging relative mass. If this approach is correct, the accepted samples will reflect the mass ratio the participant is asked to judge. In Conditions 1 and 2, the accepted samples should all have a mass ratio similar to 1:1 and 1:2, respectively. If there is noise in the judgments, the accepted samples will not reflect these ratios exactly, but will still cluster around them.

According to the heuristic model of Gilden and Proffitt (1989), participants base their relative mass judgments on one of two sources of information. First, according to the angle heuristic, the ball that scatters at the greater angle is judged to be the less massive ball. Second, according to the velocity heuristic, the ball with the greater post-collision speed appears to be the lighter ball. Participants base their decision on whichever of these two heuristics is most salient. Because the collisions in this experiment are all linear, only the velocity heuristic applies (which makes this heuristic equivalent to the final-speed hypothesis suggested by Todd & Warren, 1982). A corollary of the velocity heuristic is that if the post-collision speeds of the two balls are equal, they should be judged to be of equal mass. If this model is correct, participants in Condition 1 should accept samples that have approximately equal post-collision velocities.

It is unclear what the heuristic model predicts for Condition 2, in which the participants are asked to accept samples for collisions in which Ball 2 had twice the mass of Ball 1. One possibility is that participants would judge Ball 2 to be twice as massive at Ball 1 if its post-collision velocity was twice that of Ball 1. This suggestion is very speculative as Gilden and

Proffitt (1989) provided no scale relating differences in post-collision velocity to relative mass. A weaker, but more defensible suggestion is that, although participants can use the velocity heuristic to judge whether the two ball have equal mass and, if not, which is heavier, they have no developed model for quantifying the relative masses of the balls. In this case, the accepted samples should span the range of all collisions in which the post-collision velocity of Ball 2 is greater than that of Ball 1, regardless of the absolute difference in velocity.

Method

Participants. Four University of Massachusetts students and postdoctoral researchers (P1-P4) participated in this study for pay. They received $11 per session for 10 sessions (5 from each condition). P1 had previous knowledge of pMCMC prior to the experiment, but all participants were naïve to the purpose of the experiment.

Stimuli. The collisions were presented on a computer CRT. Each ball was a 2.3 cm diameter circle (approximately 2.7° of visual angle) of constant color. Because the participants did not need to identify a particular ball in Condition 1 – the participants were asked to judge which collision looked more like the balls had equal mass – the balls were both red. In Condition 2, however, the participants were asked to base their decisions on a mass ratio in which the second ball had twice the mass of the first, and not vice versa. Because it is important for the participants to be able to identify the two balls, Balls 1 and 2 were colored red and green, respectively. The background was neutral gray. A sample display from Condition 2 is displayed in the left panel of Figure 1.

In each collision, the two balls traveled horizontally across the vertical center of the screen and collided at the horizontal center of the screen. Positive and negative velocities are towards the right and left, respectively. The pre-collision velocity of Ball 1 was always 6.2 cm/s. The pre-collision velocity of Ball 2 and the log mass ratio of the two balls were determined as previously discussed. The participant saw the two balls move for 1 second both before and after the collision.

Procedure. There were three MCMC chains, which were interleaved as previously described. On each trial, the participant saw two collisions presented serially. The two collisions were labeled Collision 1 and Collision 2 on otherwise blank screens immediately preceding their presentation. One of the two collisions was the accepted collision from the previous trial of the chain. The other collision was sampled from the proposal distribution. The two collisions were presented in random order. The pre-collision velocity of Ball 2 was selected from a uniform probability distribution over the range -12.3 cm/s through 6.2 cm/s. In each condition, the uniform log mass ratio proposal distribution was centered on the mass ratio of interest: $\ln(1) = 0$ for Condition 1 and $\ln(0.5) \approx -0.69$ for Condition 2. In both cases, the distributions ranged 1.25 above and below their means in the natural log-ratio scale. Thus, the mass ratios in Condition 1 were greater than 1:3.5 and less than 3.5:1. In Condition 2 the mass ratios ranged between 1:7.0 and 1.8:1. The post-collision velocities of the balls were calculated using conservation of energy and momentum (Halliday, Resnick, & Walker, 2001). The presentation order of the two collisions was randomized. For the first trial of each chain, both collisions were sampled from the proposal distribution.

After the two collisions were viewed, the participant had the option of viewing the

collisions again if needed. Before the experiment, participants were discouraged from reviewing collisions unless they felt they hadn't paid attention on the first viewing. If the participant declined to review the collisions, he or she answered the question "Was Collision 1 or Collision 2 closer to a 1:1 mass ratio?" in Condition 1 and "Was Collision 1 or Collision 2 closer to a 1:2 (red:green) mass ratio?" in Condition 2. Before the experiment, these questions were explained in detail to the participants who declared that they understood what was being asked of them. No feedback was given.

Each chain consisted of 333 trials for a total of 999 trials per condition. Each condition was split over 5 sessions of approximately 200 trials each. Each session took approximately 20 minutes. All 4 participants participated first in Condition 1 and then in Condition 2.[6]

Results

The results of Experiment 1, Condition 1 are displayed in Figure 2 and the results of Condition 2 are displayed in Figure 3. Each location in the figures represents a collision. Because the collisions have only two free parameters – Ball 1's pre-collision velocity is fixed, and the post-collision velocities are completely determined by the pre-collision velocities and the mass ratio – all of the data lie in a plane. There are many possible choices for the axes, but to aid ease of visualization with regard to the velocity heuristic, the x- and y-axes in the figures were selected to be the post-collision velocities of Ball 1 and Ball 2, respectively. Recall that positive and negative velocities indicate motion to the right and left, respectively.

Because it is fixed, the pre-collision velocity of Ball 1 is not represented in the figures. Because the pre-collision velocity of Ball 2 is a linear function of the post-collision velocities,

several qualitatively significant lines of constant Ball 2 pre-collision velocity have been drawn on the graphs. These solid lines indicate perceptually salient regions in which Ball 2 was initially moving twice as fast as Ball 1 in the opposite direction (-12.3 cm/s); was moving equally as fast as Ball 1 in the opposite direction (-6.2 cm/s); was stationary (0 cm/s); and was moving equally as fast as Ball 1 in the same direction (6.2 cm/s). In the last case, which was never actually shown to a participant, no collision would occur. To provide more detailed information, the grayscale shading of each sample on the graph indicates the Ball 2 velocity (ranging from light gray indicating 6.2 cm/s to black indicating -12.3 cm/s).

The graphs also contain noteworthy dashed lines of constant mass ratio. All collisions along one of these lines have identical mass ratios. If the participants' mass ratio beliefs are physically accurate, their samples should cluster along the 1:1 line in Condition 1 and the 1:2 line in Condition 2. The 3.5:1 and 1:7 lines are present to indicate the largest and smallest mass ratios used across both conditions.

Each graph has a light gray triangle indicating the area of the proposal distribution for that condition, i.e., all of the collisions presented to the participants fell within these triangles. It is important to note that, due to the proposal distribution, the participants observed collisions from all regions of the triangle, as shown in Figure 4. If the participants had been behaving randomly, the accepted samples would be similarly spread across the entire proposal region. To the extent to which the participant is systematically accepting collisions, the accepted samples should cluster in particular regions of the space.

The circles in Figure 2 and Figure 3 represent the accepted samples from each of the four participants' $P(\text{Collision} \mid \text{Perceived } 1:1)$ and $P(\text{Collision} \mid \text{Perceived } 1:2)$ functions, respectively, collapsed across all three MCMC chains. The size of the sample points represent the number of

times the participant accepted that collision before accepting a new proposal – large points indicate individual collisions that a participant strongly preferred. Thus, modes in the distribution can be indicated both by clusters of samples and the presence of large sample points.

As described in the MCMC section, we followed the common practice of eliminating the beginning of each chain and keeping all of the remaining samples. The graphs do not show the first 30 samples from each chain because they are more indicative of the proposal distribution and the initial sample x_0 than of the participants' beliefs about collisions. We judged that the chains for each participant were well-mixed by visually verifying that all regions of the space that contained a substantial number of samples also contained samples from each of the three chains.

It is possible that participants used different strategies at different points in the experiment. To test for this possibility, we plotted the accepted samples from the first and second half of each observer's data. These plots did not indicate any substantial changes in distributions across the course of the experiment, suggesting that participants used a relatively stable process throughout the experiment. Interestingly, most observers rejected newly proposed samples more frequently in the second half of their trials. This behavior could result from a strategy of choosing the most familiar collision on some trials, or it might result from an increased probability of being near a mode of the distribution in the second half. In most cases, the samples from the second halves covered the same collision regions as those from the first halves, so this behavior is not a concern.

Discussion

First consider the results from Condition 1 displayed in Figure 2. Interpretation of the results will be based on a comparison of these data to distributions generated by models that instantiate the direct perception and velocity heuristic approaches. To create these benchmarks, we ran two simulated participants through the same experiment as the human participants. The first simulated participant acted in accord with the direct-perception approach. That is, this observer judged mass ratio according to the true mass ratio, but it was assumed that the judgment was noisy.[7] In particular, this direct-perception simulated observer accepted samples according to Barker's rule using the probability density function

$$P(\text{Collision} \mid \text{Perceived } 1:1) = N(\ln(\text{mass ratio});0,0.17), \tag{8}$$

where $N(x;\mu,\sigma)$ is the density of a Gaussian distribution with mean μ and standard deviation σ at point x. The standard deviation is the mean standard deviation when this model was fit to the participants from Cohen (2006, Experiment 1). For example, given Collisions 1 and 2 with mass ratios 1.1 and 1.3, the probability that this observer would accept Collision 1 over Collision 2 is

$$\frac{N(\ln(1.1);0,0.17)}{N(\ln(1.1);0,0.17)+N(\ln(1.3);0,0.17)} = \frac{2.01}{2.01+0.71} = 0.74.$$

The second simulated observer judged mass ratio in accord with the velocity heuristic model. This observer judged balls to be of equal mass when the absolute values of their post-collision velocities were equal, but, again, with noise. In particular, this velocity heuristic simulated observer accepted samples using

$$P(\text{Collision} \mid \text{Perceived } 1:1) = N(|y|-|x|;0,1.72), \tag{9}$$

where x is Ball 1's post-collision velocity and y is Ball 2's post-collision velocity. The standard deviation was determined by fitting a normal distribution to the data from a pilot study in which P1-4 were shown the same collisions as in Experiment 1, but were asked to judge whether the post-collision velocities were equal. For example, given Collision 1 with post-collision velocities

(cm/s) of -3 and 6 and Collision 2 with post-collision velocities of -5 and 6, the probability that this observer would accept Collision 1 over Collision 2 is

$$\frac{N(|6|-|-3|;0,1.72)}{N(|6|-|-3|;0,1.72)+N(|6|-|-5|;0,1.72)}=\frac{0.05}{0.05+0.20}=0.21.$$

The samples produced from these two simulated observers are displayed Figure 5.

Visual comparison of the human participants' data in Figure 2 and the simulated data in Figure 5 reveals a number of interesting similarities and differences. Participants 1 and 4 appear to have accepted many samples when $y=|x|$, as predicted by the heuristic model, with a slight bias towards accepting samples in the lower region of the triangle. Participant 2 also looks very similar to the heuristic model, but avoided all samples near $(6.2,6.2)$, the point where the $y=x$ line intersects the proposal region. Participant 3 had a strong bias for collisions with equal and opposite post-collision velocities and, although this participant did accept samples for which Ball 1 stopped after the collision, accepted very few other samples in the middle of the proposal region.

To quantitatively determine which model better describes human performance, we compared the data from Condition 1 to models of the direct perception and heuristic approaches. The models used in this analysis are identical to (8) and (9), but with the standard deviations free to vary. That is, the direct perception model is given by

$$P(\text{Collision}\,|\,\text{Perceived}\,1:1)=N(\ln(\text{mass ratio});0,\sigma_{dp}) \tag{10}$$

and the velocity heuristic model is given by

$$P(\text{Collision}\,|\,\text{Perceived}\,1:1)=N(|y|-|x|;0,\sigma_{vh}) \tag{11}$$

where σ_{dp} and σ_{vh} are parameters that determine the relative noise in each judgment. To make the models mathematically comparable, the samples were always represented in (Ball 1 post-

collision velocity, Ball 2 post-collision velocity) space and both models were normalized (i.e., scaled to integrate to one) over the triangular proposal region.

The models were fit to the accepted samples from each human and the two simulated observers. Maximum likelihood was used as a measure to determine the best-fitting standard deviations in (10) and (11). Five fits were run for each model on each data set and the highest scoring standard deviations were used for subsequent inter-model comparisons. The log likelihood fit values and best fitting parameters are given in Table 1.[8] The lower the fit value, the better the model accounted for the data.

To test the validity of this modeling approach, the two models were first fit to the data generated by the simulated observers for which the correct model is known. As desired, the data generated by the direct-perception and velocity-heuristic simulated observers were better described by their associated models. The two models were then fit to the data from each of the four participants of Experiment 1 in Condition 1 (P1-P4). The results are highly consistent. For every participant, the velocity heuristic model was a more likely explanation of the samples. This analysis suggests that the distribution of samples of equal mass collisions is better described by a heuristic strategy in which people judge whether the post-collision velocities of the two balls are equal.

Next consider the data from Condition 2 in which participants were asked to determine which collision looked more like a 1:2 mass ratio. These data are shown in Figure 3. Recall that if the participants acted according to the direct perception model, the accepted samples should cluster around the 1:2 constant mass-ratio line. Consistent with the Condition 1 results, it is clear from Figure 3 that none of the participants acted according to this model. Indeed, it is striking that the participants were generally much more likely to accept a 1:7 collision than a 1:2

collision.

What strategy, then, are participants using to accept 1:2 samples? Consider Figure 6, which plots each participant's Condition 1 and Condition 2 samples on the same graphs. The grayscale shading of the points now indicates a sample's condition rather than the Ball 2 pre-collision velocities. There was little overlap in the accepted samples across Conditions 1 and 2.[9] More interesting, and in agreement with the weak version of the Gilden and Proffitt (1989) heuristic model, participants tended to accept any sample as having a mass ratio of 1:2 if it was below the region of 1:1 samples (see Figure 6). This region corresponds to the area for which $|y| < |x|$, indicating that Ball 2 moved more slowly than Ball 1 after the collision. Taken together, the results of this experiment suggest that, although participants seem to possess a model for equal mass collisions (albeit an incorrect one) and can use this model to judge which ball is more massive, they cannot attach a quantitative value to the mass ratio.

Experiment 2

The participants in colliding balls experiments must rely on the ball velocities to determine mass ratio. Given the results of Experiment 1, it is reasonable to ask if the inability to accurately determine mass ratios can be attributed to difficulty in accurately observing the velocities. In the first experiment (as in almost all colliding ball experiments), the background was untextured and there were no marks indicating horizontal locations. There is evidence that the visual system is more sensitive to motion against textured backgrounds (Palmer, 1999). Adding texture to the background could make the starting position of the two balls and the collision location more salient and could aid in the estimation of the ball velocities.

The goal of Experiment 2 was to explore the possibility that the addition of texture and markers to the display would improve velocity estimates and, therefore, improve mass-ratio judgments. The experiment was identical to Experiment 1, Condition 1 except that the collisions were no longer displayed on a gray background. Instead, the collisions were viewed against a background that approximated granite; the point of collision was marked; and vertical lines were equally spaced across the screen. The right panel of Figure 1 contains an example frame from this experiment.

Method

Participants. Two University of Massachusetts graduate students (P5 and P6) participated in the study for pay. They received $7 per session for 5 sessions.[10] They were naïve to the purpose and methodology of the experiment.

Stimuli. The stimuli were identical to those of Experiment 1, Condition 1 with one modification. In Experiment 1, the balls collided on a neutral gray background. In this experiment, the background was textured to look roughly like granite. A bold, black vertical line marked the center of the screen, the point of collision, and five equally spaced, vertical lines were placed on each side of the point of collision.

Procedure. The procedure was identical to that of Experiment 1, Condition 1.

Results

The results are displayed in Figure 7, using the same graph style as Experiment 1. Just as with Experiment 1, the first 30 samples in each chain were discarded.

Discussion

The results of this experiment were qualitatively identical to those of Experiment 1, Condition 1. The results of Participant 5 look very similar to those of Participant 3 from Experiment 1; both participants almost exclusively preferred to accept collisions with equal and opposite pre- and post-collision velocities. The results of Participant 6 are comparable to those of Participant 2 from Experiment 1. The direct perception and velocity heuristic models were fit to the data from these two participants. The results are given in Table 1. Just as in Experiment 1, Condition 1, the heuristic model explained these observer's data better than the direct perception model. This result suggests that the textures and tick marks failed to move observers closer to accurate mass perception. Even in a situation in which a great deal of care was taken to ensure that the velocities were as detectable and discriminable as possible, the participants still did not act in accord with the direct perception model.

General Discussion

The results and analysis of these two experiments suggest that a model based on the Gilden and Proffitt heuristic model explains 1:1 mass ratio perception better than the direct perception model. Neither experiment's results are in agreement with the predictions of the

accurate, direct perception view. Participants accepted samples with incorrect mass ratio, did not accept samples from appropriate regions, and demonstrated strong biases for particular regions of the collision space.

Consistent with the results of Todd & Warren (1982), the results of Experiment 1, Condition 1 and Experiment 2 suggest that participants may have relied on an assortment of heuristics to judge mass ratios of 1:1. All participants accepted samples that had nearly equal pre- and post-collision velocities. Subjectively, these collisions are especially indicative of equal mass: Both balls move together at the same rate, collide, and then move apart at the same rate. It is interesting to speculate that this type of collision would imply equal mass under a very wide range of conditions – for example, with the addition of friction or slippage. Surprisingly, only Participant 3 consistently accepted collisions in which Ball 2 was stationary before the collision and Ball 1 stopped moving after the collision. This collision type is highly salient and correctly specifies a 1:1 mass ratio. Instead, four of the participants (Participants 1, 2, 4, and 6) preferred to accept collisions in which Ball 2 began at rest, but, after the collision, Balls 1 and 2 moved apart with equal velocity. Ball 2 is always more massive in such collisions.

Both of these heuristics, based on judging balls to have equal mass when their post-collision speeds are equal, are implied by the Gilden and Proffitt (1989) velocity heuristic. Indeed, the data from Participant 6 are especially consistent with the velocity heuristic, as are the data from Participant 2 when the post-collision velocity of Ball 1 is negative. The other participants, however, failed to accept collisions in some regions predicted by the velocity heuristic. Participant 4, for example, did not accept any samples near the top edge of the proposal triangle. Different participants seem to rely on different heuristics. Others have concluded that people rely on heuristics to judge relative mass, but data from the pMCMC provide a clear

starting point for exploring such strategies and individual variation.

Perhaps the most surprising result is that participants tended to accept any sample as having a mass ratio of 1:2 if Ball 2 seemed more massive than Ball 1 (Experiment 1, Condition 2). These data suggest that each participant used an (incorrect) rule to determine if the two balls are of equal mass. If the two balls were judged to be of different masses, the participant could determine which ball was heavier, but not by how much.

It is possible that the inability to discriminate between 1:2 collisions and other cases in which Ball 2 is more massive than Ball 1 is a problem of having an unanchored scale. The accepted samples from such a participant would cluster around a single mass ratio, although it might not be the correct one. After training, the participant's scale would be anchored and this cluster would shift to be centered on the 1:2 line. Given the range and configuration of the accepted samples, there is little indication of this pattern in the data for P2-4. The accepted samples from Experiment 1 Condition 2 were generally spread over a wide range and did not lie on any particular mass ratio line. A broad range of collision mass ratios were tested, therefore it is unlikely that the absence of the unanchored participant pattern was due to experimental limitations. There is a hint that the data from P1 lie near a single mass ratio. Although it is unlikely that this participant used an unanchored scale to determine a 1:2 mass ratio when the 1:1 results so strongly suggest a heuristic strategy, future work will be needed to determine the viability of such strategy switching.

The idea that participants could determine which ball is more massive, but not by how much, supports a new interpretation of Experiment 2 results from Cohen (2006). During the first phase of that experiment, participants were asked to assign one of 10 mass ratios (ranging from 1:3.5 to 3.5:1) to each of 102 collisions. Half of these collisions had a mass ratio of less than 1

and half had a mass ratio greater than 1. Although the results showed that participants were relatively good at determining whether the mass ratio was greater or less than 1, they were very poor (before training) at determining the absolute value of the log mass ratio. Indeed, participants, for the most part, appeared to have a relatively flat function relating the perceived and actual absolute values of the log mass ratios.

Following the bulk of research in the colliding balls paradigm, the focus of this paper has been to describe participants' intuitive models of mass ratios. There have been a number of recent papers that address the effect of training on mass ratio judgments (e.g., Cohen, 2006, Jacobs, et al., 2000; Jacobs et al., 2001). Consistent with the current results, the general finding is that, before training, participants are inaccurate and tend to use heuristics. Training has been shown to improve performance. Some researchers have concluded that the more accurate performance after training suggests that participants are acting in accord with the direct perception view (e.g., Jacobs, et al., 2000); that is, participants have learned to detect mass. Another possibility is that participants may have learned a small, but effective, set of heuristics. Although it is difficult to establish such heuristics a priori, especially for a wide range of stimuli, a participant can accurately classify many equal mass collisions with the two simple heuristics discussed above: collisions with equal pre- and post-collision velocities and collisions in which Ball 2 starts out stationary and Ball 1 stops moving after the collision. As discussed above, however, pMCMC also revealed a number of individual differences. Psychological MCMC might allow us to develop more nuanced descriptions of human mass perception.

References

Cohen, A. (2006). Contributions of invariants, heuristics, and exemplars to the visual perception of relative mass. *Journal of Experimental Psychology: Human Perception and Performance, 32,* 574-598.

Gelman, A., Carlin, J., Stern, H., & Rubin, D. (2004). *Bayesian data analysis*. Chapman & Hall/CRC: New York.

Gibson, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.

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

Halliday, D., Resnick, R., & Walker, J. (2001). *Fundamentals of Physics*. John Wiley & Sons, Inc: New York.

Hastings, W. (1970). Monte Carlso sampling methods using Markov chains and their applications. *Biometrika*, 57, 97-109.

Hecht, H. (1996). Heuristics and invariant in dynamic event perception: Immunized concepts or nonstatements? *Psychonomic Bulletin & Review, 3*, 61–70.

Jacobs, D. M., Michaels, C. F., & Runeson, S. (2000). Learning to perceive the relative mass of colliding balls: The effects of ratio scaling and feedback. *Perception & Psychophysics, 62*, 1332–1340.

Jacobs, D. M., Runeson, S., & Michaels, C. F. (2001). Learning to visually perceive the relative mass of colliding balls in globally and locally constrained task ecologies. *Journal of Experimental Psychology: Human Perception and Performance, 27,* 1019 –1038.

Luce, R. D. (1959). *Individual Choice Behavior*, Wiley.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., & Teller, E. (1953). Equations of state calculations by fast computing machines. *Journal of chemical physics, 21,* 1087-1092.

Palmer, S. (1999). *Vision Science: Photons to Phenomenology*. The MIT Press: Cambridge, MA.

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

Runeson, S. (1977). On the possibility of smart perceptual mechanisms. *Scandinavian Journal of Psychology, 18*, 172–179.

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

Sanborn, A. N. Personal communication. March 4, 2008.

Sanborn, A. N. & Griffiths, T. L. (2008). Markov chain Monte Carlo with people. *Advances in Neural Information Processing Systems, 20*. http://books.nips.cc/. Neural Information Processing Systems Foundation, La Jolla, CA.

Sanborn, A., Mansinghka, V., & Griffiths, T. (2006, November). *A Bayesian framework for modeling intuitive dynamics*. Poster session presented at the annual meeting of the Psychonomic Society, Houston, TX.

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

Author Note

M.G.R. was supported by NIMH Grant MH16745.

The authors would like to thank Adam Sanborn and Thomas Griffiths for their pioneering work on psychological MCMC and their helpful discussions and insights and Erik Learned-Miller for the initial idea behind Experiment 2.

Footnotes

[1] To make clear that it is the human's judgment of mass ratio that is being measured and not the true mass ratio, the phrase "Perceived mass ratio" is used in all equations. That is not to say that the subject necessarily directly perceives mass. "Judged mass ratio" would also be an appropriate phrase.

[2] Currently, we have no technique for estimating the probability density in (6), which, in turn, would allow us to estimate the probability density of (4), but the experimental paradigm described below does not make any assumptions about its values. Therefore, its estimation can be set aside for future work without detracting from the generality of the results presented here.

[3] Although the term "Markov chain" is most familiar as describing a class of model for psychological or physical processes, that is not its meaning in this context. The MCMC algorithm uses a Markov chain to draw samples, but it is not a model of the underlying distributions and does not assume that those distributions are described by Markov chains. In fact, as described below, MCMC is useful because it makes very weak assumptions about the distribution of interest. Therefore, we can use it to learn about (7) without imposing any sort of restrictions on the subject's mechanism for mass ratio perception.

[4] Of the four subjects, two stated that they were unaware that Ball 1's velocity was always fixed. The other two subjects were not asked.

[5] Past collision experiments have presented collisions on a vertical screen (e.g., Jacobs et al., 2001). The collisions appear natural. Indeed, viewing a pool ball collision from above on a television would give exactly this view.

[6] P1 performed the trials of Condition 1 in a self-paced manner on his laptop computer. There was no appreciable impact on his results compared to the results of the other subjects in this

condition. In Condition 2, he participated under the same conditions as the other three participants.

[7] A more sophisticated model might also incorporate the idea that the perception of mass ratio would be more difficult to the extent that the velocities or relative velocities are more difficult to detect (e.g., Cohen, 2006).

[8] The log likelihood calculation assumes that all the samples are independent. Neighboring samples in an MCMC chain, however, are correlated. To ensure that the correlations did not influence the results, we refit the models using only every 20th sample from each chain – a common procedure for increasing the independence of MCMC samples (see Gelman et al. 2004). The log likelihood values for these "thinned" samples favored the same models as the full log likelihood values reported in Table 1.

[9] If the mass ratio range were extended in Condition 1, it is possible that the accepted samples from Participant 1 may have overlapped across conditions. Even if this were the case, however, the bulk of the 1:2 samples from this participant were far from the 1:1 samples.

[10] Due to experimenter error, P5's first session consisted of 504 trials. To compensate, his second session was shortened to only contain 120 trials, and then followed by two regular-length sessions.

Table 1

Maximum log likelihood fit values and parameters for the two simulated observers and each observer of Condition 1, Experiment 1 and Experiment 2.

Observer	Direct Perception Model		Velocity Heuristic Model	
	-Ln L	σ_{dp}	-Ln L	σ_{vh}
DP	3803.5	0.19	4764.5	2.78
VH	4332.0	4.02	3454.3	1.48
P1	4753.6	1.16	3904.0	2.36
P2	4342.3	0.36	3935.1	2.44
P3	4156.5	0.29	3906.5	2.36
P4	4745.8	1.01	3664.5	1.83
P5	4517.2	0.46	4192.8	3.32
P6	4704.3	0.73	4139.7	3.10

Notes. DP is the direct perception simulated observer; VH is the velocity heuristic simulated observer; σ_{dp} and σ_{vh} are the standard deviations for the direct perception and velocity heuristic models, respectively; Ln L is the maximum log likelihood fit value.

Figure Captions

Figure 1. Left: A frame from a two-ball collision video from Experiment 1, with neutral gray, untextured background. Right: A video frame from Experiment 2 with textured backgrounds and horizontal position markers.

Figure 2. The pMCMC samples from Participants 1-4's $P(\text{Collision} \mid \text{Perceived } 1:1)$ distribution (Experiment 1, Condition 1). The x- and y-axes represent the post-collision velocities of the two balls. Positive values indicate rightward motion and negative values indicate leftward motion. The size of each circle represents the number of times it was repeated in the accepted sample sequence. The grayscale value of each circle represents the pre-collision velocity of Ball 2 (Ball 1's pre-collision velocity was fixed at 6.2 cm/s). The dashed lines are lines of constant mass ratio, and the solid lines are lines of constant Ball 2 pre-collision velocity. The gray triangle indicates the region that the proposal distribution samples covered.

Figure 3. The pMCMC samples from Participants 1-4's $P(\text{Collision} \mid \text{Perceived } 1:2)$ distribution (Experiment 1, Condition 2). The x- and y-axes represent the post-collision velocities of the two balls. Positive values indicate rightward motion and negative values indicate leftward motion. The size of each circle represents the number of times it was repeated in the sample sequence. The grayscale value of each circle represents the pre-collision velocity of Ball 2 (Ball 1's pre-collision velocity was fixed at 6.2 cm/s). The dashed lines are lines of constant mass ratio, and the solid lines are lines of constant Ball 2 pre-collision velocity. The gray triangle indicates the region that the proposal distribution samples covered.

Figure 4. The samples from the proposal distribution for Participant 1 in the 1:1 and 1:2 conditions of Experiment 1.

Figure 5. The pMCMC samples from the simulated direct perception (left) and velocity heuristic (right) observers. The x- and y-axes represent the post-collision velocities of the two balls. Positive values indicate rightward motion and negative values indicate leftward motion. The size of each point represents the number of times it was repeated in the sample sequence. The grayscale value of each point represents the pre-collision velocity of Ball 2 (Ball 1's pre-collision velocity was fixed at 6.2 cm/s). The thin dashed lines are lines of constant mass ratio, and the thin solid lines are lines of constant Ball 2 pre-collision velocity. The gray triangle indicates the region that the proposal distribution samples covered.

Figure 6. The pMCMC samples from Participants 1-4's $P(\text{Collision} \mid \text{Perceived } 1:1)$ and $P(\text{Collision} \mid \text{Perceived } 1:2)$ distributions plotted together (Experiment 1). The x- and y-axes represent the post-collision velocities of the two balls. Positive values indicate rightward motion and negative values indicate leftward motion. The size of each circle represents the number of times it was repeated in the sample sequence. The dashed lines are lines of constant mass ratio, and the solid lines are lines of constant Ball 2 pre-collision velocity.

Figure 7. The pMCMC samples from Participants 5 and 6's $P(\text{Collision} \mid \text{Perceived } 1:1)$ distributions (Experiment 2). These participants saw the collisions against a textured background with horizontal positioning marks. The x- and y-axes represent the post collision velocities of the two balls. Positive values indicate rightward motion and negative values indicate leftward motion. The size of each circle represents the number of times it was repeated in the sample sequence. The grayscale value of each circle represents the pre-collision velocity of Ball 2 (Ball 1's pre-collision velocity was fixed at 6.2 cm/s). The dashed lines are lines of constant mass ratio, and the solid lines are lines of constant Ball 2 pre-collision velocity. The gray triangle indicates the region that the proposal distribution samples covered.

Figure 1

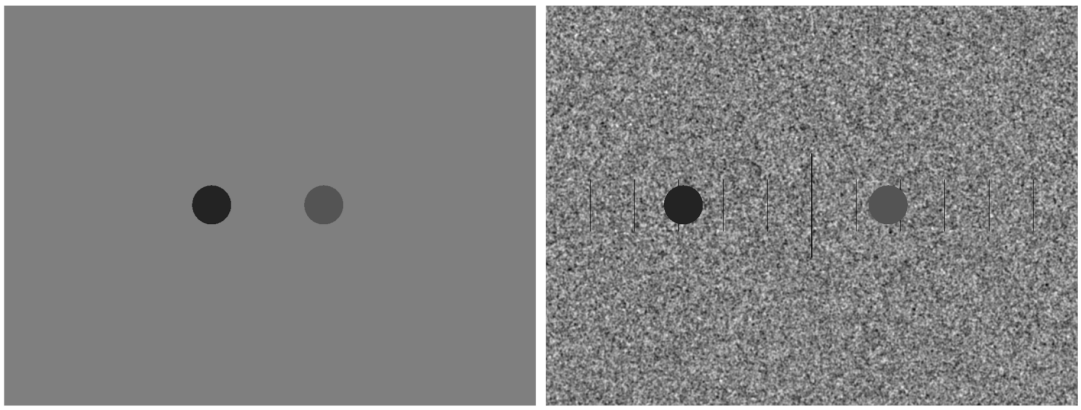


Figure 2

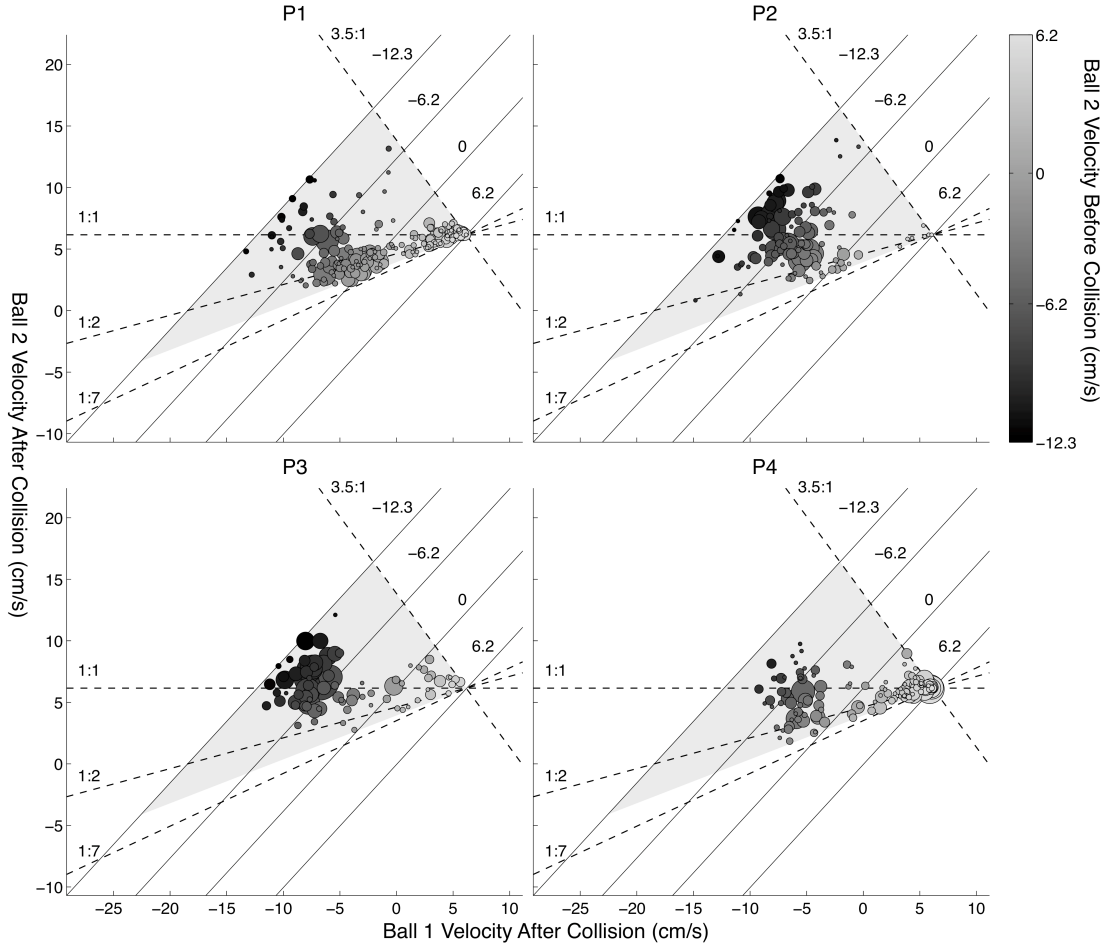

P1
P2
P3
P4
3.5:1
−12.3
−6.2
0
6.2
1:1
1:2
1:7
Ball 2 Velocity After Collision (cm/s)
Ball 1 Velocity After Collision (cm/s)
Ball 2 Velocity Before Collision (cm/s)
6.2
0
−6.2
−12.3

Figure 3

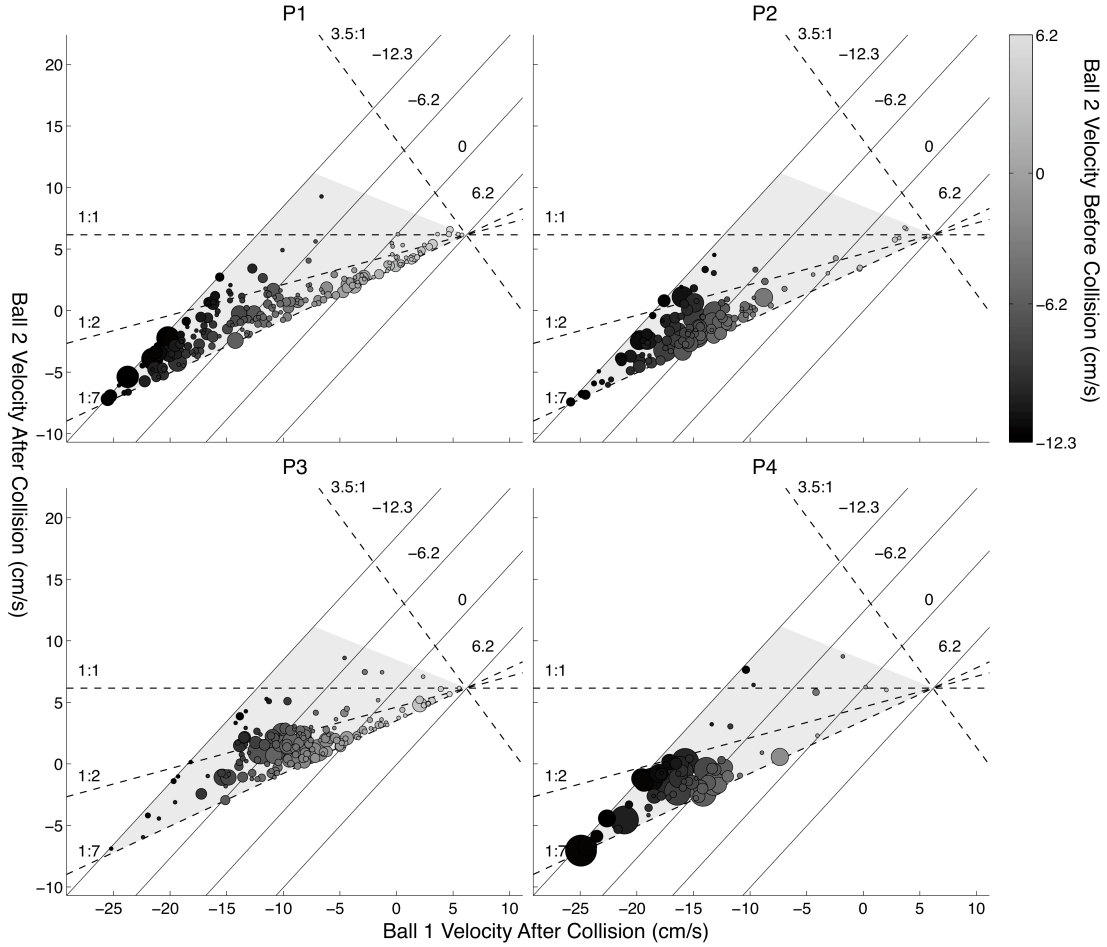

P1
P2
P3
P4
3.5:1
−12.3
−6.2
0
6.2
1:1
1:2
1:7
Ball 2 Velocity After Collision (cm/s)
Ball 1 Velocity After Collision (cm/s)
Ball 2 Velocity Before Collision (cm/s)
6.2
0
−6.2
−12.3

Figure 4

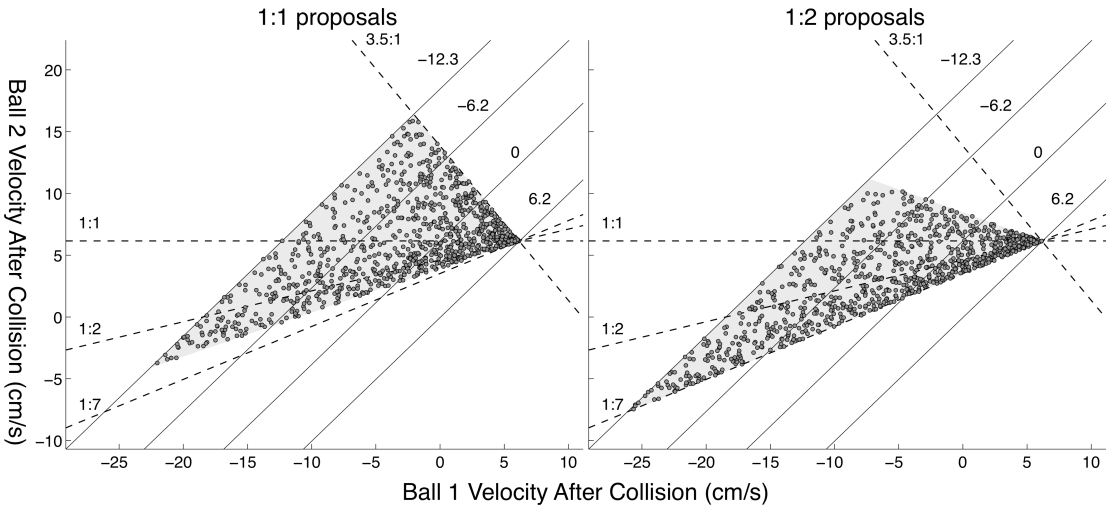

1:1 proposals
1:2 proposals
3.5:1
−12.3
−6.2
0
6.2
1:1
1:2
1:7
Ball 2 Velocity After Collision (cm/s)
Ball 1 Velocity After Collision (cm/s)

Figure 5

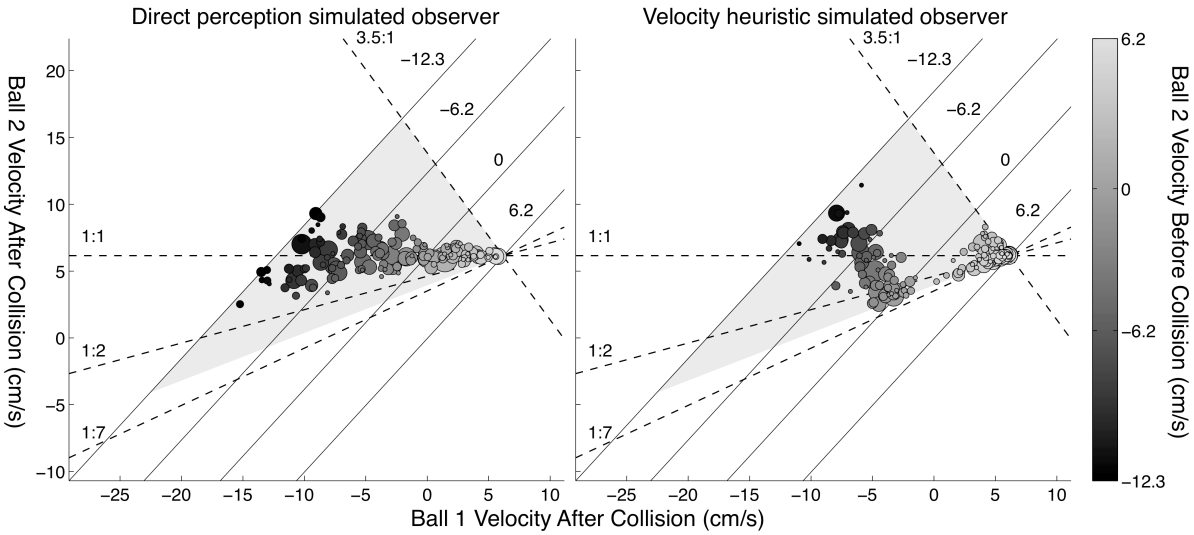

Direct perception simulated observer
Velocity heuristic simulated observer
3.5:1
−12.3
−6.2
0
6.2
1:1
1:2
1:7
Ball 2 Velocity After Collision (cm/s)
Ball 1 Velocity After Collision (cm/s)
Ball 2 Velocity Before Collision (cm/s)
20
15
10
5
0
−5
−10
−25
−20
−15
−10
−5
0
5
10
6.2
0
−6.2
−12.3

Figure 6

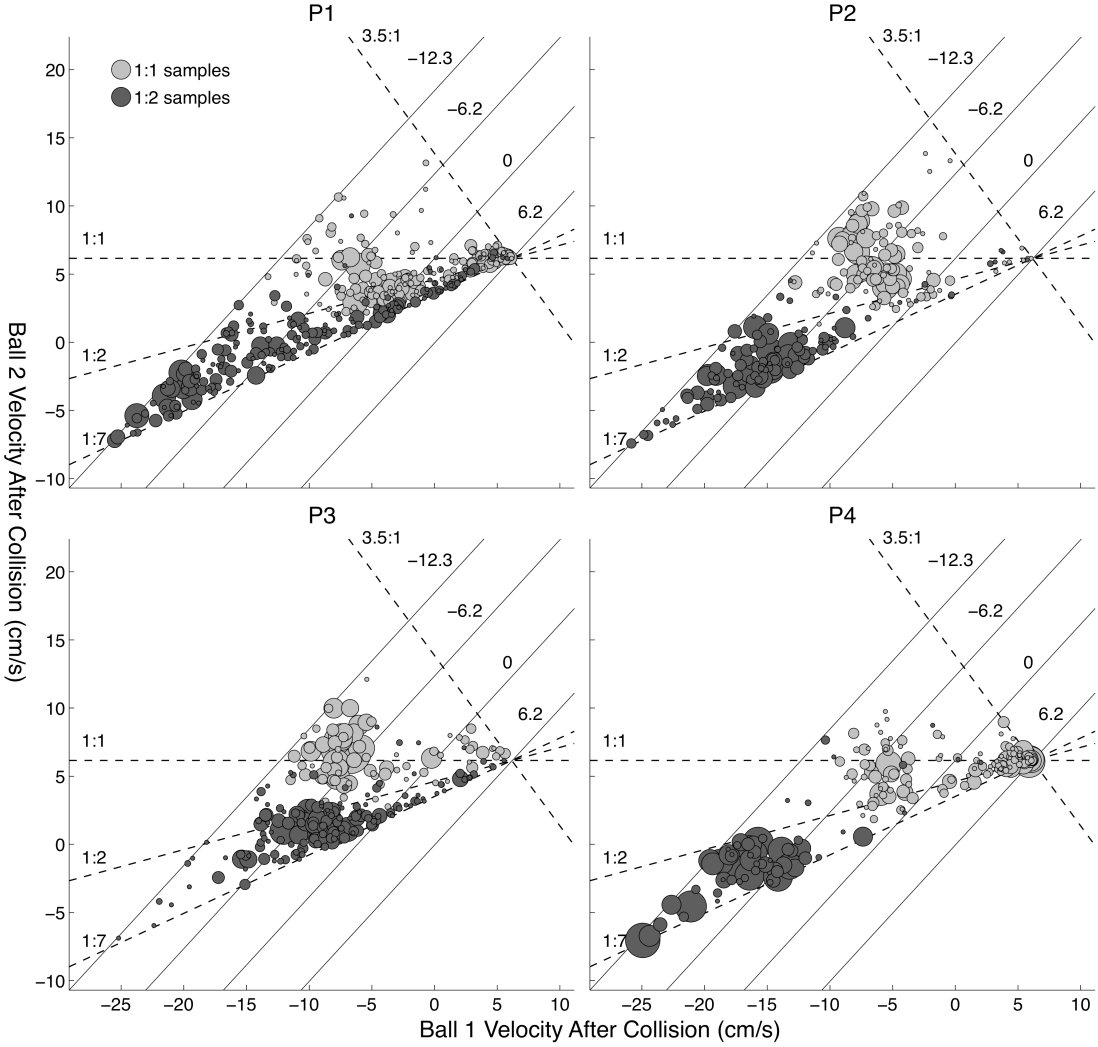
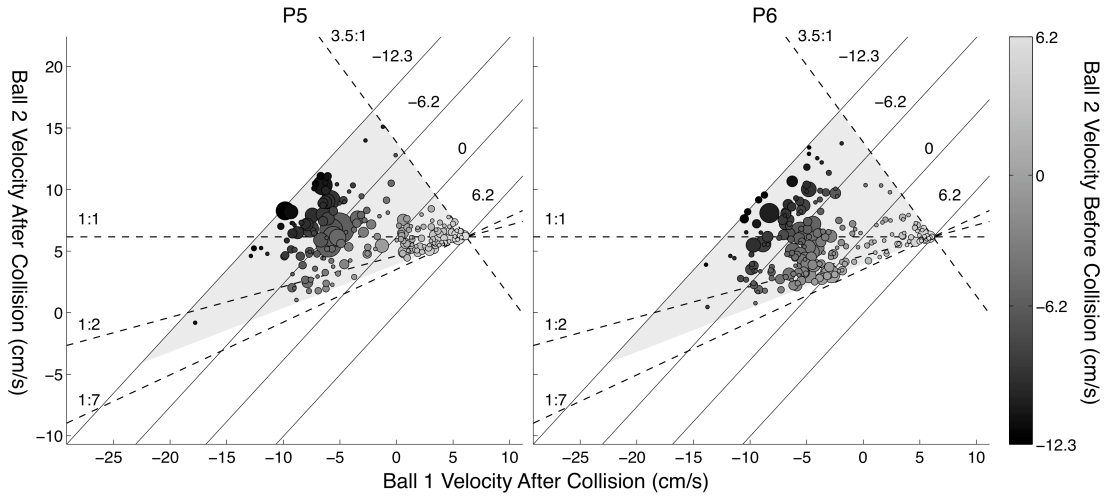

P1
P2
P3
P4
1:1 samples
1:2 samples
3.5:1
−12.3
−6.2
0
6.2
1:1
1:2
1:7
Ball 2 Velocity After Collision (cm/s)
Ball 1 Velocity After Collision (cm/s)

Figure 7


P5
P6
3.5:1
−12.3
−6.2
0
6.2
1:1
1:2
1:7
Ball 2 Velocity After Collision (cm/s)
Ball 1 Velocity After Collision (cm/s)
Ball 2 Velocity Before Collision (cm/s)
6.2
0
−6.2
−12.3