Journal of Experimental Psychology:
Human Perception and Performance
2006, Vol. 32, No. 3, 574–598

0096-1523/06/$12.00 DOI: 10.1037/0096-1523.32.3.574

Contributions of Invariants, Heuristics, and Exemplars to the Visual Perception of Relative Mass

Andrew L. Cohen
University of Massachusetts at Amherst

Some potential contributions of invariants, heuristics, and exemplars to the perception of dynamic properties in the colliding balls task were explored. On each trial, an observer is asked to determine the heavier of 2 colliding balls. The invariant approach assumes that people can learn to detect complex visual patterns that reliably specify which ball is heavier. The heuristic approach assumes that observers only have access to simple motion cues. The exemplar-based approach assumes that people store particular exemplars of collisions in memory, which are later retrieved to perform the task. Mathematical models of these theories are contrasted in 2 experiments. Observers may use more than 1 strategy to determine relative mass. Although observers can learn to detect and use invariants, they may rely on either heuristics before the invariant has been learned or exemplars when memory demands and similarity relations allow.

Keywords: dynamic properties, relative mass, collision dynamics, computational modeling

The purpose of this research is to begin to explore the potential contributions of invariants, heuristics, and exemplars to the perception of dynamic properties as realized by the perception of relative mass in the colliding balls task. 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. The observer's task is either to determine which of the two balls is heavier or to make a quantitative estimate of mass ratio. Recent evidence suggests that observers can become quite adept at determining the relative mass of the two balls (Jacobs, Michaels, & Runeson, 2000; Jacobs, Runeson, & Michaels, 2001; Runeson, Juslin, & Olsson, 2000), and the focus of research has been on what mechanisms observers use to attain this level of performance. More generally, the central issue has been how people are able to determine the *dynamics* of an event, such as a collision, by viewing only the *kinematics* of the event.

Kinematics is the branch of mechanics pertaining to the study of pure motion, such as position, angle, velocity, and acceleration. Dynamics (or, more specifically, kinetics) is the branch of mechanics that examines the forces acting on moving objects. In the present experimental context, dynamics refers more generally to the study of any "hidden" causal factor. Under this taxonomy, a visual presentation of a collision contains only kinematics. The observer can see, for example, the relative velocities of the two balls before and after the collision. The observer is not privy to the underlying dynamic properties of the collision, the two masses, cause the balls to ricochet in a particular pattern.

This work was supported by Grant MH48494 from the National Institute of Mental Health to Indiana University. Thanks to Robert Nosofsky, Geoffrey Bingham, Jerome Busemeyer, Andrew Hanson, John Kruschke, and Richard Shiffrin for helpful discussions, suggestions, and insights.

Correspondence concerning this article should be addressed to Andrew L. Cohen, Department of Psychology, University of Massachusetts, Amherst, MA 01003. E-mail: acohen@psych.umass.edu

There have been two main, contrasting approaches to how observers recover dynamic properties from the kinematics of an event. The heuristic theorists (Gilden & Proffitt, 1989, 1994; Proffitt & Gilden, 1989; Todd & Warren, 1982) claim that the perception of dynamic properties is primarily based on inference. The basic idea is that observers can use only imperfect, rudimentary motion cues that must be supplemented with high-level cognitive processes that implement their commonsense notions of the physical world. Such cognitive processes are usually assumed to act as a problem-solving or 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 scattered more, that is, reflected at a greater angle.

The direct perception approach 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. Such information is often called an *invariant* because it remains unchanged across most of the situations an organism is likely to encounter. With some experience, organisms are assumed to be able to detect and use invariants directly, without the need for cognitive enhancement (Runeson, 1977a).

THE PERCEPTION OF DYNAMIC PROPERTIES

Because dynamic properties are important for the guidance of action, supporters of the direct perception approach have hypothesized that such properties might also be perceived directly. Heider and Simmel (1944) and Michotte (1946) provided some of the first demonstrations that observers might be sensitive to the underlying causal properties of a visual event. Although these early studies manipulated the kinematics of events to achieve a particular dynamic impression, they did not always create events in accord with an underlying physical interpretation (Runeson, 1977b), and in the work of Heider and Simmel (1944) there was little mention of the link between the kinematics and underlying causal factors of the events (but see Michotte, 1946). The key element in such work as it relates to the theory of direct perception is finding the information that most parsimoniously or, ideally, uniquely links the kinematics of the event to the dynamics. Following the postulation of Runeson et al. (2000) that unique perceptual properties emerge when the "kinematic patterns of unfolding events specify, that is, provide sufficient information about, their underlying dynamic properties" (p. 525), several researchers have attempted to find this link.

There is a body of evidence that observers are adept at detecting properties of human motion that are related to the underlying dynamics (although see Gilden, 1991), such as the identity and gender of a point-light walker (e.g., Cutting & Kozlowski, 1977), the mass of a lifted object (e.g., Bingham, 1987), and the words of a point-light talker (Lachs, 2002). Although these studies have shown that people can perform adequately at perceiving causal properties of an event, the search for specifying information has, in general, proved more complicated. Gilden and Proffitt (1989) argued that without a kinematic theory, it is difficult to fully evaluate these experiments as a whole because it is unclear what constraints have been imposed in the experimental designs and even what designs are necessary to contrast theories of the perception of dynamic properties. For the direct perception approach to gain an experimental foothold, it is necessary to uncover how the kinematics of an event specify the dynamics.

THE COLLIDING BALLS TASK

In contrast to human motion, the relation between the kinematics and dynamics of colliding balls is well defined. Runeson (1977b) showed that the kinematics of colliding balls uniquely specifies the mass ratio of the two balls. A paradigmatic collision is shown in Figure 1. The two balls begin at the top of the figure, proceed down, collide, and ricochet. Create a reference frame for the collision where the point of contact is the origin. Call the line connecting the centers of the balls at the moment of contact the *collision axis* and the perpendicular the *sweep axis*. Let u_{ax} and u_{bx} be the projection of the precollision velocities of Balls A and B onto the collision axis, respectively. Likewise, let v_{ax} and v_{bx} be the projection of the postcollision velocities of Balls A and B onto the collision axis, respectively. From the conservation of momentum equation, and assuming no spin, slippage, energy loss, or friction, it is straightforward to show that the relative mass of the two balls is given by

$$\frac{m_a}{m_b} = \frac{|v_{bx} - u_{bx}|}{|v_{ax} - u_{ax}|}, \tag{1}$$

where m_a and m_b are the masses of Balls A and B, respectively. It is impossible to determine m_a or m_b individually from the kinematics of the collision, only their ratio is specified. Equation 1 defines a dynamic variable of the collision, the mass ratio, solely in terms of the kinematics of the collision, the pre- and postcollision velocities.

Todd and Warren (1982) were among the first to empirically test the idea that observers can directly detect the relative mass of colliding balls. Although sympathetic to the direct perception approach, they concluded that a simpler heuristic provided a better description of human behavior. This research sparked a debate between those who believed that observers can use invariants to determine the relative mass of colliding balls (Runeson, 1995; Runeson & Vedeler, 1993) and those who believed that observers are only able to use heuristics (Gilden & Proffitt, 1989, 1994; Proffitt & Gilden, 1989). After nearly two decades of debate, the two theories proved surprisingly difficult to differentiate experimentally.

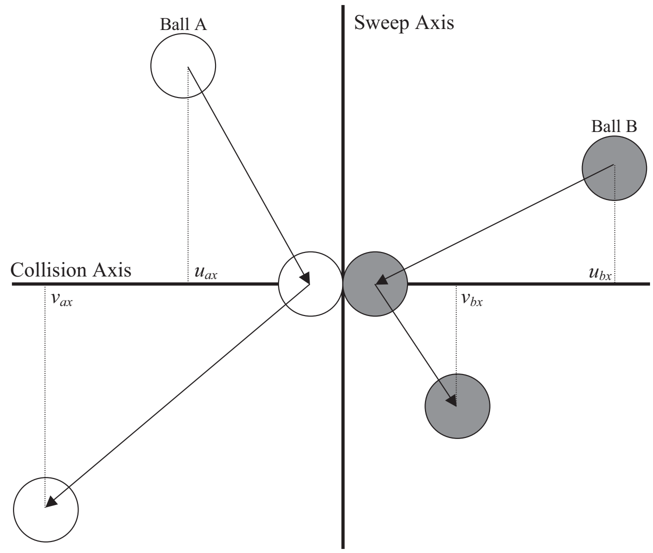

Figure 1. A schematic two-ball collisions. The balls begin at their topmost position and follow the paths defined by the arrows. u_{ax}, v_{ax}, u_{bx}, and v_{bx} are the pre- and postcollision collision axis velocities of Balls A and B, respectively.

The majority of the earlier studies were done without feedback; that is, the observers received no training in the detection of relative mass (for an exception, see Runeson & Vedeler, 1993, Experiment 3, occluded condition). It may be that without training in a potentially unfamiliar task, observers do not yet have the perceptual experience needed to detect the specifying information (Hecht, 1996; Jacobs et al., 2000, 2001; Michaels & de Vries, 1998; Runeson et al., 2000). In a related task, Michaels and de Vries (1998) assessed the ability of observers to perceive the pulling force exerted by human stick figures. Unlike previous experiments, observers in this study received feedback during a training phase. The researchers found a dramatic improvement in performance and a shift from reliance on simple to complex kinematic variables with practice.

Recent evidence (Jacobs et al., 2000, 2001; Runeson et al., 2000) suggests that there might also be such a shift in the perception of relative mass in the colliding balls task. For example, in both Runeson et al. (2000) and Jacobs et al. (2000), observers received feedback during a training phase of a colliding balls experiment. Both studies found a general improvement in the perception of relative mass with training. Before training, observers tended to rely on simple, nonspecifying variables, that is, variables that do not specify relative mass. After training, the data of at least some of the observers were in alignment with the actual mass ratios, which was taken to suggest that the observers had learned the specifying, mass ratio variable. Jacobs et al. (2001) replicated these results, in part, and further suggested that observers may only shift to specifying variables if their initial strategy fails.

With the addition of practice to the experimental repertoire, such studies have provided strong evidence that existing versions of the heuristic theory do not adequately account for posttraining human performance. For example, Runeson et al. (2000) directly compared Gilden and Proffit's (1989) cue-heuristic model and an invariant-based model in a collision learning experiment. Although the pretraining results were fairly well explained by the cue-heuristic model, for many observers, the posttraining results showed clear deviations from this model and a good correspondence with the invariant model. Particular heuristic theories are discussed below.

This recent emphasis on learning, however, also places a burden on those who believe in direct perception. There is currently no clear, well-defined explanation of how observers learn to use the specifying information. Based on the work of Turvey and colleagues (e.g., Saltzman & Kelso, 1987; Turvey, 1990) in the domain of skill acquisition, Runeson et al. (2000) offered a preliminary dynamic systems account of how the specifying information might be learned. According to the theory, the novice observer relies on simple cues and inferences. Practice allows the observer to randomly explore different, potentially more efficient and complex cues. Coupled with a perceptual system flexible enough to detect and use these new cues, the observer eventually settles on information "good enough" to perform the task. The key argument is that observers must learn to use invariants and that until this learning occurs, other, less direct strategies, such as heuristics or exemplars, must be used.

The main prediction of the invariant model as it relates to the present work is that, because the participants are learning to detect mass ratio directly, performance should change globally, across all collisions. An implementation of the invariant model that incorporates this assumption was described by Runeson et al. (2000). This model is formulated in more detail below, but the intuition is that the mass of each collision is picked up directly, as in Equation 1, with noise. Because the noise is assumed constant for all collisions, performance on all collisions with the same mass ratio should be identical, improve with practice, and decrease as the mass ratio nears unity. It is possible to relax this prediction by assuming that collision-specific properties affect the detection of mass ratio, but there is currently no theory that addresses this issue, and such a step would greatly complicate the model.

The prediction that learning is global is hard to evaluate from extant data and, indeed, there are some indications that it may not hold. The collisions in the learning experiments were systematically selected to contrast theories of the perception of mass ratio and may not be representative of the space of possible collisions. Without a representative sample of the collision space, it is difficult to determine the global nature of the learning. Furthermore, experimental results are almost always reported as means collapsed over all collisions of the same mass ratio. This manner of reporting data only illustrates trends and hides performance on individual collisions.

The only exception to the presentation of means in a learning experiment is in Figures 9A and 9B of Runeson et al. (2000). In this figure, reproduced here as Figure 2, the pre- and posttest percentage correct for each collision is plotted against a confidence measure. The present argument focuses on the performance measure on the abscissa. The individual collisions are not labeled, so it is still difficult to assess performance as a function of training, mass ratio, or individual collision parameters, but two features are worth noting. First, there is a clear overall increase in performance with practice. Second, and more importantly, there is at least one collision for which performance *decreases* as a function of training. Performance on this stimulus not only becomes worse with training but also remains well below chance, approximately 15% correct. A straightforward interpretation of the invariant model cannot account for the differential learning of this stimulus.

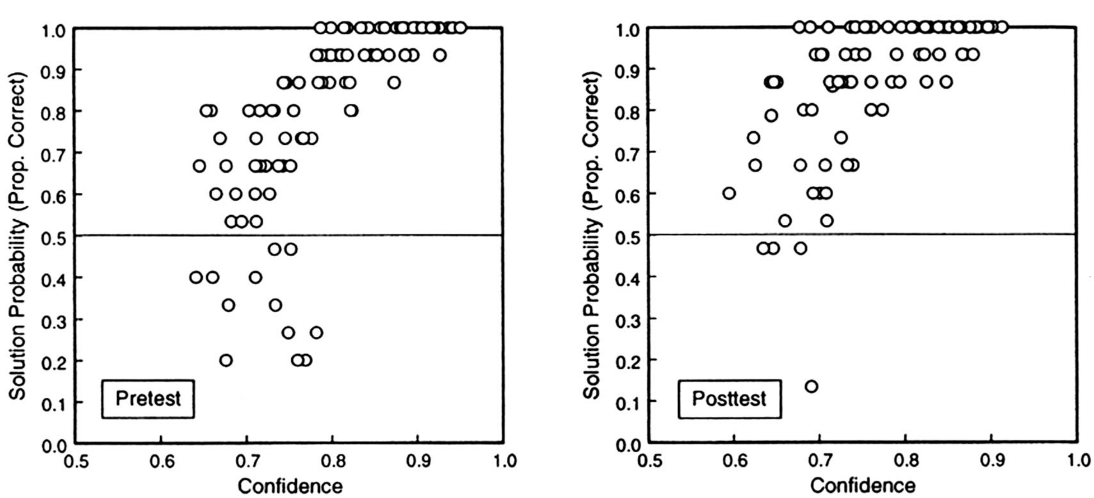

Figure 2. Figures 9A and 9B of Runeson, Juslin, and Olsson (2000). Pre- and posttest measures were taken before and after training, respectively. From "Visual Perception of Dynamic Properties: Cue Heuristics Versus Direct-Perceptual Competence," by S. Runeson, P. Juslin, and H. Olsson, 2000, *Psychological Review, 107,* p. 546. Copyright 2000 by the American Psychological Association. Reprinted with permission.

It is possible to explain this result and remain within the direct perception theoretical framework. One potential explanation is that observers learned to use an "incomplete invariant," so named because the invariant may either only approximate the specifying information or only correctly specify the information under certain circumstances (Runeson et al., 2000; Runeson & Vedeler, 1993). In the present situation, observers may have learned an incomplete invariant that worked for all but the one collision. Hecht (1996) warned against the use of incomplete invariants because, without a well-developed theory, their use gives the direct perception approach extreme predictive power. It is left to future work to explore this possibility.

It turns out that the mass ratio of this collision is near unity (1.17).[1] Thus, another potential explanation for this result is that observers were not confident of their use of the invariant and so resorted to a heuristic that was incorrect for this particular stimulus. That is, observers may have realized that use of the invariant led to an unstable response for this stimulus and so relied on a different process. These data were averaged across 15 observers, and it seems unlikely (although possible) that most observers resorted to a poor heuristic only for this one stimulus. If learning were indeed global, all collisions, including the other three equally close to a mass ratio of 1, should have been affected by training equivalently. The key idea here is that even observers who have learned to rely on an invariant may abandon that strategy if the invariant is difficult to detect. For example, the invariant described by Equation 1 is difficult to detect for mass ratios near unity. Such a possibility is discussed in Experiment 2.

AN EXEMPLAR-BASED APPROACH

Another, more parsimonious explanation of Runeson et al.'s (2000) result is that training affected collisions on an individual or local level. The exemplar-based approach is one of the most widely used learning approaches to perception that incorporates such an assumption. The main tenet of exemplar-based models is that people store particular instances of events in memory, called *exemplars*, and that these exemplars are later retrieved to perform a particular task.[2] Exemplar-based models have been successfully applied to identification (e.g., Lockhead, 1972), old–new recognition (e.g., Nosofsky, 1988), recall (e.g., Hintzman, 1987), problem solving (e.g., Ross, 1987), automaticity (e.g., Logan, 1988; Palmeri, 1997), function learning (DeLosh, Busemeyer, & McDaniel, 1997), and same–different tasks (Cohen & Nosofsky, 2000). The present research continues to explore a potential role of exemplar retrieval by examining performance in the perception of dynamic properties.

Classification (Estes, 1994; Hintzman, 1986; Kruschke, 1992; Lamberts, 1998; Medin & Schaffer, 1978; Nosofsky, 1986) is one of the most successful applications of the exemplar-based approach to perception. In a typical classification task, participants first learn to associate a set of stimuli with one of a discrete number of category labels. In a subsequent transfer phase, participants are asked to assign category labels to a new set of stimuli. According to exemplar models of perceptual classification, people represent the learned categories by storing individual exemplars in memory along with their associated category label and then classify objects on the basis of their similarity to these stored exemplars.

To date, the tasks used to study the perception of dynamic properties have not been viewed as classification tasks. There is reason to suspect, however, that the tasks used to study the perception of dynamic properties might fall under the heading of a classification task and therefore be amenable to similar modeling techniques. In discussing his theory of the perception of dynamic events, Runeson (1977b) postulated that "perception has a descriptive system of its own, consisting of concepts (categories, variables, properties) in which our perception of the environment is structured" (pp. 10–11). This conjecture is mirrored in the surface structure of these tasks in which dynamics are treated as categories of information. For example, in a collision experiment, observers are asked to base their responses on the relative mass of the two balls. When learning is included in the experimental design of a perception of dynamic properties task, observers are trained to associate each of a set of stimuli with a label. Even when learning is assumed to have occurred preexperimentally, observers are then asked to assign one of a discrete set of labels to a new set of stimuli. This experimental design is almost identical to that of a classification task.[3]

The main difference between such a learning task and a typical classification task resides in the nature of the stimuli and their mapping to the labels. Classification stimuli are usually simple, static, unfamiliar visual forms. For example, colors (Nosofsky, 1987), semicircles with radial lines (Nosofsky, 1986), dot patterns (Shin & Nosofsky, 1992), and simple geometric shapes (Nosofsky, 1984) have all been used in classification experiments. Labels are usually assigned arbitrarily in that there is no lawful relationship between labels and individual stimuli. In contrast, the stimuli in the perception of dynamic properties tasks are relatively complex and familiar moving forms in which the mapping from stimuli to label is based on the underlying physics of each stimulus. More recently, exemplar models have been applied to more complex stimuli such as faces (e.g., Busey & Tunnicliff, 1999) and natural language categories (e.g., Storms, De Boeck, & Ruts, 2001). The application of exemplar-based models to the perception of dynamic properties is a continuation of this trend to more complex stimuli.

Recall that, according to exemplar models of classification, category membership is assigned on the basis of the similarity structure of the stored exemplars. If exemplar models are to be successfully applied to the perception of dynamic properties, it is vital that the category structure be tightly coupled to the similarity relationships between events; that is, the relative similarities of the events must be well controlled. The exemplar model would fail, for example, if there were no systematic relationship between kinematic similarity and dynamics. There is preliminary evidence

[1] I thank Sverker Runeson for providing this information. He also pointed out that the major decrement in performance for this collision is isolated to the fourth occurrence in the second posttest block. The question still remains as to why this collision would be differentially affected by factors such as fatigue.

[2] Exemplar models predict that learning an instance affects responses to other exemplars to differing degrees, that is, locally. Note that global learning is also possible in exemplar models. For example, attention may shift to new stimulus dimensions (e.g., Nosofsky, 1986).

[3] When the observers are asked to make a quantitative judgment, for example, judging relative mass in a colliding balls experiment, the experimental surface structure is very similar to a function-learning experiment.

that dynamic events may indeed have a systematic similarity structure. In particular, Bingham, Schmidt, and Rosenblum (1995) found that observers used similar descriptors to describe items with similar underlying dynamics. For example, patch-light displays of both stirred and splashed water were commonly described as “liquid,” “floating,” “water,” and “leaves,” whereas patch-light displays of struck and rolling balls were rarely described by such terms.

The preceding suggests that the perception of dynamic properties may be modeled as a classification task. Both tasks are experimentally treated in the same manner, and dynamic categories seem connected by similarity. In the next section, a formal exemplar-based model for the perception of dynamic properties is proposed, the basic structure of which is closely based on the generalized context model of Nosofsky (1986). The model assumes that an exemplar and its associated feedback or category label are stored for each training collision.[4] The exemplars are assumed to reside in a multidimensional space, the dimensions of which represent the psychological dimensions along which the collisions are compared. Similarity between collisions is assumed to be based on the surface features of the collisions and is a decreasing function of distance in the space. The dimensions of the space are neither restricted to be simple cues as in the heuristic approach nor the specifying information as in the direct perception approach but may lie anywhere in between.[5] The dimensions are based on the kinematics of the collisions, but exactly what dimensions to use is a complex issue that is discussed below. When a test item is presented, it is compared with all exemplars stored in memory, and the judgment is based on its relative overall similarity to each category.

In contrast to the direct perception approach, this model predicts that learning is local to the training items, in which locality is defined by the similarity relations between exemplars.[6] For example, Figure 3 (which anticipates part of the design of Experiment 1) illustrates a schematic of the similarity space between six training and four transfer collisions. The closer two items are in the space, the more similar they are. In the upper left region of the space, Ball A is heavier. In the lower right region of the space, Ball B is heavier. Absolute mass ratio increases with distance from the dividing line. The exemplar model predicts that observers should perform well on Collisions 1, 2, and 3 because they are similar to members of the same category and dissimilar from members of the opposite category. Observers should perform poorly on Collision 4 because it is more similar to members of the opposite category. In the same way, this model could explain the decreased performance in the one collision from Runeson et al. (2000). Of course, the actual similarity relations between collisions are likely to be more complex than in this simple illustration. Experiment 1 explores whether such relations are possible. A pure direct perception model would predict that performance on Collisions 3 and 4 should be similar because they have approximately equal absolute mass.

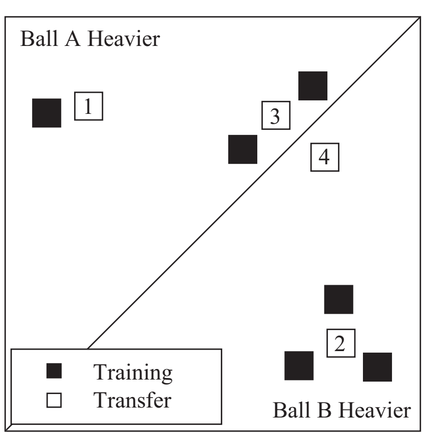

Figure 3. A schematic, hypothetical collision space. Each square represents a collision. Similarity between collisions increases as the distance between the two collisions decreases. Ball A is heavier for collisions to the upper left of the diagonal line, and Ball B is heavier for collisions to the lower right of the diagonal line. Absolute mass ratio increases with distance from the dividing line.

In summary, previous work has demonstrated that before training observers rely on relatively simple heuristics to determine the relative mass of two colliding balls. Practice with feedback allows observers to “tune up” to an invariant that directly specifies the relative mass of the two balls. There is reason to believe that even observers who have learned to rely on invariants may switch to other strategies, such as heuristics or exemplars, when the invariant is below threshold, that is, when the invariant cannot be detected (Kreegipuu & Runeson, 1999; Runeson et al., 2000).

The remainder of this article is organized as follows. The next section formally presents two invariant-based models and an exemplar-based model for the perception of relative mass in the colliding balls task. Experiment 1 qualitatively contrasts the predictions of the three models in a design that tightly couples the similarity relations between training and transfer collisions. Experiment 2 explores what happens when this similarity coupling is relaxed. Heuristic models are also fit to the pretraining data from both experiments. In general, the results of Experiment 1 strongly support the idea that exemplars are a viable strategy for determining the heavier of two colliding balls. The results of Experiment 2 suggest that well-trained observers rely on mass invariants but use less direct strategies when the invariant is difficult to detect.

[4] There are at least two reasons to believe that this experimental framework has relevance to natural human vision and the study of dynamic properties in general. First, as discussed previously, many studies have provided feedback for repeated visual events. This research continues along that line. Second, the modern environment includes circumstances that are very similar to the present experimental framework. In particular, with training, people can become experts at video games in which objects repeatedly collide in idealized physical worlds and feedback is given in the form of success or failure.

[5] Note that the exemplar model also contains processing assumptions. Thus, as discussed later, even if the dimensions include specifying information, the exemplar model is not necessarily identical to or a generalization of the direct perception approach.

[6] In contrast to the learning mechanisms of the invariant approach, the learning mechanisms of the exemplar model are very well specified. This specification is one of the main theoretical advantages of the exemplar approach over the invariant approach. These learning mechanisms are not fully explored in the present work (but see Cohen, 2002, for some suggestions).

MODELS

One of the key contributions of this research is the formulation of quantitative models of the perception of dynamic properties that are faithful to the invariant and exemplar approaches. In this section, two invariant models and an exemplar-based model for the detection of relative mass in the colliding balls task are developed. These three models are designed to predict the heavier of the two balls. Extensions of the models to the quantitative predictions of relative mass will be discussed in Experiment 2. A heuristic model will be developed in the Discussion sections of the experiments.

A Strong Mass Invariant Model

Overview

The strong mass invariant model is a straightforward implementation of the direct perception approach to relative mass detection in the colliding balls task and is based on the invariant model of Runeson et al. (2000). Recall that the relative mass of the two balls in a collision experiment is specified by Equation 1. Although not implemented in the model, it is assumed that training allows the observer to detect relative mass as a perceptual primitive with noise. Specifically, when a collision is presented, the observer detects relative mass as a single, analytically complex variable and makes a judgment based on a single sample from a distribution centered on the true relative mass. The probability that Ball A is judged heavier is given by the proportion of this distribution over the region where Ball A is more massive.

The strong mass invariant model makes two further assumptions. First, perceived relative mass is assumed to lie on a logarithmic scale. There is a host of evidence that lifted weight is perceived on a logarithmic scale going back to Weber's work on the just noticeable difference. Because most of the colliding balls studies select collisions to be equidistant on a logarithmic scale, there has also been a tacit acceptance of the logarithmic scale in viewed relative mass. The possibility that viewed relative mass may best be modeled on a ratio scale is explored in Experiment 2. (The results will suggest that the precise nature of the scale is not critical for the model.) Second, the perceptual noise is assumed to be normal with a single variance term for all collisions regardless of relative mass. Changes in these two assumptions would alter the relative performance of collisions of different relative masses but would not affect comparative performance for collisions of the same relative mass.

This model makes two main predictions. First, as the balls become closer in mass, the judgment process should become noisier. As the relative mass of the two balls approaches one, there is a greater chance that the noise will push the detection process across the equal mass boundary. Second, performance for all collisions with the same relative mass should be equal. This prediction follows from the assumption that the relative mass judgment is based only on the output of Equation 1 and not the specific kinematics of each collision.

This model is a strong implementation of the direct perception view in that observers are assumed to detect and use the output of Equation 1 directly. Weaker versions are certainly possible. For example, observers may learn to use incomplete invariants, or properties of the individual collisions may affect the pickup of relative mass. Unfortunately, these theories are currently unspecified and so are not implemented here. The possibility that each of the individual terms of Equation 1 are picked up separately with noise and then combined is discussed in the Modeling section of Experiment 1.

Formal Properties

The strong mass invariant model, I_S, assumes that an observer can directly perceive the mass ratio of a collision as in Equation 1. The perceived mass ratio is a function of this invariant perturbed by noise. Relative mass is assumed to lie on a logarithmic scale, and the noise is assumed to be normal with mean 0 and variance σ^2. Given these assumptions, the probability that an observer would declare Ball A as heavier on viewing Collision i is given by

$$P(\text{Ball A Heavier} \mid i) = \int_0^\infty N\left(z, \ln\frac{|v_B - u_B|}{|v_A - u_A|}, \sigma^2\right) dz, \qquad (2)$$

where N is the probability density of the normal distribution with mean $\ln(|v_B - u_B|/|v_A - u_A|)$ and variance σ^2 at value z. The mean of the distribution is the natural logarithm of Equation 1. The variance parameter, σ^2, is the one free parameter in the model. Overall performance is coupled to the variance. As the variance decreases, performance increases, particularly for collisions with relative mass nearer to one.

An Angle-Change Invariant Model

Overview

Any model that correctly specifies relative mass must be mathematically equivalent to Equation 1. This formula is not, however, the only means to determine relative mass. For example, Runeson (1995) and Runeson et al. (2000) demonstrated that vector algebra may also be used to arrive at the same value. It turns out that there is another intuitive, geometric way to determine which ball of a two-ball collision is heavier. Consider the collision in Figure 4. At every point in time throughout the collision, draw a dot at the middle of the line connecting the centers of the two balls. When taken over the course of the entire collision, these dots form two straight lines, one before and one after the collision. The two balls do not have to move at the same velocity. These lines are shown in bold in Figure 4. There is a systematic relationship between these two lines and the masses of the two balls. If the precollision line were to continue unchanged, it would take the path as shown by the dashed line in Figure 4. Let α, measured clockwise, be the angle between this line and the actual path. If α is zero, that is, the two lines lie on top of each other, the masses of the two balls are equal. If α is negative, Ball A is heavier. If α is positive, as shown in Figure 4, Ball B is heavier. A proof is given on my Web page (http://people.umass.edu/alc). It is important to note that the degree of angle change, although correlated with relative mass, does not specify relative mass. The direction of angle change specifies which ball is heavier. For example, α can be altered, without

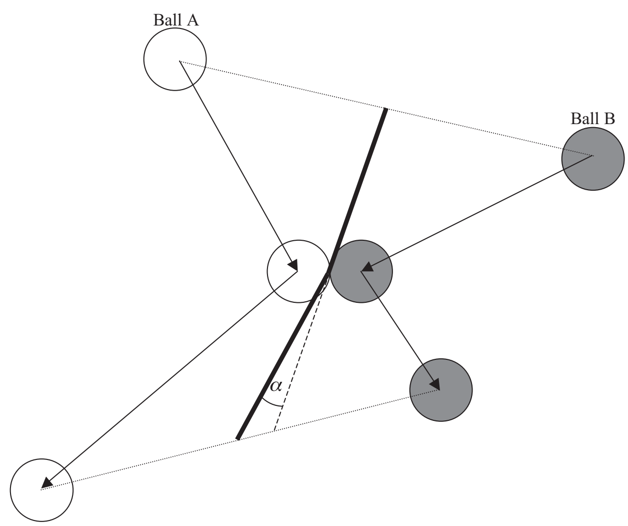

Figure 4. A schematic two-ball collision. The bold lines were generated by plotting the points midway between the two balls at every moment in time. The dashed line is the continuation of the precollision bold line. The angle change between the two bold lines, α, is measured clockwise.

changing relative mass, simply by changing the sweep axis velocities.

The angle-change invariant model follows the same logic as the strong mass invariant model with different information. Training allows the observer to attune to the invariant. Once attunement occurs, the observer detects the direction of α with noise. Note that it is the direction of α that is detected, not relative mass. Because α is not tied directly to relative mass, this model and the strong mass invariant model do not make the same predictions for all collisions. For example, it is possible to have a collision with a high relative mass but low $|\alpha|$. Experiment 1 explores this issue further. The angle change is assumed to lie on a linear scale from $-180°$ to $180°$ and the noise is assumed to be normal, centered on α with variance σ^2. The probability that the observer judges Ball A heavier is the proportion of the distribution where α is negative.

The angle-change invariant model has a number of theoretical advantages over the strong mass invariant model. First, it is intuitive that α may be a perceptual primitive. Rather than an abstract mathematical formula, α is a straightforward geometric quantity of the kinematics of a collision. Second, because α can be explained to observers, it is possible to measure the detection of α independently of relative mass. Independent detection measures allow for the comparison of relative mass using the perceptual, not physical, value of α. Third, this model does not require any sort of coordinate system. The minimum requirement is the pickup of three points: a single bisection point before and after the collision, and the point of collision itself. The major disadvantage of the angle-change invariant model is that the direction of α is not an invariant for relative mass, only for determining which ball is heavier. It is difficult to see how this model could be successfully applied to quantitative judgments of relative mass. This issue is explored further in Experiment 2. Unfortunately, as will be seen, the angle-change invariant model fairs poorly relative to the other invariant models.

Formal Properties

Like the strong mass invariant model, the angle-change invariant model, I_A, assumes that observers base their mass judgment on the direct, noisy pickup of an invariant, the direction of α from Figure 4. The angle change, α, is assumed to lie on a linear scale and the noise is assumed to be normal with mean 0 and variance σ^2. Given these assumptions, the probability that an observer would declare Ball A as heavier on viewing Collision i is given by

$$P(\text{Ball A Heavier} \mid i) = \int_{-180}^{0} N(z,\alpha,\sigma^2)dz, \qquad (3)$$

where N is the probability density of the normal distribution with mean α and variance σ^2 at value z. The angle change, α, can either be assumed from the kinematics of the collision or determined through independent scaling, for example, observers could be asked to directly judge the direction of angle change. The variance parameter, σ^2, is the one free parameter in the model. Overall performance is based on both α and σ^2. Under the assumption that the direction of α is detected correctly, performance increases as $|\alpha|$ increases and σ^2 decreases.

An Exemplar Model

Overview

The exemplar model for the perception of relative mass is based on the generalized context model (GCM) of Nosofsky (1986). According to the GCM, people represent categories by storing individual exemplars of the category in memory and classify test items on the basis of their relative summed similarity to exemplars from each category. Exemplars are represented as points in a multidimensional psychological space. The similarity between exemplars is an exponentially decreasing function of distance in the space (Shepard, 1987). The probability of classifying an item as a member of Category A is based on the relative summed similarity of the item to the exemplars of Category A divided by the summed similarity of the item to all category exemplars.

Applying the GCM to the perception of dynamic property experiments requires two preliminary steps. First, the relevant, contrasting categories must be identified. For example, the categories in the point-light walker experiments (e.g., Cutting & Kozlowski, 1977) would be identity or gender, and the categories of a lifted weight experiment (e.g., Bingham, 1987) would be the stimulus weights. As an initial attempt, the categories in a colliding balls experiment are assumed to be defined by the heavier ball. Thus, there are two categories, the category in which Ball A is heavier and the category in which Ball B is heavier.

The second step is to identify the dimensions from which event similarity arises. These dimensions can be any independent psychologically relevant properties of the events ranging from simple cues to complex variables. A dimension can be based on a static property of the event, such as the relative size of the balls in a colliding balls experiment. It may also be necessary to form the value of a dimension over the course of the event. For example, in Bingham's (1987) lifted weight experiments, peak and average velocities and lift duration are possible dimensions. In a colliding balls experiment, the speeds of the balls, the relative angles of the trajectories, and the difference in exit speeds are some of the potential nonstatic dimensions.

The strongest theory of event similarity would be based on a lawful or theoretically motivated mapping from the event kinematics to the psychological dimensions. Unfortunately, no such theory exists. The exemplar model is usually applied to simple stimuli with well-specified dimensions to avoid the need for such a theory. Even for simple stimuli, however, the dimensions are not always apparent. For example, it is not always clear whether the similarity relations between rectangles are best described by their height and width or area and shape (e.g., Krantz & Tversky, 1975). Identifying psychological dimensions is even more complex for events such as collisions. Even if restricted to psychologically plausible dimensions, the possible number of dimensions is very large. (See Medin, Goldstone, & Gentner, 1993, and Goldstone, 1994, for overviews of the vices and virtues of the use of similarity in cognition and categorization.)

An empirical approach based on multidimensional scaling (MDS) is taken as a first attempt at determining the psychological dimensions of collisions (see Busey & Tunnicliff, 1999, for a similar approach to faces). Untrained observers were asked to rate the similarity, using an unspecified criterion, between each pair of collisions. An MDS analysis was applied to the similarity ratings. MDS (e.g., Shepard, Romney, & Nerlove, 1972) is a technique that attempts to recreate a geometric configuration of objects given only the distances between the objects. For example, given the flying distances between all pairs of U.S. cities, MDS would recover the relative positions of the cities. In a similar fashion, MDS applied to the similarity ratings between collisions can recover the geometric relationship between the collisions.

This model is not a generalization of the invariant model. As formulated here, the two models make different predictions even if the invariant is assumed to be one of the exemplar model dimensions. For example, an experiment could be devised in which observers are given practice determining the heavier of two balls on numerous collisions with mass ratios ranging from 4:1 to 1:1 and from 1:2 to 1:4. Notice the asymmetry in the training regions. Because of the global learning assumption, the strong mass invariant model predicts that observers should be equally good at judging collisions with mass ratios of 1.1:1 and 1:1.1.[7] Even if mass ratio were the sole dimension, the exemplar model would predict that collisions with mass ratios of 1.1:1 would be classified better than collisions with mass ratios of 1:1.1. This prediction follows because the collisions with masses ratios of 1.1:1 are more similar to other collisions with the same feedback. In contrast, the collisions with mass ratios of 1:1.1 are more similar to collisions with the opposite feedback. This extreme, hypothetical example again highlights the global versus local nature of the models.

Formal Properties

Individual collisions are represented as points in an M-dimensional psychological space. Let x_{im} denote the coordinate value of Collision i on Dimension m. The distance between Collisions i and j is given by the weighted euclidean distance metric,

$$d_{ij} = \sqrt{\sum_{m \in M} w_m (x_{im} - x_{jm})^2}, \tag{4}$$

where the w_m ($w_m \geq 0$, $\Sigma\ w_m = 1$) are attention-weighting parameters that serve to stretch distances along diagnostic dimensions and shrink distances along nondiagnostic dimensions (Nosofsky, 1986).[8] The similarity of Collision i to Exemplar j is an exponentially decreasing function of distance,

$$s_{ij} = \exp(-c \times d_{ij}), \tag{5}$$

where c is a sensitivity parameter that scales the distances in the space (Shepard, 1987). If c is high, objects close together in the space will be judged as dissimilar, whereas if c is low, even relatively distant objects will be judged similar. The probability that Ball A is judged heavier on viewing Collision i is given by

$$P(\text{Ball A Heavier} \mid i) = \frac{\sum_{a \in A} S_{ia}}{\sum_{a \in A} S_{ia} + \sum_{b \in B} S_{ib}}, \tag{6}$$

where A and B are the set of all collisions in which feedback was given that Ball A or Ball B was heavier, respectively.[9] Training serves to form the association between individual collisions and their feedback.

There are M potential free parameters in the model, $M - 1$ attention weights, w_m, and the sensitivity parameter, c. Two versions of the model will be fit. The first is a one-parameter restricted version, E_R, with the weights fixed at $1/M$. The second is the full model with the dimension weights free to vary, E_D. Most implementations of the GCM also include bias parameters for the competing categories. Although it is reasonable to assume that ball labels might cause a bias, other collision-specific properties such as ball size (Natsoulas, 1960) and relative initial ball velocity (Runeson & Vedeler, 1993) might result in a bias.

[7] What is being claimed here is that once an invariant is learned, it is used for all mass judgments, not that training has no affect. Indeed, training can affect which invariant (incomplete or not) is learned. In the present example, it is assumed the strong mass invariant model is learned.

[8] It is possible that the euclidean metric is not appropriate for this space; however, it is a fair starting point given the success of this metric for simpler stimuli.

[9] There is a sense in which using the artificial labels attached to the balls, such as Ball A, to assign category membership is problematic. For example, showing the same collision twice with the labels reversed is unlikely to alter a mass judgment. A more complete theory would base the categories on the role each ball plays within the collision. Imagine a situation in which an observer is told that, in Collision *i,* Ball A is heavier. The observer then notices that Balls B and A in Collision *j* act much like Balls A and B in Collision *i*, respectively. The observer should judge Ball B as heavier in Collision *j*, but if only the ball labels are used as categories, Ball A will be judged heavier. This issue can be viewed as one of alignment (e.g., Goldstone, 1998). The experiments reported here were designed to minimize alignment problems by applying the labels as consistently as possible. For example, the orientation of the collision axis and the location of the collision were fixed, and the left and right balls at the moment of collision were consistently labeled across all collisions. Unfortunately, this problem could still arise, for example, in mirror images of collisions. As a first approximation to a solution, it was assumed that if both balls entered from the left in Collision *i* and from the right in Collision *j*, the left ball in Collision *i* was aligned with the right ball of Collision *j*.

EXPERIMENT 1

Experiment 1 has two main goals. The first is to determine whether observers can adopt an exemplar-based strategy to determine the relative mass of colliding balls even when an invariant strategy is viable. The second goal is to contrast the predictions of the exemplar and two invariant models and to collect data suitable for quantitative modeling. The basic idea is to train observers on one set of collisions and then test them on another set that distinguishes the predictions of the three models. An outline of the design is shown in Table 1. Each row of Table 1 represents a set of stimuli on which the relevant model will perform poorly (low percentage correct), and each column holds a set of stimuli on which the relevant model will perform well (high percentage correct). The contents of a cell give the conditions under which these performance measures hold. For example, to obtain low performance from the angle-change invariant model and high performance from the strong mass invariant model, a collision must possess low absolute angle change and high absolute mass ratio. The conditions in three of the nine cells are impossible to satisfy, as it is impossible to have a model predict both low and high performance. Filling in the remaining six cells of the table creates a set of pairwise tests between the three models.

Each of the three models bases performance on a different aspect of the collisions. The strong mass invariant model predicts that performance will increase with absolute mass, $|\ln MR|$, as in Equation 2 (the natural logarithm, ln, is taken to be consistent with past work and later measurements). The angle-change invariant model couples performance to $|\alpha|$ from Equation 3. The exemplar model predicts good performance when sim_C is high and poor performance when sim_I is high. Essentially, sim_C is high if the similarity relations (Equation 5) between collisions lead the observer to make the correct choice. Sim_I is defined likewise, except the similarity relations point to the incorrect choice.

To create transfer collisions with the appropriate sim_C and sim_I values, one must carefully manipulate the similarity relations between the training and transfer collisions (as was done for the example in Figure 3). Similarity between collisions, as given by Equation 5, decreases exponentially, which means that nearby collisions will affect summed similarity much more than distant collisions. Sim_C and sim_I were manipulated primarily by associating each transfer stimulus with a single, perceptually similar training stimulus of the appropriate mass ratio and, at the same time, keeping them relatively distant from the other training collisions. When sim_C was high, the transfer collision and its associated training collision had the same mass relations. When sim_I was high, the two collisions had opposite mass relations.

Recall that the strong mass invariant model bases its predictions solely on the collision axis velocities as in Equation 1. It is possible that a nonzero sweep axis velocity interferes with the detection of the collision axis velocities. To further test the strong mass invariant model under the best possible circumstances, I included three additional transfer collisions that had the same absolute mass ratio and were linear, that is, they had no sweep axis velocity. The similarity relations between these three additional transfer collisions and the training collisions were not systematically manipulated. Because the angle-change invariant model is undefined for linear collisions, this model cannot be tested on the linear collisions.

In addition to the 9 transfer collisions (the 6 from Table 1 and the 3 with no sweep axis velocity), there were 30 training collisions. Training Collisions 1–6 were selected to conform to the appropriate similarity relations with Transfer Collisions 31–36, respectively. To facilitate possible learning of the invariants, I randomly selected the other 24 training items to span the mass ratio and angle-change range and cover the similarity space of collisions. On each trial, the observer viewed a collision and was asked to judge which of the two balls was heavier. The experiment is broken into five parts: pretraining test, training, posttraining test, extended training, and postextended training test. To measure pretraining performance, observers first judged which ball was heavier on all 39 collisions 12 times before receiving any training. They were then trained to criterion on the 30 training stimuli and were afterward asked to make 10 more judgments of all 39 of the collisions. The relatively small number of training items may have biased people toward memorization. To further increase the possibility of learning the invariants, observers were again trained on the 30 training items and an additional 200 new collisions before being tested again on the original 39. To quantitatively assess the similarity relations between collisions, a different set of observers rated the similarity between the original 39 collisions.

Method

Participants

Eighty-two Indiana University Bloomington undergraduate students participated in the similarity-rating task for course credit. Each of the similarity-rating sessions took between 10 and 20 min. Four Indiana University Bloomington graduate students participated in the remainder of

Table 1
Transfer Stimulus Design of Experiment 1

Low P(C)	High P(C): I_S	I_A	E
I_S		$\|\alpha\| \uparrow$, $\|\ln MR\| \downarrow$, Group B	$sim_C \uparrow$, $\|\ln MR\| \downarrow$, Group C
I_A	$\|\ln MR\| \uparrow$, $\|\alpha\| \downarrow$, Group A		$sim_C \uparrow$, $\|\alpha\| \downarrow$, Group C
E	$\|\ln MR\| \uparrow$, $sim_I \uparrow$, Group A	$\|\alpha\| \uparrow$, $sim_I \uparrow$, Group B	

Note. P(C) = predicted proportion correct; MR = the mass ratio of the collision; sim_I and sim_C = the similarity of the collision to the incorrect and correct categories, respectively; α = the angle change of the collision; $\downarrow$ and $\uparrow$ = a low and high value, respectively; I_S = the strong mass invariant model; I_A = the angle change invariant model; E = the exemplar model.

the experiment. Each session took between 45 and 90 min to complete. These 4 participants were paid. The payments are described in more detail in the *Procedure* section. None of the participants were familiar with the issues under study.

Apparatus

The collisions were rendered on a digital computer and displayed on a 28.0 × 20.5 cm monitor at a resolution of 1,024 × 768 pixels and a refresh rate of 85 Hz. To avoid any potential effects of inferences involving "gravity" on the perception of the collisions (Kaiser & Proffitt, 1987), the monitor was placed in a brace that held the screen parallel to the ground plane. Participants sat in a chair looking down at the screen at an angle of approximately 45° and a line-of-sight distance of approximately 50 cm. The participants were not restrained in any way.

Stimuli

Thirty computer-simulated two-ball collisions were *training* collisions and 9 were *transfer* collisions. A separate, *extended* set of 200 collisions was included in the extended training sessions. In each collision, two balls moved across a computer screen, collided, and moved away from each other according to the physical equations given by Halliday, Resnick, and Walker (1993). Because no rotation was visible, the balls looked more like hockey pucks than rolling balls.

The point of collision was the center of the computer screen, and the collision axis was always horizontal on the screen and perpendicular to the participant's line of sight. The participants saw the balls move for 800 ms before and 800 ms after the collision.[10] The balls appeared and disappeared at speed; that is, there was no "natural start" (Runeson, 1974). The balls always had a positive sweep axis velocity. The ball on the left during the moment of collision was red and the ball on the right was blue. The screen was black. Each ball was approximately 1.25 cm in diameter. The balls were assumed perfectly elastic, and no deformation was displayed. No friction or slip was simulated.

To approximate a representative sample from the collision parameter space, the actual collision parameters were selected by hand from a set of pseudorandom collisions that conformed to the design constraints outlined above and illustrated in Table 1. It is possible to find collisions for all six cells of Table 1. Because each transfer collision needs to be related to the training collisions in a very specific way, however, it proved difficult to fill each cell with transfer collisions and still find enough training collisions with the appropriate similarity relations. To cut down on the total number of collisions, certain pairs of the cells in Table 1 were collapsed and two collisions were selected from each group. These pairs of cells are labeled Group A, B, and C in Table 1. What this means, for example, is that whereas the strong mass invariant model predicts good performance for the collisions in Group A, both the angle-change invariant model and exemplar model predict poor performance (compared with the strong mass invariant model) for these same collisions. Figure 5 schematically pairs these six transfer stimuli (Collisions 31–36) with their associated, similar training stimuli (Collisions 1–6, respectively). It should be intuitively apparent from Figure 5 that the training and transfer pairs of stimuli were extremely similar. In Figure 5, the ball represented by the circle or the square is heavier if the collision is labeled with a plus or minus sign, respectively.[11]

All of the collisions were initially selected to have 1 of 10 equally spaced ln mass ratios between −1.25 and 1.25. The ln ratios were −1.25, −0.97, −0.70, −0.42, −0.14, 0.14, 0.42, 0.70, 0.97, and 1.25. There were three collisions with each mass ratio. For each log mass ratio there was at least one collision in which the precollision collision axis velocities of both balls were of the same sign and at least one collision in which these velocities were of different signs. The similarity relations between Training Collisions 1–6 and Transfer Collisions 31–36 were discussed above and are illustrated in Figure 5. The linear Transfer Collisions 37–39 were identical

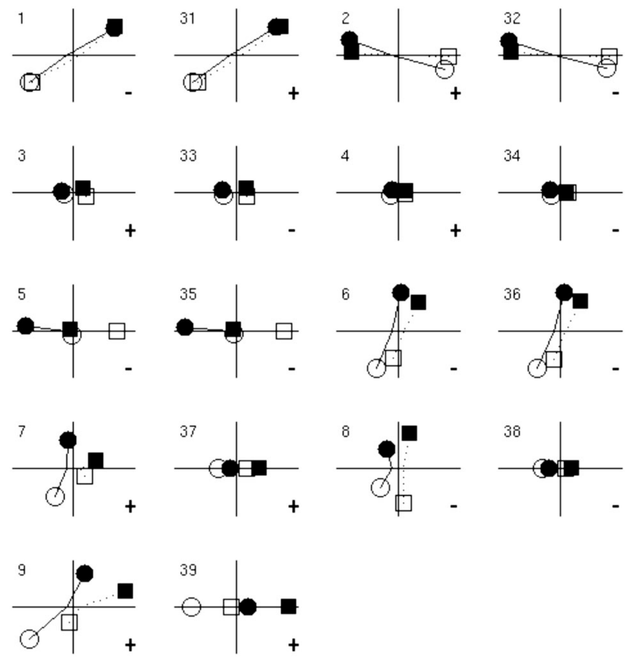

Figure 5. Schematic of the transfer collisions (31–39) and their associated training collisions (1–9). The filled and unfilled shapes represent precollision and postcollision velocities, respectively. The circle and square are the red and blue balls, respectively. The origin of the axes is the point of collision, and distance along the axes represents velocity. The red or blue ball is heavier if the collision is labeled with a plus or minus sign, respectively. The horizontal and vertical axes represent the collision axis and sweep axis velocities of the balls, respectively. The solid and dotted lines represent possible trajectories for the red and blue balls, respectively. (For the lines to represent actual trajectories, the starting positions along the lines would need to be appropriately scaled by the 800-ms pre- and postcollision epoch, pixel size, screen size, etc.)

to Training Collisions 7–9 with the sweep axis velocities set to 0°/s. These collisions are also shown in Figure 5.[12]

In Table 1, high absolute ln mass ratios were defined as above 0.97. Low absolute ln mass ratios were defined as 0.14 or below. These ratios correspond to good and poor performance in prior experiments. There is no good theory of the detectability of this type of angle change, and no such theory is offered here. As an approximation, an angle change over 25° was considered high and an angle change under 2° was considered low. Similarity was defined using Equation 5 in which all attention weights were set equal and $c = 1$. Recall that the psychological dimensions are to be determined from similarity ratings. Similarity ratings between stimuli cannot be collected before the stimuli are chosen, but the stimuli cannot be

[10] For a few collisions with fast moving balls, a ball may have been off the screen for a brief time.

[11] Collisions with high absolute mass ratios and high similarity to the incorrect category can be realized because of the ratio form of Equation 1, which makes it possible to have very small changes in collision axis velocities, the numerator and denominator of Equation 1, but still have a high mass ratio. That is, the numerator and denominator of Equation 1 can be arbitrarily close to each other but still have a very large ratio, for example, $|1.035 - 1.000|/|1.010 - 1.000| = 3.5$.

[12] The collisions parameters for the training and transfer collisions are reported on my Web page.

chosen before the similarity relations are known. To break this cycle, I initially used the six physical velocities as a surrogate for the psychological dimensions in Equation 4. From subjective reports, any similarity value above 0.05 was considered high and any value below 0.01 was considered low. Recall that these similarity values were determined by Equation 5, not the observer ratings (and are on a different scale than the observer ratings).

To summarize the model predictions of the nine transfer stimuli, the strong mass invariant model predicts better performance than either the angle-change invariant or exemplar models for Collisions 31 and 32 (Group A of Table 1) and relatively poor performance (off-ceiling) on the remaining collisions. Likewise, the angle-change invariant model makes the highest percentage correct prediction of the three models for Collisions 33 and 34 (Group B) and relatively low performance on Collisions 31, 32, 35, and 36. The exemplar model makes the highest percentage correct prediction of the three models for Collisions 35 and 36 (Group C) and predicts poor performance for Collisions 31–34. No strong predictions are made for Collisions 37–39.

The extended training set was a collection of 200 collisions selected randomly as outlined above with the following constraints. The collision axis velocities ranged from −26°/s to 26°/s, and the sweep axis velocities ranged from 0°/s to 26°/s. Half of these collisions had precollision velocities of the same sign and half had precollision velocities of different signs. Within each of these sets, there were 10 collisions created from each of the original 10 mass ratios. This set of collisions was designed to span a large portion of the possible collision parameters.

Procedure

Similarity Ratings

The similarity-rating task was run separately from all of the other sessions. To familiarize the participants with the range of possible stimuli, each participant first viewed the 30 training collisions and 9 transfer collisions, one at a time, serially. The participant was then shown a random set of 125 of the possible 741 pairings of 2 different collisions.[13] The 2 collisions in each pair were shown in serial in a random order. As is the case throughout the remainder of the article, stimulus order was randomized separately for each participant. Participants judged the similarity of the 2 collisions on a scale of 1 to 7 (1 = *very dissimilar*, 7 = *very similar*). They were asked to base their judgment on "how similar the two collisions looked to each other" and were urged to use the full range of ratings. Each pair could be viewed as many times as the participant desired before responding.

Mass Judgments

Pretraining test. During each of the following experiment phases, participants were told that they would see two balls move toward each other, collide, and then move away from each other. They were asked to think of the collisions as two steel balls colliding on a steel table. Participants were asked to judge which ball was heavier, that is, which ball would weigh more on a scale. Participants could view the collisions as many times as desired before responding. Frequent breaks were suggested. Different observers participated in the similarity-rating task and all other sessions.

The pretraining phase was broken into two sessions. During each session, each of the 39 training and transfer collisions was seen once per block for six blocks for a total of 234 trials. The order of trials was randomized. No feedback was given. Participants were paid $10 per session.

Training. In all training sessions, observers received feedback on the training collisions. After the feedback, the participant could view the collision as many times as desired. The transfer collisions were also shown during the training blocks, but participants received a message of "Thank You" on these trials. Training was broken into a series of graded sessions; each session was more demanding than the previous one. During the first two sessions of training, participants saw each collision of the 30 training collisions and the 9 transfer collisions once per block for 5 blocks. In the next series of sessions, participants were asked to reach a certain level of training before proceeding on to the next session type. Each block consisted of the 30 training collisions and 9 transfer collisions. Each of these sessions lasted a maximum of 10 blocks. If the criterion was not met by the 10th block, the criterion session was repeated. The criterion only applied to the training collisions. The first criterion was 24 correct per block for 3 blocks in a row. The second and third criteria were 27 correct per block for 2 and 3 blocks in a row, respectively. The fourth criterion was 29 correct per block for 2 blocks in a row. The final criterion was 30 correct for 1 block. Participants were informed of their percentage correct for the most recent block. During the criterion blocks, participants were asked to verify each response before it was registered. Participants were paid $10 per session.

Posttraining test. During two posttraining test sessions, each of the 30 training and 9 transfer stimuli were seen once per block for five blocks. Feedback was given for each of the 30 training stimuli to ensure that participants remembered the training collisions. Feedback of "Thank You" was given for each of the 9 transfer stimuli. Participants were asked to verify each response before it was registered. Participants were paid $10 per session.

Extended training. Extended training was broken into two sessions of two phases. In the first phase, each of the 200 extended training collisions was shown once with feedback. In the second phase, each of 30 training stimuli were shown once per block until the participant reached a criterion of 30 out of 30 for 1 block or until 10 blocks had been completed. Feedback was supplied on every trial. Participants were paid $5 plus a $5 bonus if they reached 90% correct on the extended training collisions.

Postextended training test. During two postextended training sessions, each of the 30 training and 9 transfer collisions were shown once per block for five blocks. Feedback was given on the training trials and a message of "Thank You" was given on the transfer trials. Participants were paid $7 plus a $10 bonus if they got 97% correct on all trials, both training and transfer, across all blocks.

Results

Similarity Ratings

For all further analyses of the similarity data, the ratings were considered independent of the order in which the two collisions were viewed in each pair. To evaluate how consistent the ratings were across participants, I correlated the ratings of each participant with the ratings of the other 81 participants on the appropriate stimuli. The mean correlation coefficient was .57, $t(81) = 39.5$, $p < .001$. Across participants, the ratings were correlated, but there were also strong individual differences. The ratings of only 1 participant were uncorrelated with the others (−.10) and were removed from further analysis.

Each pair was seen an average of 13 times. The collisions give a fair approximation to the desired similarity structure. The mean similarity ratings to pairs (1, 31), (2, 32), (3, 33), (4, 34), (5, 35), and (6, 36) were 6.9, 6.4, 6.6, 6.9, 6.9, and 6.9, respectively. The maximum similarities of Collisions 31–36 to collisions of the undesired category (i.e., training collisions with the opposite mass relations) were 5.7, 5.1, 5.4, 5.1, 4.1, and 5.9, respectively, far below the similarities to the desired categories. There were also some deviations from the desired similarity structure. For example,

[13] The first 6 participants only saw 75 pairs.

it was hoped that Collisions 37–39 would be somewhat distant from all other collisions; instead, the maximum similarities of Collisions 37–39 were 6.8, 6.5, and 6.0 to Collisions 3, 3, and 5, respectively.

An MDS solution for the collisions was derived by fitting the standard euclidean model to the ratings averaged across the participants. The Formula 1 stress values for the 1–6 dimensional solutions are 0.39, 0.23, 0.16, 0.12, 0.10, and 0.09, respectively. These solutions account for 56%, 74%, 82%, 86%, 88%, and 90% of the variance in the data, respectively. There is no clear elbow in these data; however, the four-dimensional solution (given on my Web page) seems to do the best job of describing the data with the fewest parameters and will be used to create the similarity space for the exemplar model.

Because of the coarse nature of the similarity measure and the individual differences in the similarity ratings, it should be emphasized that the MDS solution used here is only a rough approximation to the similarity relations between stimuli. Tests that rely on fine-grained similarity relations will have to wait for more precise similarity measures. The orientation of the solution is discussed in the Modeling section.

Mass Judgments

The pretraining, posttraining, and postextended training data from all participants for Collisions 1–6 and 31–39 are given in Table 2. In Table 2, ln mass ratios less than zero indicate a heavier blue ball, and ln mass ratios greater than zero indicate a heavier red ball. The first two blocks of the first session of the pretraining phase were to allow the participants to become familiar with the task and were not analyzed. Before training, performance for all participants was only moderately in line with the actual mass ratios. Over all 39 collisions, Participants 1–4 reached .72, .65, .63, and .68 proportion correct, respectively, and .80, .67, .70, and .71 proportion correct on the pretraining training collisions, respectively, which suggests that, before training, the invariants were either not used or not used well. All participants learned to classify the training stimuli. Participants 1–4 achieved .99, .99, .98, and .98 proportion correct, respectively, on the training collisions of the postextended training sessions.

Although linear Collisions 37–39 had approximately the same absolute mass ratio, performance varied greatly across the three collisions in contrast to the qualitative prediction of the strong mass invariant model. Before training, all 4 participants performed relatively well on Collision 38 and poorly on Collision 39. Performance was mixed on Collision 37. Performance was much more varied after extended training. Participant 1 had very low performance on Collision 38 and high performance on Collisions 37 and 39. Participant 2 had very low performance on Collisions 37 and 39, but high performance on Collision 38. Participant 3 performed well on Collisions 38 and 39 and relatively poorly on Collision 37. Participant 4 performed poorly on Collisions 38 and 39 and well on Collision 37.

Modeling

Because participants have seen a wider range of stimuli and have had more experience with the collisions, the postextended training data should most strongly benefit the invariant models.

Table 2

Proportion Correct for Each Participant for Fifteen Collisions of Each Phase of Experiment 1

Collision	ln MR	Participant 1			Participant 2			Participant 3			Participant 4		
		Pretraining	Posttraining	Postextended	Pretraining	Posttraining	Postextended	Pretraining	Posttraining	Postextended	Pretraining	Posttraining	Postextended
						Training collisions							
1	−0.45	0	1.0	1.0	.2	1.0	1.0	.2	1.0	1.0	.7	1.0	1.0
2	1.02	.9	1.0	1.0	.8	.4	1.0	.7	1.0	1.0	.2	1.0	1.0
3	0.14	.9	1.0	1.0	.7	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
4	0.14	.5	1.0	1.0	.4	1.0	1.0	.8	.7	1.0	0	1.0	1.0
5	−0.14	0	1.0	1.0	0	.5	1.0	.1	1.0	1.0	.1	1.0	1.0
6	−0.13	1.0	1.0	1.0	.6	.8	.8	.9	1.0	.8	1.0	1.0	1.0
						Transfer collisions							
31	0.98	.9	0	0	.3	0	0	.8	0	0	.1	0	0
32	−1.11	.1	0	0	.7	.3	0	.4	0	0	.7	0	0
33	−0.14	.3	0	0	.6	0	0	0	.2	0	.6	0	0
34	−0.14	.5	0	0	.5	0	0	.3	.3	0	.8	0	0
35	−0.14	0	1.0	1.0	.3	.7	.9	0	1.0	.9	.3	1.0	1.0
36	−0.14	1.0	1.0	1.0	1.0	.9	.8	.7	.8	.8	1.0	1.0	.9
37	0.41	.3	.9	1.0	.5	.3	0	.7	.4	.6	.7	1.0	1.0
38	−0.42	.9	.4	.1	.9	1.0	.9	.7	.8	1.0	1.0	0	0
39	0.42	0	1.0	.9	.3	0	0	0	1.0	1.0	.2	0	0

Note. $MR = Mass_{Red}/Mass_{Blue}$. If $\ln MR < 0$, the blue ball is heavier; otherwise, the red ball is heavier.

Because relative training with each individual stimulus should be slight, the extended training should not greatly affect the predictions of the exemplar model, particularly if the extended training collisions conform to the desired similarity structure of Experiment 1. The additional training will weaken the qualitative predictions of the exemplar model to the extent that these collisions do not conform to this similarity structure. All of the models were fit to the postextended training data by maximizing proportion of variance accounted for (PVAF) on proportion correct over all of the stimuli. (PVAF is given by [SST − SSE]/SST, will be 1.0 if the model accounts for the data perfectly, and will be negative if the model fits worse than the mean.) Because the angle-change invariant model is not defined for linear Collisions 37–39, the PVAF for this model is over Collisions 1–36 only. The modeling results are shown in Table 3. A summary of the fit measures and the best-fitting parameters are provided on my Web page. Because there was little variation in the data, the PVAF is particularly low. The raw fits in Table 3 are more informative for understanding how well the models explain the data.

Invariant Models

Comparison of the data and predictions of the invariant models (I_S and I_A) in Table 3 reveals that neither of these models fits the data particularly well. Given the nature of the training, it is not surprising that these models tend to underpredict performance on the training collisions. That is, participants were urged to memorize the training collisions, but the invariant models were forced to predict performance based solely on the appropriate information value. For this reason, performance on the transfer collisions is more diagnostic.

Contrary to the predictions of both the strong mass and angle-change invariant models (compared with the predictions of the exemplar model), all participants performed well on Collisions 35 and 36. These collisions, however, were very similar to the well-practiced Collisions 5 and 6, and memorization may also have played a key role here, therefore the stimuli on which the invariant models predict good performance provide a stronger contrast. The strong mass invariant model predicts good performance on Collisions 31 and 32, and the angle-change invariant model predicts good performance on Collisions 33 and 34. All participants' performances were at floor on Collisions 31–34.

Because of the strong role of memorization in training, it may seem unfair to fit the invariant models to both the training and transfer data, but the fits become even worse if only the transfer data are used. Because the noise in the invariant models is centered on correct, specifying information, none of the invariant models can predict less than chance performance on any collision. The participants are at floor for approximately half of the transfer collisions. The best the invariant models can do in this situation is to predict chance performance via an infinite noise variance.

A Weak Mass Invariant Model

There is a sense in which the stimuli used to create the high absolute mass ratio and high similarity to the incorrect category collisions from Group A of Table 1 might be unfair to the invariant approach. Each of the balls in these collisions had a very small absolute collision axis velocity change. It may be that observers have learned to detect mass ratio directly but that noise is added in each component of Equation 1 rather than externally (Runeson et al., 2000). According to this weak mass invariant model, I_w, the probability that an observer would declare Ball A as heavier on viewing Collision i is the probability that

$$\ln\frac{|v_{bx} - u_{bx} + X_b|}{|v_{ax} - u_{ax} + X_a|} \tag{7}$$

is greater than zero, where X_a and X_b are random variables distributed normally with mean 0 and variance σ^2 and u_{ax}, v_{ax}, u_{bx}, and u_{bx} are the pre- and postcollision collision axis velocities of Balls A and B, respectively. The variance parameter is the only free parameter in the model. One nice feature of this model is that, unlike the strong mass invariant model, it predicts low performance for collisions of high absolute mass ratios, but with very little collision axis velocity change such as Collisions 31 and 32.[14]

To fit the model, 10,000 simulations were run for each collision.[15] The fit results are given in Table 3. The fit summary and best-fitting parameters can be found on my Web page. Comparison of the Data and I_W columns of Table 3 reveals that, in general, although the weak mass invariant model provides better fits to the data than the strong mass invariant model (most notably on Collisions 31 and 32), it still fits much worse than the restricted exemplar model.

Exemplar Model

Restricted Model

The restricted exemplar model has one free parameter, the sensitivity parameter of Equation 5. The attention weights of Equation 4 are all set to 0.25. Because the exemplar model stores the training collisions as exemplars and they are relatively distinct from each other, it is not surprising that this model predicts perfect training collision performance (compare the Data and E_R columns of Table 3 for Collisions 1–6). Again, the key test lies in the transfer collisions (compare the Data and E_R columns of Table 3 for Collisions 31–39). The exemplar model predicts the data of Participant 1 nearly perfectly, accounting for all of the qualitative aspects of the data, including the transfer collisions. This model does a fair job for Participants 2–4 and fails only for a subset of Collisions 37–39 for each participant. Recall that Collisions 37–39 were the linear collisions that were not experimentally coupled to any of the training collisions.

[14] It should be noted that velocity direction thresholds for sine wave gratings are approximately 5% of the initial velocity (Mckee & Watamaniuk, 1994). The velocity changes for the red balls in Collisions 1, 2, 31, and 32 were approximately 7%, 4%, 6%, and 8%. The velocity changes for the blue balls were approximately 4%, 11%, 15%, and 3%. The velocity changes for one of the balls in Collisions 1, 2, and 32 were at or below this threshold, but both balls are above threshold for Collision 31. The 2 participants who were asked, however, reported that they could distinguish training from transfer collisions before training was complete.

[15] The model was simulated because, at the time, a closed-form solution was not available. After these analyses reported here were completed, it was discovered that it may be possible to derive the predictions of the weak mass invariant model analytically. For example, see Rice (1995).

Table 3
Model Predictions for Each Participant for Fifteen Collisions of Experiment 1

			Participant 1						Participant 2						Participant 3						Participant 4					
				Model prediction						Model prediction						Model prediction						Model prediction				
Collision	ln MR	α	Data	I_S	I_W	I_A	E_R	E_D	Data	I_S	I_W	I_A	E_R	E_D	Data	I_S	I_W	I_A	E_R	E_D	Data	I_S	I_W	I_A	E_R	E_D
Training collisions																										
1	−0.45	−0.3	1.0	1.0	.5	.6	1.0	1.0	1.0	.8	.5	.6	1.0	1.0	1.0	1.0	.5	.6	1.0	1.0	1.0	.8	.5	.6	1.0	1.0
2	1.02	0.4	1.0	1.0	.6	.7	1.0	1.0	1.0	1.0	.6	.6	1.0	1.0	1.0	1.0	.6	.6	1.0	1.0	1.0	1.0	.6	.6	1.0	1.0
3	0.14	20.4	1.0	.8	.7	1.0	1.0	1.0	1.0	.6	.6	1.0	1.0	1.0	1.0	.7	.7	1.0	1.0	1.0	1.0	.6	.6	1.0	1.0	1.0
4	0.14	24.5	1.0	.8	.6	1.0	1.0	1.0	1.0	.6	.6	1.0	1.0	1.0	1.0	.7	.6	1.0	1.0	1.0	1.0	.6	.6	1.0	1.0	1.0
5	−0.14	−1.0	1.0	.8	.8	.8	1.0	1.0	1.0	.6	.8	.7	1.0	1.0	1.0	.7	.8	.7	1.0	1.0	1.0	.6	.7	.8	1.0	1.0
6	−0.13	−1.4	1.0	.8	.6	.9	1.0	1.0	.8	.6	.6	.8	1.0	1.0	.8	.7	.6	.8	1.0	1.0	1.0	.6	.6	.9	1.0	1.0
Transfer collisions																										
31	0.98	1.4	0	1.0	.7	.9	0	0	0	1.0	.6	.8	0	0	0	1.0	.7	.8	0	0	0	1.0	.6	.9	0	0
32	−1.11	−0.3	0	1.0	.6	.6	0	0	0	1.0	.6	.6	0	0	0	1.0	.6	.6	0	0	0	1.0	.6	.6	0	0
33	−0.14	−26.3	0	.8	.7	1.0	0	0	0	.6	.6	1.0	0	0	0	.7	.7	1.0	0	0	0	.6	.6	1.0	0	0
34	−0.14	−30.5	0	.8	.6	1.0	0	0	0	.6	.6	1.0	0	0	0	.7	.6	1.0	0	0	0	.6	.6	1.0	0	0
35	−0.14	−0.8	1.0	.8	.8	.8	1.0	1.0	.9	.6	.8	.7	1.0	.9	.9	.7	.8	.7	1.0	1.0	1.0	.6	.7	.7	1.0	.9
36	−0.14	−1.3	1.0	.8	.6	.9	1.0	1.0	.8	.6	.6	.8	1.0	1.0	.8	.7	.6	.8	1.0	1.0	.9	.6	.6	.8	1.0	1.0
37	0.41		1.0	1.0	1.0		1.0	1.0	0	.8	.9		1.0	0	.6	1.0	.9		1.0	.6	1.0	.8	.9		1.0	.7
38	−0.42		.1	1.0	.9		0	0	.9	.8	.8		0	0	1.0	1.0	.9		0	1.0	0	.8	.8		0	0
39	0.42		.9	1.0	1.0		1.0	1.0	0	.8	1.0		1.0	0	1.0	1.0	1.0		1.0	1.0	0	.8	1.0		1.0	.7

Note. The data are the postextended training test data. Data and predictions are proportion correct. MR = $Mass_{Red}/Mass_{Blue}$; α is the relative amount of angle change in angle-change invariant model and is undefined for the last three linear collisions; I_S and I_W = the strong and weak mass invariant models; I_A = the angle-change invariant model; E_R and E_D = the restricted and full exemplar models.

Dimensions

Before turning to the fits of the full exemplar model, it is necessary to determine how the dimensions of the psychological space were selected. A four-dimensional solution was selected from the MDS analysis of the similarity ratings. For the attention weights of Equation 4 to have psychological significance, it is necessary to rotate the MDS solution so that the dimensions are psychologically relevant. There is currently no way to determine these dimensions a priori. Seven potential dimensions were subjectively determined from an ADDTREE clustering analysis (Sattath & Tversky, 1977) of the collisions based on the similarity ratings. These dimensions were verticality (i.e., how up and down the collision appeared), overall speed, entrance grouping (i.e., how close the two balls were at the start of the collision), asymmetry (i.e., how different the paths of the two balls were), entrance angle, exit speed difference, and absolute speed change difference. Formulas for these potential dimensions are given my Web page. In addition, ln mass ratio was also included as an eighth potential dimension.

To get a general idea of the relation between the MDS solution and the potential dimensions, the MDS solution was rotated to maximally correlate with each individual potential dimension, that is, one dimension at a time. The correlations are high for verticality (.89), overall speed (.92), entrance grouping (.93), asymmetry (.91), entrance angle (.82), and absolute speed change difference (.88) and moderate for exit speed difference (.65). Note that the correlation between the best-fitting MDS dimension and ln mass ratio was very low (.11), suggesting that ln mass ratio was not a psychological dimension. If ln mass ratio were a dimension of the similarity space, the exemplar model would lose much of its appeal because it would, in part, predict mass ratio from mass ratio.

The MDS solution was then rotated so that its four dimensions have the maximal average correlation with each set of four potential dimensions. The best set of four dimensions, D1–D4, was verticality (.87), overall speed (.89), entrance grouping (.92), and asymmetry (.82), respectively. Because of the additional constraints of rotating to fit four dimensions at once, the correlations are slightly lower than for the individual rotations. Overall, the correlations are fairly high. Each dimension was scaled to have a mean of zero and a standard deviation of one.

Dimension Weighting Model

The full exemplar model had four free parameters, the sensitivity parameter of Equation 5 and the three attention weight parameters of Equation 4. The best-fitting parameters can be found on my Web page. Compare the Data and E_D columns of Table 3, particularly for the transfer collisions. There was no benefit of the extra three parameters for Participant 1. By placing most of the weight on Dimension 1, the rest on Dimension 3, and setting the sensitivity high, the full model can account for the pattern of data of Participant 3, including mediocre performance on Collision 37 and high performance on Collision 38. By placing most of the weight on Dimensions 3 and 4 and setting sensitivity high, the full model does a better job of fitting the data of Participants 2 (including low performance on Collisions 37 and 39) and 4 (including lower performance on Collision 39). It is interesting to speculate that each of the participants attended to different dimensions or aspects of the collisions to make relative mass judgments.

Discussion

Invariant and Exemplar Models

Experiment 1 contrasted the predictions of the invariant and exemplar models through manipulation of the information available in the transfer collisions and their similarity relations to the training collisions. In agreement with the exemplar model predictions, all participants performed well on the training stimuli, were at floor on Collisions 31–34, and were near ceiling on Collisions 35 and 36. For a relatively small number of well-learned stimuli, the exemplar model far outperforms any of the invariant models. It seems that, under these circumstances, observers are able to adopt exemplar-based strategies to determine the relative mass of two colliding balls.

The linear Collisions 37–39, however, provide a challenge for the exemplar model in this context. For example, even the full exemplar model cannot account for the good performance of Participants 2 and 4 on Collisions 38 and 39, respectively. It must be emphasized that because the similarity relations between Collisions 37–39 and the training collisions were not controlled, the predictions of the exemplar model are not tightly constrained. This situation does suggest, however, that an exemplar model of relative mass judgment may only be appropriate when the similarity relations between training and transfer stimuli are closely controlled. It is also interesting to note that postextended training data for the only two problematic collisions were almost identical to the preextended training data. That is, Participant 2 was 90% correct on Collision 38 both before and after training, respectively. Similarly, Participant 4 was 20% and 0% correct on Collision 39 before and after training, respectively. Given the individual differences in the similarity ratings, it is possible that these two collisions resided in a relatively distant region of the collision space for Participants 2 and 4 and so they may have abandoned an exemplar process and relied on pretraining strategies. Experiment 2 addresses these issues further.

Heuristic Model

Although this experiment was not explicitly designed to directly contrast the invariant or exemplar models with heuristic models, heuristics are an important class of models and should be considered as a viable alternative to any of the models presented above, especially for the pretraining data before observers have learned to rely on either experimentally defined exemplars or a specifying invariant. A modified version of the heuristic model from Gilden and Proffitt (1989) was fit to the pretraining data. According to this heuristic model, 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 greatest angle is judged to be lighter. Second, according to the velocity heuristic, the ball with the greatest postcollision velocity appears to be lighter. Observers base their decision on whichever of these two heuristics is most salient. Gilden and Proffitt (1989) suggested that angles greater than 90° provide salient angle information and speed ratios over 2 provide salient velocity information. In the present implementation, the angle and velocity ratio that determine salience were parameters that were adjusted to best account for the data.[16] If only one heuristic is salient, the observer uses that heuristic. If both are salient, each observer consistently relies on a single fixed heuristic (particular to that individual). If neither heuristic was salient, the participant was assumed to guess which ball was heavier on each trial.

The model output was first fit to the pretraining data of each participant from Experiment 1. Because both the output of this baseline version of the heuristic model and the experimental data were essentially categorical, the fit measure was the number of categorical responses, "red," "blue," or "guess," in agreement between the participant and model. Each participant's response for each collision was classified as "red" or "blue" if more than two thirds of the 10 responses were "red" or "blue," respectively. Otherwise, the response was classified as a "guess." In contrast to past work (e.g., Runeson et al., 2000), the heuristic fails to account for these pretraining data. The model correctly classifies only 24, 22, 24, and 28 out of 39 collisions for Participants 1–4, respectively.[17] For completeness, the heuristic model was also fit to the posttraining data. The model correctly classifies 28, 31, 29, and 29 of the 39 collisions for Participants 1–4, respectively.[18] As noted above, the relatively poor correlation between the specifying information and pretraining data suggests that the invariant was not used before training. By definition then, some sort of heuristic strategy (other than the Gilden & Proffitt, 1989, model) was used (see also Hubbard, 2004). The unconstrained nature of the heuristic model, however, makes the search for the appropriate heuristics difficult and will be left for future research.

EXPERIMENT 2

There is some evidence that trained observers may use different modes of processing when mass ratio is near and far from one. Kreegipuu and Runeson (1999) trained observers on a set of collisions of various mass ratios. After training, the observers were asked which ball was heavier and to state whether they "inferred" or "saw" this response. The authors found a general change from "inferred" to "saw" responses with practice. They took this result to suggest a switch from an inferential to a perceptual, invariant-based mode of reasoning. Although there was an overall trend toward "saw" responses, observers also reported more "inferred" responses for the collisions of mass ratio near one.

The first goal of Experiment 2 is to explore the possibility that observers may use two modes of reasoning to determine relative mass. In line with Kreegipuu and Runeson (1999), one possibility is that observers may use a more perceptual-based mode for high absolute mass ratios and a more inferential-based mode for mass ratios near one. Alternatively, as suggested by Experiment 1, observers may rely on exemplar processes, but only for those transfer collisions that are highly similar to training collisions.

Although the results of Experiment 1 favored the exemplar model, the similarity relations between most of the transfer and training stimuli were tightly coupled, and those stimuli for which such a relation did not hold were somewhat problematic for the exemplar model. In particular, Collisions 37–39 were not well predicted by the exemplar model for Participants 2 and 4. The second goal of Experiment 2 is to test the generality of these results by exploring a much larger region of the collision space and relaxing the similarity relations between training and transfer collisions. Furthermore, the strong mass invariant model makes the claim that observers can determine not only which ball is heavier but also by how much. To test this aspect of the model, I compared the predictions of the invariant and exemplar models in a mass ratio identification task. Each trial proceeded as in Experiment 1, except that observers were asked to make a quantitative judgment of the relative mass of the two balls. Although this task has been used frequently in recent work (Jacobs et al., 2000, 2001), only averages over collisions of the same mean mass ratio have been reported. In the present research, the models are fit to the performance of each individual collision. Straightforward extensions of the models to quantitative identification of mass ratio are introduced in the Models section and the details are given in the Appendix.

The angle-change invariant model predicted high performance on Collisions 33 and 34 of Experiment 1. All of the participants were at floor on these collisions. Although these two collisions had a relatively high absolute α, $-26.3°$ and $-30.5°$, respectively, later pilot studies found that it might have been subjectively difficult to detect the direction of α. If it is the case that the physical value of α is a poor predictor of the perceived value of α, the modeling evidence against the angle-change invariant model from Experiment 1 is greatly weakened. The final purpose of Experiment 2 is to collect estimates of α independently of relative mass and use these estimates to predict performance in the angle-change invariant model.

Thirty training collisions and 72 transfer collisions were selected to span the collision space, mass ratio range, and α range. On each trial, the participant was asked to label the collision with 1 of 10 mass ratios. The experiment was broken into four phases: pretraining test, training, posttraining test, and angle detection. Judgments were collected before and after a training phase. During training, the participants learned the training collisions to criterion. In an angle-detection phase, participants judged the direction and absolute magnitude of α for each collision.

[16] In Gilden and Proffitt's (1989) Experiment 2, one ball was stationary before the collision and so the scatter angle only applied to the ball in motion. In the present implementation, both balls were in motion before the collision. Scatter was considered salient if one ball scattered more than the other ball by an amount determined by the salience parameter.

[17] The angle and speed salience parameters for Participants 1–4 were 20, 90, 90, and 90 and .1, .1, .1, and .1, respectively. According to the model, all 4 participants relied on angle information if both heuristics were salient. For Participants 2–4, unless there was a conflict, all attention was essentially placed on the velocity ratio heuristic. More than 18 correct is not to be expected by chance using a binomial with $p = 1/3$ and $p = .05$.

[18] The angle and speed salience parameters were 20 and .1, and the default heuristic was angle change for all 4 participants. Note that the overall fits for all 4 participants for both the pretraining and posttraining data changed very little over a large range of parameter values. Representative values are given. For contrast, the strong mass invariant model (fit on PVAF) correctly predicted 34, 25, 34, and 25 of the 39 collisions for Participants 1–4, respectively, where correct was defined in the same manner as for the heuristic model (most of the model "mistakes" were in the correct direction, but classified as guesses). The restricted exemplar model (fit on PVAF) correctly predicted 39, 36, 38, and 38 of the 39 collisions for Participant 1–4, respectively.

Method

Participants

Two Indiana University Bloomington graduate students and one non-Indiana University mathematics professor were paid for their participation in this study. The payment scale is discussed in the *Procedure* section.

Apparatus

The experimental testing conditions were identical to Experiment 1.

Stimuli

There were 30 training collisions and 72 transfer collisions. The collisions were selected as in Experiment 1 under the following additional constraints. The events were selected to span the range of angle changes. The minimum and maximum angle changes were $-103°$ and $81°$, respectively. The mean angle change was $-2°$ with a standard deviation of $25°$. The correlation between angle changes and ln mass ratios was .71, $p < .01$. As a partial replication, the final two transfer stimuli were Collisions 1 and 31 from Experiment 1.

Procedure

Pretraining Test

The participants were given the same general introduction to the collisions as in Experiment 1. The pretraining phase was broken into two sessions. The first session was broken into two parts. To gain a general familiarity with the range of collisions, participants first simply viewed each of the 102 collisions. They were then asked to judge the relative mass of the two balls in each of the 102 collisions once. Participants registered their relative mass judgments via a button press. The 10 buttons corresponded to the 10 mass ratios used to create the collisions. The buttons were labeled by the mass ratios, the percentage mass differences, and labels R5, . . . R1, B1, . . ., B5. Participants could view the collisions as many times as they liked before responding. They were paid $5 for this session.

In the second session, participants saw each of the 102 collisions once per block for four blocks. They received $8 plus a $2 bonus if they reached over 90% correct. No feedback was given in the pretraining sessions.

Training

Each trial of the training sessions proceeded as in the pretraining phase except that participants saw only the 30 training collisions, could change their responses before registering them, received feedback, and after feedback could again view the collisions as many times as they liked. The training phase was also broken into two sessions. During the first session, participants saw each of the 30 training collisions once per block for 10 blocks and were paid $8. In the second session, participants saw each of the training collisions once per block and were tested until they got at most one wrong for two blocks in a row. They were tested in a maximum of 10 blocks per session. If they did not reach this criterion before the end of 10 blocks, the session repeated up to four times. They were paid $15, $12, $10, or $9 depending on which of the four sessions they reached criterion.

Posttraining Test

The posttraining phase proceeded exactly as the second session of the pretraining phase. Participants were paid $8 per session and were given a $2 bonus if they got over 90% of the training trials correct and $2 if they reached over 90% correct overall.

Angle Detection

The task in the angle-detection phase was different. After a verbal explanation of α from the angle-change invariant model, each participant was shown five varied collisions with positive and negative α. Each collision was first displayed with a yellow dot that followed the midway location between the two balls, that is, a dot moving along the bold line in Figure 4. Then the collision was displayed without the dot, and the participants were encouraged to visualize the direction and magnitude of α.

Participants then saw all of the collisions without the yellow dot. They responded whether α angled toward the red or blue ball and by how much. They were asked to determine the size of α and respond on a 5-point scale from 1 (*barely angled*) to 5 (*strongly angled*). No feedback was given. In each of two sessions, participants saw each of the 102 collisions once per block for four blocks. They received $8 plus a $2 bonus if they correctly determined the direction of α on 90% of the trials. None of the bonus criteria in any phase of this experiment was reached (see below for details).

Results

Mass Judgments

The pretest data for Participant 1 are displayed in the three left panels of Figure 6. The pretest data for Participants 2 and 3 were similar and can be found on my Web page (the few individual differences are discussed below). The size of each bubble indicates the proportion of observed ln mass ratio responses for each true ln mass ratio. For example, there are 10 collisions with a ln mass ratio of 1.25 and 10 repetitions for each collision for a total of 100 responses. If a participant responded 1.25 on 50 of those trials, 0.97 on 25, and 0.70 on the remaining 25, there would be two equally sized bubbles at 0.97 and 0.70 and a bubble twice the size at 1.25. Perfect performance would be indicated by a positive slope diagonal of large bubbles. The main benefit of these plots is that they illustrate the variance within each mass ratio. The top-left panel shows pretraining performance for all collisions. The middle- and bottom-left panels show pretraining performance for the training and transfer collisions, respectively.

Participants came to the experiment with some ability to perform the task. The pretest correlations between actual and observed mean ln mass ratio were .63, .71, and .72 for Participants 1–3, respectively. Although the correlations were moderately high, proportion correct was not. Participants 1–3 achieved .15, .21, and .20 proportion correct over all collisions, respectively. For the training collisions, Participants 1–3 reached .18, .18, and .21 proportion correct, respectively. Note that there was sometimes a very high variance within an individual ln mass ratio. Some of the pretraining responses indicated systematic, extreme judgment deviations. For example, Participant 2 often predicted a ln mass ratio of 1.25 for a collision with a ln mass ratio of -1.25. As these extreme mistakes occur relatively often—for instance, across the 3 participants, there were 23 cases in which the mean response to a particular stimulus was 4 or more scale points from the correct response—it is unlikely that these deviations are just the result of an occasional mispressed button.

The posttraining data for Participant 1 are illustrated in the three right panels of Figure 6 (and on my Web page for Participants 2 and 3). The posttest correlations between actual and observed mean ln mass ratio were .74, .79, and .79 for Participants 1–3, respectively. The overall posttraining proportion correct for Participants 1–3 were .29, .33, and .33, respectively. Percentage

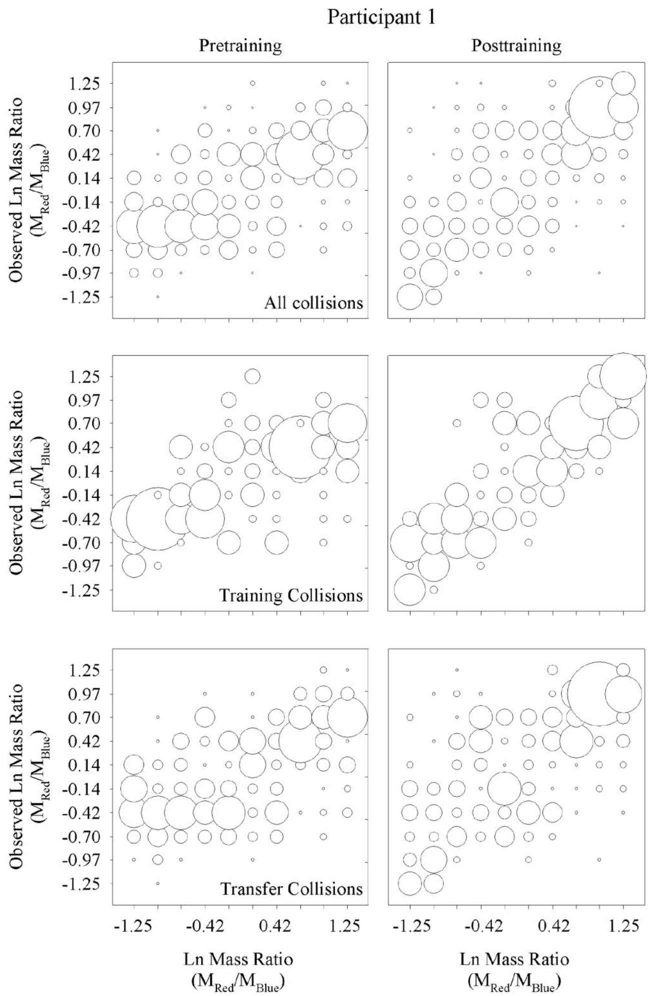

Figure 6. Pre- and posttraining data for all collisions, training collisions, and transfer collisions from Experiment 2 for Participant 1. The size of each bubble indicates the proportion of ln mass ratio responses given for each true ln mass ratio.

correct is a relatively insensitive measure of learning in this task because it does not take the magnitude of the mistake into account. A more sensitive measure is the absolute deviation from correct. The absolute deviation from the correct response was computed for the mean of each collision for pre- and posttraining. Then, for each collision, the pretraining deviation was subtracted from the posttraining deviation. A negative score would indicate learning. For Participant 1, the mean difference score for the training and transfer collisions were −0.74 and −0.28, respectively. For Participant 2, the difference scores were −0.84 and −0.36, respectively. For Participant 3, the difference scores were −0.07 and 0.05, respectively. A *t* test revealed a significant improvement in the training stimuli for Participant 1, $t(29) = 3.51$, $p < .01$, and Participant 2, $t(29) = 4.76$, $p < .001$, but not Participant 3, $t(29) = 0.24$, $p < 1$, and for the transfer stimuli of Participant 1, $t(71) = 1.97$, $p = .05$, and Participant 2, $t(71) = 2.11$, $p < .05$, but not Participant 3, $t(71) = 0.27$, $p < 1$.

All 3 participants learned to judge the relative mass of the training collisions very well. Participants 1 and 2 also performed well on the transfer collisions. Participant 3's quantitative performance on the midrange transfer stimuli actually declined, as these judgments were relatively extreme. It is interesting that all 3 participants showed improved performance on the collisions with extreme mass ratios and decreased performance on collisions with moderate or low absolute mass ratios, perhaps owing to a greater willingness to use more of the rating scale.

Angle Detection

Although she claimed to understand the instructions and tried to do well, Participant 1 had a very difficult time with the angle-detection task. Her correlation between the true angle and the average judged angle was only .08, $p < .05$. Participant 2 did well at this task and achieved a correlation of .58, $p < .01$. Participant 3 reached a correlation of .22, $p < .001$. Participants 1–3 detected the correct angle-change direction on .58, .75, and .62 proportion of the trials, respectively, $t(815) = 33.69$, $p < .001$, $t(815) = 49.77$, $p < .001$, and $t(815) = 36.28$, $p < .001$. There are clear individual differences in this task. The formal modeling will help determine if the differences in angle detection relate to differences in mass judgments.

Modeling Results

Extensions of the models from Experiment 1 applied to the quantitative prediction of relative mass were fit to the data using a PVAF measure. The strong mass ratio invariant model I_{SR} is similar to the strong mass invariant model with the altered assumption that perceived mass ratio lies on a ratio, not a logarithmic, scale. As it would require collecting data for 5,151 cells of a similarity matrix, the number of stimuli made it impractical to construct an MDS space for the collisions in Experiment 2. Instead, the collisions were assumed to reside in a similarity space defined by the four dimensions (verticality, overall speed, entrance grouping, and asymmetry, as defined on my Web page) selected for Experiment 1. Details of these models are given in the Appendix. The fits for Participant 1 are illustrated in Figure 7. The qualitatively similar fits for Participants 2 and 3 are given on my Web page. A fit summary is outlined in Table 4, and the best-fitting parameters are given on my Web page.

The mass invariant models did a fair job of predicting the data. There are a few trends worth noting. First, comparison of columns I_S and I_{SR} in Table 4 reveals that there was only a slight change in fit when mass ratio was switched from a logarithmic to a ratio variable, so the strong ratio mass invariant model is not discussed further. Second, as can be seen in Figure 7 (Panels I_S and I_{SR}), the strong mass invariant models cannot quite predict the range of responses in that the predictions tend toward the mean. Third, although the overall fit was good, the strong mass invariant models predict a few extreme systematic errors for Participant 2. In particular, one stimulus with a ln mass ratio of −1.25 and a predicted ln mass ratio of −0.87 had an observed ln mass ratio of 1.18. This collision was very similar in form to Collisions 31 and 32 of Experiment 1. The weak mass invariant model does a better job of predicting the range of responses and does not suffer from such extreme mispredictions (cf. Panels I_S and I_W of Figure 7). The angle-change invariant model was not able to predict the data of any participant, accounting for at most 37% of the variance.

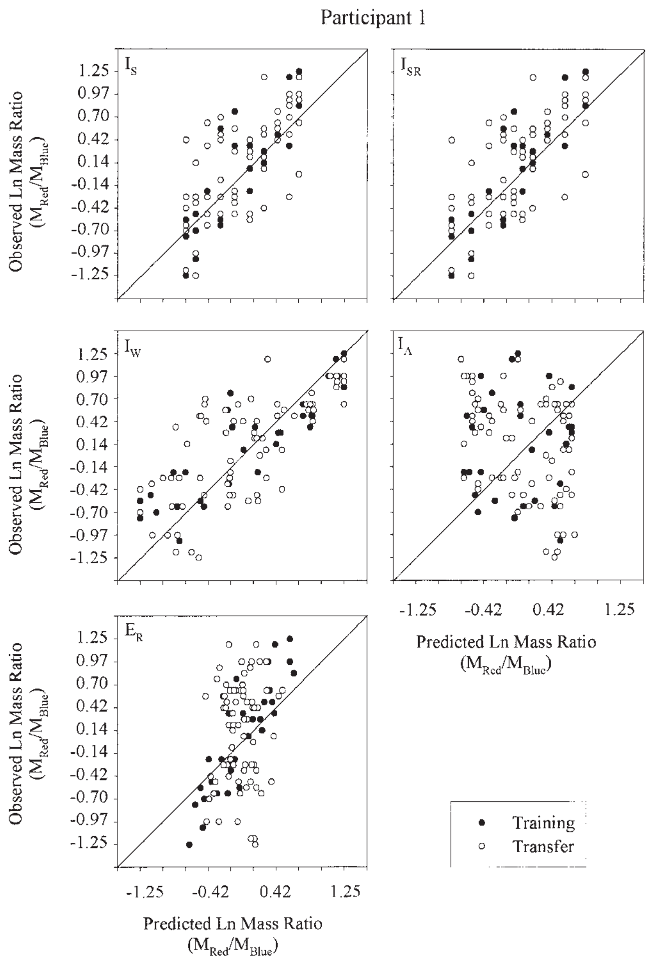

Figure 7. Model fits to the posttraining data from Experiment 2 for Participant 1 for the strong, strong ratio, and weak mass invariant models, I_S, I_{SR}, and I_W, respectively; the angle-change invariant model, I_A; and the restricted exemplar model, E_R. A perfect fit would lie along the diagonal.

Likewise, the restricted exemplar model also fails to predict the data, accounting for at most 25% of the variance. Although the restricted exemplar model well predicted trends in the training data (row $PVAF_{Train}$ of Table 4), the model failed on the transfer data (row $PVAF_{Trans}$ of Table 4) because it was forced to fit the transfer collisions using only the sparse and relatively dissimilar training collisions.

Mixture Models

As discussed previously, observers may use two modes of reasoning to determine the relative mass of colliding balls. To explore such a possibility, I also fitted a second class of models, identical to those fit above, but augmented with the observers' pretraining data, to the posttraining data. The basic idea is that for any of the models discussed above, the probability that an observer will rely on pretraining strategies, as reflected by their pretraining data, is based on the strength of the output of the original model. For example, for the strong mass invariant model, as the value of Equation 1 decreases, the probability that an observer will rely on whatever strategy they used before training will increase. That is, observers might learn to use the invariant but revert to their pretraining strategies when their confidence in the invariant is low (the data in Figure 1 suggest that posttest confidence and performance are positively correlated). Juslin and Olsson (1997) suggested that confidence might be low if multiple samples of the invariant are inconsistent. Likewise, observers may rely on an exemplar process but revert to their pretraining strategies when a transfer collision is dissimilar from all of the training collisions.[19] A potential heuristic model of pretraining strategies is discussed in the Discussion section. Details of these mixture models are given in the Appendix. The pretraining data (the pretraining-strategies-only model P) can be used to predict the posttraining data. Each mixture model adds additional processing assumptions beyond the pretraining data alone; that is, the pretraining data are a component of each mixture model. Thus, there is a benefit of these additional processing assumptions to the extent that the PVAF of the mixture model is greater than the proportion of the posttraining data predicted by the pretraining data.

It is important to keep in mind that the focus of the present research was on how observers *learn* to detect dynamic properties, not what strategies they already possess. It is uncertain whether a model of pretraining strategies beyond the use of the pretraining data would benefit the study of learning in the present context. In fact, a model of pretraining performance, although interesting in its own right, might actually act to hide trends in the data that indicate learning. For example, if both the pre- and posttraining performance of Collision *i* are predicted to be high, but Collision *i* is an idiosyncratic exception to some pretraining model and pretraining performance is actually low, the pretraining model might mask the effect of training. For these reasons, the incorporation of pretraining data into the mixture models is the best choice to selectively uncover the effect of training. Of course, a more complete explanation of relative mass detection would need to account for pretraining data. The heuristic models, discussed below, were fit to the pretraining data as a first step in this direction.

The fits of the mixture models for Participant 1 are illustrated in Figure 8 (and on my Web page for Participants 2 and 3). A fit summary is outlined in Table 4 (and the best-fitting parameters are given on my Web page). The angle-change invariant model still cannot account for either the training or transfer data. Indeed, the best this model can do is to rely almost solely on the pretraining data (compare columns I_A and P of Table 4). The other three mixture models, however, improve the fits over the pretraining strategies alone. The main advantage of these models over the pretraining-strategies-only model is in the prediction of the training collisions (rows $PVAF_{Train}$ in Table 4). The mass invariant models also somewhat improve the fits of the transfer collisions over the pretraining strategies model, especially the weak mass invariant model for Participants 2 and 3. The main advantage of these models over their invariant-only submodels is reflected in the transfer data (rows $PVAF_{Trans}$ in Table 4). The exemplar model

[19] A mixture of exemplars and pretraining strategies may imply a type of rule-plus-exception model (Erickson & Kruschke, 1998; Nosofsky, Palmeri, & McKinley, 1994; see also Allen & Brooks, 1991, for evidence that exemplars affect performance even when an easy, perfectly predictive rule is available).

Table 4
Proportion of Variance Accounted for Over All Collisions ($PVAF_{All}$), the Training Collisions ($PVAF_{Train}$), and the Transfer Collisions ($PVAF_{Trans}$) for the Best-Fitting Models for All Participants of Experiment 2

	Model										
Statistic	I_S	I_{SR}	I_W	I_A	E_R	T	I_{SP}	I_{WP}	I_{AP}	E_{RP}	P
Participant 1											
$PVAF_{All}$	0.52	0.54	∼0.41	−0.88	0.15	0.30	0.70	∼0.72	0.60	0.69	0.60
$PVAF_{Train}$	0.71	0.74	∼0.60	−0.52	0.63	0.66	0.73	∼0.77	0.53	0.79	0.53
$PVAF_{Trans}$	0.43	0.44	∼0.31	−1.05	−0.08	0.13	0.69	∼0.69	0.63	0.64	0.63
Participant 2											
$PVAF_{All}$	0.63	0.61	∼0.68	0.37	0.20	0.55	0.71	∼0.79	0.63	0.68	0.62
$PVAF_{Train}$	0.87	0.86	∼0.88	0.47	0.75	0.91	0.88	∼0.92	0.73	0.94	0.74
$PVAF_{Trans}$	0.51	0.49	∼0.57	0.32	−0.08	0.37	0.63	∼0.72	0.58	0.55	0.57
Participant 3											
$PVAF_{All}$	0.61	0.59	∼0.60	0.37	0.25	0.48	0.65	∼0.66	0.51	0.62	0.50
$PVAF_{Train}$	0.82	0.77	∼0.76	0.13	0.79	0.81	0.79	∼0.79	0.39	0.84	0.41
$PVAF_{Trans}$	0.53	0.49	∼0.51	0.43	−0.03	0.34	0.60	∼0.61	0.56	0.56	0.54

Note. I_S, I_{SR}, and I_W = the strong, strong ratio, and weak mass invariant models; I_A = the angle-change invariant model; E_R = the restricted exemplar model; T = the true model; I_{SP}, I_{WP}, I_{AP}, E_{RP} = the appropriate models augmented with pretraining strategies; P = the pretraining-strategies-only model. The fits for I_W and I_{WP} are approximate because they were simulated.

drastically improves the fits of both the training and transfer collisions over the exemplar process-only submodel. This improvement is quite marked for the transfer data in which, for example, the PVAF increased from −0.08 to 0.64 for Participant 1.

Across all 3 participants, the fits of the mixed mass invariant and the exemplar models are very similar. The main difference is that the exemplar model fits the training collisions better than the invariant models and the invariant models fit the transfer collisions better. For Participant 2, the strong mass invariant model and the exemplar model produce numerically similar fits; however, the former model still suffers from a few extreme mispredictions.

Discussion

Invariant and Exemplar Models

Experiment 2 explored a large region of the collision space and relaxed the similarity relations between training and transfer collisions. Neither the angle-change invariant model[20] nor the restricted exemplar model was able to account for the data. In contrast to Experiment 1 and in line with past work (e.g., Runeson et al., 2000), the most parsimonious explanation of the single-process modeling results is that the participants relied on a process akin to the weak or strong mass invariant model. The weak mass invariant model gives the best explanation of the data without the extreme mispredictions that occur for the strong mass invariant model. None of the single-process models, however, do better than an overall moderate job of predicting the data. Furthermore, the only advantage of these models over the use of pretraining data is in the fit of the training stimuli. This result is perhaps not surprising, as the task required participants to learn the training collisions and all but the angle-change invariant model predict good performance on these collisions. The relatively poor fit on the transfer collisions suggests that all of these single-process models are lacking an essential underlying process.

The substantial improvement in fit for the mixed models, especially on the transfer stimuli, suggests that participants may have reverted to their pretraining strategies when either confidence in the invariant was weak or a collision was dissimilar from the training collisions. This possibility is especially promising for the transfer collisions in which extreme mispredictions in the data continued even after extensive training. Looking back at Experiment 1, Participants 2 and 4 may indeed have used two processes, an exemplar process for the tightly coupled Collisions 31–36 and a different strategy on Collisions 37–39. The fit of the mixed-exemplar model to the transfer collisions, however, is entirely due to the addition of pretraining stimuli. The pretraining data account for .99 proportion of the variance for the transfer collisions for all 3 participants. One of the main strengths, and indeed one of the key benchmarks, of exemplar models has been its ability to predict performance on transfer items (e.g., Nosofsky, 1986). Thus, although this analysis does not necessarily rule out the exemplar

[20] Given the apparent individual differences between participants in the angle detection task, it might be interesting to explore whether those observers who can detect angle change can also be trained to use the information to determine qualitative judgments of relative mass. If observers can be taught to base their mass ratio judgments solely on angle change, presumably, their performance in mass detection should match their performance on angle-change detection.

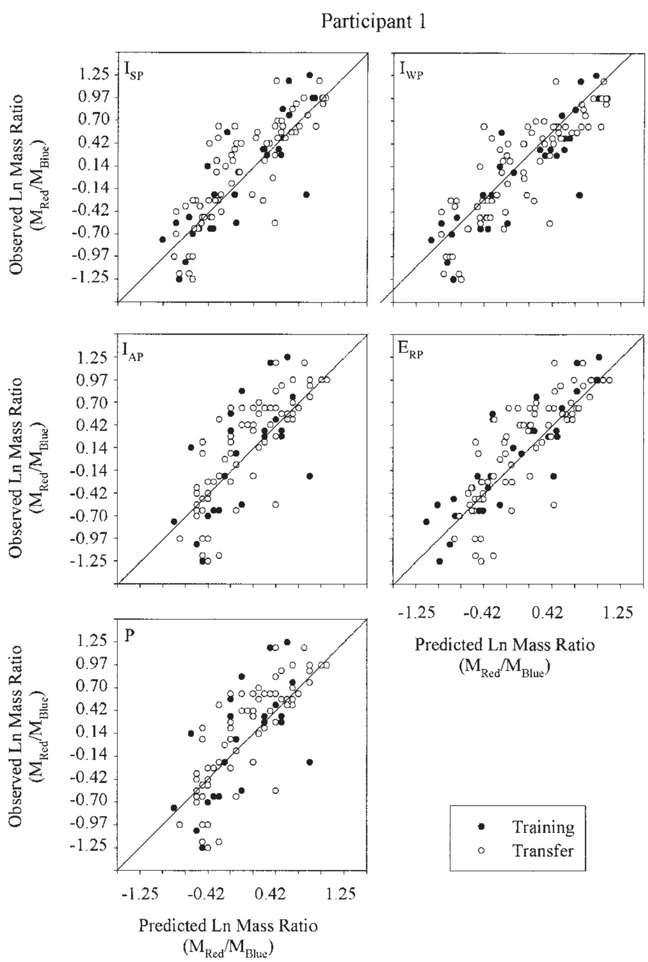

Figure 8. Model fits to the posttraining data from Experiment 2 for Participant 1 for the strong and weak mass invariant models, I_{SP} and I_{WP}, respectively; the angle-change invariant model, I_{AP}; and the restricted exemplar model, E_{RP}, with the inclusion of the pretraining strategies and the pretraining-strategies-only, P, model. A perfect fit would lie along the diagonal.

model (it may be that participants memorized the training collisions and relied almost wholly on pretraining strategies for the transfer collisions), its appeal is greatly diminished.

The dual-process invariant models fair better at predicting the transfer data. For example, the pretraining data account for only .81, .74, and .83 proportion of the variance for the weak-invariant transfer collision predictions for Participants 1–3, respectively. A partial correlation analysis was run to assess the contribution of the weak-invariant process to the prediction of the transfer data above and beyond that of the pretraining data. The correlation between the best-fitting weak-invariant predictions and the posttraining data with the pretraining data partialed out was .47, .63, and .44 for Participants 1–3, respectively, demonstrating that the weak invariant does provides a substantive improvement over the pretraining data. In summary, although these data do not allow sharp discriminations between the mixed mass invariant and exemplar models, the weight of the evidence suggests that observers may rely on processes akin to those in the dual-process mass invariant models.

It must be noted that even the best-fitting model accounted for less than 80% of the variance in the data, suggesting that there are other important aspects of the data not accounted for by the models developed and tested here. Recall that all of the participants improved on only those transfer collisions with high absolute mass ratios. One possibility is that with training, participants simply learned to use the response scale more effectively. In particular, before training, participants may have been reluctant to use the edges of the response scale. Indeed, during debriefing, Participant 2 said that she thought a 1:3.5 collision would look more extreme than any of the collisions she viewed and so almost never used that response. During training, she learned how extreme a 1:3.5 collision appeared and it helped her scale her responses. Such a process could help account for the change in transfer collision performance from pre- to posttraining without relying on the use of an invariant. An experiment with extreme mass ratio transfer collisions might help determine whether participants used such a response strategy.

Heuristic Model

The heuristic model described earlier was fit to the pretraining data of each participant from Experiment 2. Recall that the data from Experiment 2 were mass ratio judgments. As in Gilden and Proffit (1989), perceived mass ratio was assumed to be proportional to salience. For example, if two decisions are based only on velocity ratio and the ratio in the first collision is higher than in the second, there will be a perceived higher mass discrepancy in the first collision. A third parameter was fit to place angle difference and velocity ratio on the same scale. The R^2 between the data of Participants 1–3 and this heuristic model were .53, .27, and .43, respectively.[21] In contrast, the PVAF for the best-fitting strong invariant model for the pretraining data were .39, .49, and .50 for Participants 1–3, respectively.[22] The R^2 of the best-fitting heuristic model to the posttraining data were .23, .30, and .29 for Participants 1–3, respectively, far worse than the strong mass invariant model.[23] These results suggest that whereas Participant 1 may have relied on heuristics before training, none of the participants solely relied on the heuristics after training. The possibility still remains, however, that heuristics could have been used in conjunction with invariant or exemplar processes, perhaps as part of the pretraining data in the mixture models, or that other, as yet unspecified, heuristics may have been used.

---

[21] The best-fitting angle difference salience, velocity ratio salience, and scaling parameters for Participants 1–3 were 3.46, 1.18, and 16.29; 12.93, 1.12, and 28.05; and 2.92, 1.44, and 14.35, respectively. If both heuristics were salient, the better default heuristic was velocity ratio for Participants 1 and 3 and angle difference for Participant 2. Note that because these model predictions and the data were on a different scales, the fit value was calculated as R^2, not (SST − SSE)/SST.

[22] The best-fitting variance parameters for Participant 1–3 were 1.37, 0.98, and 1.04, respectively.

[23] The best-fitting angle difference salience, velocity ratio salience, and scaling parameters for Participants 1–3 were 2.13, 1.44, and 13.60; 7.92, 2.41, and 0.00; and 35.52, 1.13, and 39.91, respectively. If both heuristics were salient, the better default heuristic was velocity ratio for Participants 1 and 2 and angle difference for Participant 3.

GENERAL DISCUSSION

The main focus of this research was to explore the potential contributions of invariants, heuristics, and exemplars to the perception of dynamic properties in the colliding balls task. The invariant models were based on the direct perception approach and assumed that observers can learn to detect and use complex, task-specifying information. Such invariants are perceived with noise. The differences between the invariant models lie in the type of information detected and the placement of the noise. The strong mass invariant model assumes that observers can pick up mass ratio as a perceptual primitive and that noise is added externally. A similar process occurs in the weak mass invariant model, except the individual component velocities that specify mass ratio are detected with noise and then combined. The angle-change invariant model bases its predictions on the noisy detection of the pre- to postcollision change in angle of the moment-to-moment midway point between the two balls.

The heuristic approach assumes that observers only have access to simple motion cues that must be supplemented with higher level cognition. Following Gilden and Proffitt (1989), the heuristic model assumes that participants base their relative mass judgments on whichever of two sources of information is more salient. First, according to the angle heuristic, the ball that scatters at the greatest angle is judged to be lighter. Second, according to the velocity heuristic, the ball with the greatest postcollision velocity appears to be lighter.

The exemplar model is built directly from the generalized context model of classification performance (Nosofsky, 1986). Particular instances of training collisions are stored in memory as exemplars along with their associated feedback. At test, a similarity measure is used to compare the transfer collision to the stored exemplars. Predicted performance is a function of the summed similarity of the transfer collision to each competing category of response. The exemplar approach essentially treats this task as a mapping from similarity relations to output. In this way, the perception of dynamic properties is treated much the same as a classification, identification, or function-learning task. Two-process models were also formulated in which predicted performance was a mixture of the original process and the pretraining data.

Experiment 1 explored the possibility that observers could rely on an exemplar process to determine the relative mass of colliding balls and allowed for both qualitative and quantitative contrast between two invariant models, a heuristic model, and an exemplar-based model. The transfer collisions of Experiment 1 were selected to contrast the predictions of the exemplar and two invariant models. Each of the three models predicted high performance for exactly two distinct transfer collisions. Neither the invariant models nor the heuristic models could account for these data. Whereas the restricted exemplar model did an excellent job accounting for these data, it did have trouble predicting the data for three linear collisions for Participants 2 and 4. It was argued that these collisions might have resided in an isolated region of the collision space for these participants and were possibly best modeled by the pretraining data.

Experiment 2 explored this suggestion further in a mass ratio judgment task by relaxing the similarity relations between the training and transfer collisions. The single-process mass invariant models did a fair job accounting for these data. The angle-change invariant model, the restricted exemplar model, and the heuristic model performed very poorly on the posttraining data. The heuristic model did a fair to moderate job of predicting the pretraining data. The addition of the pretraining data helped the mass invariant and the exemplar models considerably. It was argued that the weak mass invariant model augmented with pretraining data when the invariant was difficult to detect best accounted for the data.

The data collected in these two experiments were rich enough to begin to contrast computational models of the perception of dynamic properties and to suggest that exemplars, invariants, and heuristics all have a role to play in this task. Indeed, in this regard, these data are in general agreement with much of the past work on the detection of dynamic properties in collision events. On the one hand, experiments using untrained observers and a relatively limited range of collisions have found good agreement with simple heuristic models (Gilden & Proffitt, 1989; Runeson et al., 2000; Todd & Warren, 1982). In Experiment 2, there was some limited evidence for the use of the Gilden and Proffitt (1989) heuristic model before training. On the other hand, judgments of trained observers or observers with exposure to a larger set of collisions seem in better agreement with an invariant approach (Jacobs et al., 2000, 2001). Of the single-process models, the mass invariant model best predicted the posttraining data of Experiment 2. As discussed above, Kreegipuu and Runeson (1999) found tacit evidence for a mixture of both inferential and perceptual modes of processing. The dual-process mass invariant model far outperformed either of its submodels on the posttraining data of Experiment 2, suggesting that even observers who have learned to rely on an invariant may revert to simpler perceptual dimensions when the invariant is difficult to detect. Finally, the results of the Experiment 1 suggest that in addition to heuristics and invariants, it is important to explore the possibility that observers might also rely on an exemplar process, especially for transfer collisions that are similar to a familiar set of training collisions. It is interesting to note that, because the exemplar model relies solely on imperfect motion cues, it is a form of a heuristic model, albeit one with a very different flavor from the Gilden and Proffitt (1989) model.

References

Allen, S. W., & Brooks, L. R. (1991). Specializing the operation of an explicit rule. *Journal of Experimental Psychology: General, 120,* 3–19.

Bingham, G. P. (1987). Kinematic form and scaling: Further investigations on the visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 155–177.

Bingham, G. P., Schmidt, R. C., & Rosenblum, L. D. (1995). Dynamics and the orientation of kinematic forms in visual event recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1473–1493.

Busey, T. A., & Tunnicliff, J. L. (1999). Accounts of blending, distinctiveness, and typicality in the false recognition of faces. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25,* 1210–1235.

Cohen, A. L. (2002). *A comparison of invariant and exemplar models of the perception of dynamic properties.* Unpublished doctoral dissertation, Indiana University Bloomington.

Cohen, A. L., & Nosofsky, R. M. (2000). An exemplar-retrieval model of speeded same–different judgments. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 1549–1569.

Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society, 9,* 353–356.

DeLosh, E. L., Busemeyer, J. R., & McDaniel, M. A. (1997). Extrapolation: The sine qua non for abstraction in function learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23,* 968–986.

Erickson, M. A., & Kruschke, J. K. (1998). Rules and exemplars in category learning. *Journal of Experimental Psychology: General, 127,* 107–140.

Estes, W. K. (1994). *Classification and cognition.* New York: Oxford University Press.

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

Gilden, D. L. (1991). On the origins of dynamical awareness. *Psychological Review, 98,* 554–568.

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

Gilden, D. L., & Proffitt, D. R. (1994). Heuristic judgment of mass ratio in two-body collisions. *Perception & Psychophysics, 56,* 708–720.

Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition, 52,* 125–157.

Goldstone, R. L. (1998). Hanging together: A connectionist model of similarity. In J. Grainger & A. M. Jacobs (Eds.), *Localist connectionist approaches to human cognition* (pp. 283–325). Mahwah, NJ: Erlbaum.

Halliday, D., Resnick, R., & Walker, J. (1993). *Fundamentals of physics* (4th ed.). New York: Wiley

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

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *American Journal of Psychology, 57,* 243–259.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. *Psychological Review, 93,* 411–428.

Hintzman, D. L. (1987). Recognition and recall in MINERVA 2: Analysis of the "recognition-failure" paradigm. In P. Morris (Ed.), *Modelling cognition* (pp. 215–229). Chichester, England: Wiley.

Hubbard, T. L. (2004). The perception of causality: Insights from Michotte's launching effect, naive impetus theory, and representational momentum. In A. M. Oliveira, M. P. Teixeira, G. F. Borges, & M. J. Ferro (Eds.), *Fechner Day 2004* (pp. 116–121). Coimbra, Portugal: International Society for Psychophysics.

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.

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgment: A sampling model of confidence in sensory discrimination. *Psychological Review, 104,* 344–366.

Kaiser, M. K., & Proffitt, D. R. (1987). Observers' sensitivity to dynamic anomalies in collisions. *Perception and Psychophysics, 42,* 275–280.

Krantz, D. H., & Tversky, A. (1975). Similarity of rectangles: An analysis of subjective dimensions. *Journal of Mathematical Psychology, 12,* 4–34.

Kreegipuu, K., & Runeson, S. (1999). Becoming competent through the perceiver's eye. In M. A. Grealy & J. A. Thomson (Ed.), *Studies in perception and action V: Tenth international conference on perception and action* (pp. 20–23). Mahwah, NJ: Erlbaum.

Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review, 99,* 22–44.

Lachs, L. (2002). *Vocal tract kinematics and cross-modal information.* Unpublished doctoral dissertation, Indiana University Bloomington.

Lamberts, K. (1998). The time course of categorization. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24,* 695–711.

Lockhead, G. R. (1972). Processing dimensional stimuli: A note. *Psychological Review, 79,* 410–419.

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95,* 492–527.

Mckee, S. P., & Watamaniuk, N. J. (1994). The psychophysics of motion perception. In A. T. Smith (Ed.), *Visual detection of motion* (pp. 84–114). San Diego, CA: Academic Press.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100,* 254–278.

Medin, D. L., & Schaffer, M. M. (1978). Context theory of classification learning. *Psychological Review, 85,* 207–238.

Michaels, C. F., & de Vries, M. (1998). Higher order and lower order variables in the visual perception of relative pulling force. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 526–546.

Michotte, A. (1946). La causalidad fisica, Es un dato fenomenico? [Physical causality: Is it a phenomenal datum?]. *Revista de Psicologia General y Aplicada, 1,* 11–58.

Natsoulas, T. (1960). Judgments of velocity and weight in a causal situation. *American Journal of Psychology, 73,* 404–410.

Nosofsky, R. M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 104–114.

Nosofsky, R. M. (1986). Attention, similarity, and the identification–categorization relationship. *Journal of Experimental Psychology: General, 115,* 39–57.

Nosofsky, R. M. (1987). Attention and learning processes in the identification and categorization of integral stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 87–108.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 700–708.

Nosofsky, R. M., Palmeri, T. J., & McKinley, S. C. (1994). Rule-plus-exception model of classification learning. *Psychological Review, 101,* 53–79.

Palmeri, T. J. (1997). Exemplar similarity and the development of automaticity. *Journal of Experimental Psychology, 23,* 324–354.

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

Rice, J. A. (1995). *Mathematical statistics and data analysis* (2nd ed.). Belmont, CA: Duxbury Press.

Ross, B. H. (1987). This is like that: Of similarity effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 629–639.

Runeson, S. (1974). Constant velocity: Not perceived as such. *Psychological Research, 37,* 3–23.

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

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

Runeson, S. (1995). Support for the cue-heuristic model is based on suboptimal observer performance: Response to Gilden & Proffit (1994). *Perception & Psychophysics, 57,* 1262–1273.

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

Runeson, S., & Vedeler, D. (1993). The indispensability of precollision kinematics in the visual perception of relative mass. *Perception and Psychophysics, 53,* 617–632.

Saltzman, E., & Kelso, J. A. S. (1987). Skilled actions: A task-dynamic approach. *Psychological Review, 94,* 84–106.

Sattath, S., & Tversky, A. (1977). Additive similarity trees. *Psychometrika, 42,* 319–345.
Shepard, R. N. (1987). Toward a universal law of generalization for psychological science. *Science, 237,* 1317–1323.
Shepard, R. N., Romney, A. K., & Nerlove, S. B. (1972). *Multidimensional scaling: Theory and applications in the behavioral sciences: I. Theory.* New York: Seminar Press.
Shin, H. J., & Nosofsky, R. M. (1992). Similarity-scaling studies of dot-pattern classification and recognition. *Journal of Experimental Psychology: General, 121,* 278–304.
Storms, G., De Boeck, P., & Ruts, W. (2001). Categorization of novel stimuli in well-known natural concepts: A case study. *Psychonomic Bulletin & Review, 8,* 377–384.
Todd, J. T., & Warren, W. H. (1982). Visual perception of relative mass in dynamic events. *Perception, 11,* 325–335.
Turvey, M. T. (1990). Coordination. *American Psychologist, 45,* 938–953.

Appendix

Models for Quantitatively Judging Relative Mass

Mass Invariant Models

Strong Mass Invariant Model (I_S)

The strong mass invariant model assumes that, after training, the mass ratio of each collision is detected directly with normal noise. On viewing Collision *i*, the probability of responding that the ln mass ratio is $\mathrm{LMR_J}$ is given by

$$P(\mathrm{LMR} = \mathrm{LMR_J} \mid i) = \frac{N(\mathrm{LMR_J},\mathrm{LMR_J},\sigma^2)}{\sum_K N(\mathrm{LMR_K},\mathrm{LMR_K},\sigma^2)}, \tag{A1}$$

where *N* is the probability density of the normal distribution with mean $\mathrm{LMR_K}$ and variance σ^2. The overall mean response is then

$$\sum_K P(\mathrm{LMR} = \mathrm{LMR_K} \mid i) \times \mathrm{LMR_K}. \tag{A2}$$

This model has one free parameter, the noise variance, σ^2. When the variance is zero, this model predicts the actual ln mass ratio. For comparison, the predictions of this “true” model, T, are also presented,

Strong Ratio Mass Invariant Model (I_{SR})

The strong ratio mass invariant model assumes that viewed mass ratio lies on the ratio scale −2.50, −1.65, −1.01, −0.51, −0.15, 0.15, 0.51, 1.01, 1.65, and 2.50. Otherwise, the model is unchanged from the strong mass invariant model

Weak Mass Invariant Model (I_W)

The weak mass invariant model prediction for mass ratio is simply the average of Equation 7 across a number of simulations. As in Experiment 1, Equation 7 was simulated 10,000 times for each collision. The output of the weak mass invariant model is simply the average mass ratio across the 10,000 simulations. If the mean is less than −1.25 or greater than 1.25, the response was assumed to be –1.25 and 1.25, respectively.

Angle-Change Invariant Model (I_A)

As formulated earlier, the angle-change invariant model can only predict which ball is heavier; it cannot accurately predict mass ratio judgments. When asked to make a mass ratio judgment, however, it is reasonable to assume that observers who have learned to use α might base their decision on the absolute size of α. That is, absolute relative mass increases with $|\alpha|$. Indeed, although not perfect, such a strategy would fair reasonably well given the set of collisions used in Experiment 2. The correlation between physical angle and ln mass ratio was −.79 for the training collisions and −.71 for all of the collisions. Recall that participants were asked to judge the absolute size of α on a 10-point scale independently of mass ratio. These judgments were used as a direct prediction of mass ratio. There are no parameters in this model.

Restricted Exemplar Model (E_R)

Each of the 10 responses was considered a category to extend the exemplar model to this task. On viewing Collision *i*, the probability of responding that the ln mass ratio is $\mathrm{LMR_J}$ is given by

$$P(\mathrm{LMR} = \mathrm{LMR_J} \mid i) = \frac{\sum_{j \in J} s_{ij}}{\sum_K \sum_{k \in K} s_{ik}}, \tag{A3}$$

where s_{ij} is similarity as in Equation 5 and *K* is the set of training collisions with a ln mass ratio of $\mathrm{LMR_K}$. The overall mean response, R_e, is then given by Equation A2 where the probability is determined by Equation A3.

Unlike the invariant models, which assume that the model response relates directly to mass ratio, the output of this model is only assumed to be monotonic with mass ratio. Thus, R_e was scaled to correct for any absolute deviation from mass ratio. Specifically the scaled output of the exemplar model is given by

$$5.5 + \gamma \times (R_e - 5.5), \tag{A4}$$

where 5.5 is the center of the response scale and γ is a free parameter that acts to linearly scale R_e.

To restrict the number of parameters, the attention weights are assumed fixed. The two parameters in this model are the sensitivity parameter of Equation 5 and the scaling parameter from Equation A4. The large number of collisions makes it prohibitive to collect a full set of pairwise similarity ratings of the collisions. The collisions were assumed to reside in a space defined by the four dimensions selected for Experiment 1.

Incorporating Preexperimental Strategies

Invariant Models

Let R_m represent a model response for a particular collision and model, and let R_p represent the associated pretraining participant response. One possibility for incorporating preexperiment strategies, already mentioned for the invariant approach, is that observers may fall back on R_p when confidence in the invariant is low. More formally, on viewing Collision *i*, the probability of judging Ball A as heavier is given by

$$P(\text{Ball A Heavier} \mid i) = \frac{R_m}{R_m + \lambda} R_m + \frac{\lambda}{R_m + \lambda} R_p, \tag{A5}$$

(*Appendix continues*)

where λ is a free parameter that acts to scale the contributions of R_m and R_p relative to the value of the perceived invariant. The strong mass invariant model (I_{SP}), the weak mass invariant model (I_{WP}), and the angle-change invariant model (I_{AP}) augmented with pretraining strategies as in Equation A5 are fit to the data from Experiment 2. All three of these models have λ as a free parameter. In addition, the strong and weak mass invariant models also have a variance parameter.

Exemplar Model

The incorporation of pretraining strategies into the exemplar model was similar in spirit to the augmented invariant models. The probability of utilizing an exemplar mode of processing increases as the summed similarity of a transfer collision to the training collisions increases. Specifically, given Collision *i*, the probability of judging Ball A heavier is

$$P(\text{Ball A Heavier} \mid i) = \frac{\sum_{a\in A} s_{ia} + \sum_{b\in B} s_{ib}}{\sum_{a\in A} s_{ia} + \sum_{b\in B} s_{ib} + \lambda} R_e + \frac{\lambda}{\sum_{a\in A} s_{ia} + \sum_{b\in B} s_{ib} + \lambda} R_p, \quad \text{(A6)}$$

where λ is a free parameter that scales the contributions of R_e and R_p relative to the overall summed similarity of Collision *i* (Cohen & Nosofsky, 2000). To keep the number of parameters low, I fixed the sensitivity parameter at 11.80, the average sensitivity of the 4 participants of Experiment 1.[A1] Furthermore, only the restricted exemplar model with pretraining strategies, E_{RP}, was fit. This model has two free parameters, λ and γ.

Note that each of the mixture models has two submodels. The first is the zero-parameter, pretraining-strategies-only model, P, which is simply the pretraining data. The second is the original model, the strong or weak mass invariant model, the angle-change invariant model, or the restricted exemplar model. To determine the utility of combining these submodels, it will be informative to compare the mixture models to each of their submodels.

Received March 16, 2005
Revision received October 17, 2005
Accepted October 22, 2005 ■

[A1] For some ideas on how this model might extrapolate to mass ratios outside the training range, see Delosh et al. (1997).