Human Learning of Elemental Category Structures: Revising the Classic Result of Shepard, Hovland, and Jenkins (1961)

Kenneth J. Kurtz, Kimery R. Levering, Roger D. Stanton, Joshua Romero, and Steven N. Morris

Online First Publication, July 16, 2012. doi: 10.1037/a0029178
The seminal research of Shepard, Hovland, and Jenkins (1961) shifted the direction of scientific thinking on concept learning when it first appeared. The work has more than stood the test of time—50 years later the most important benchmark for evaluating formal accounts of classification learning is the observed ordering of the ease of learning of the SHJ (for short) elemental category structures (Anderson, 1991; Estes, 1994; Feldman, 2000, 2006; Gluck & Bower, 1988; Goodman, Tenenbaum, Feldman, & Griffiths, 2008; Kruschke, 1992; Kurtz, 2007; Lafond, Lacouture, & Mineau, 2007; Love, Medin, & Gureckis, 2004; Mathy & Brad-metz, 2004; Nosofsky, 1984; Nosofsky, Gluck, Palmeri, McKinley, & Glaubhier, 1994; Nosofsky, Palmeri, & McKinley, 1994; Pitt, Kim, Navarro, & Myung, 2006; Vigo, 2006, 2009). In addition, the SHJ types have served as the foundation for innovative behavioral investigations of unsupervised category learning (Love, 2002), cross-species category learning (Smith, Minda, & Washburn, 2004), cognitive development of category learning (Minda, Desroches, & Church, 2008), working memory capacity and category learning (Lewandowsky, 2011), eye-movements and selective attention in category learning (Rehder & Hoffman, 2005), and the role of stimulus composition in category learning (Love & Markman, 2003; Mathy & Bradmetz, 2011; Nosofsky & Palmeri, 1996). Despite critics (e.g., Murphy, 2003, 2005), the impact of the SHJ ordering on the present-day psychological literature on category learning can hardly be overstated. In what follows, we present a detailed review of existing findings and a series of new experimental results that entail a revised interpretation.

Shepard et al.’s (1961) Observed Ordering of the Six Types

Shepard et al. (1961) tested learners on the elemental types of two-choice classifications that can be constructed with stimuli composed of three binary stimulus dimensions (see Feldman, 2000, 2006, for a generalization of these category structures). The abstract form of the eight possible stimulus items (000, 001, 010,
100, 110, 101, 011, 111) can be visualized as the vertices of a three-dimensional cube. Each of the three dimensions of the underlying form of each category structure is physically realized in terms of two contrasting perceptual properties of the stimuli such as shape (e.g., circle or triangle), size (e.g., large or small), and shading (e.g., white or black).

While there are numerous possible two-way classifications, these reduce to six elemental types assuming interchangeability of the dimensions (see Figure 1). The first is Type I based on a unidimensional rule—for example, all members of one category are squares, while all members of the contrast category are triangles. Type II instantiates an exclusive-or (XOR) logical rule that separates one category based on examples that either have two specific properties or else have neither of these properties from a contrast category made up of examples that possess exactly one of the properties. For example, white squares (00-) and black triangles (11-) might comprise one category, while white triangles (01-) and black squares (10-) comprise the other. The third dimension is irrelevant to category membership. For Types III, IV, and V, all three dimensions are relevant, but there is no definitive description in terms of a perfectly predictive cue (as in Type I) or cue combination (as in Type II). Each of these category structures can be characterized in terms of a unidimensional rule supplemented by a memorized exception to the rule. Type IV also conforms to a linearly-separable family resemblance structure (Rosch & Mervis, 1975) in which two items serve as prototypes for their respective categories. The remaining category structure, Type VI, is usually described as having no statistical or rule-like regularities, so it can only be mastered via some form of paired association or item memorization (though, on systematic ways of acquiring Type VI, see Mathy, 2010; Shepard et al., 1961).

Shepard et al. (1961) tested a small sample of participants who each learned all of the six types. The dependent measure was the number of errors before reaching a strict learning criterion of four consecutive blocks of the training set without errors. The observed ordering of the ease of acquisition (from easiest to most difficult) was as follows: Type I < Type II < Types III, IV, V < Type VI. Based on this result, Shepard et al. concluded that “the most serious shortcoming of the [stimulus] generalization theory is that it does not provide for a process of abstraction (or selective attention)” (p. 29). Several researchers in the field (e.g., Kruschke, 1992, 2005; Nosofsky, 1984; Nosofsky, Gluck, et al., 1994) have made a theoretical commitment to the idea that the SHJ ordering is best understood in terms of selective attention to the dimensions that are relevant. The Type I structure requires attention to only one dimension and is the easiest to learn. Type II requires attention to two dimensions (the two that participate in the XOR relationship) and is the second easiest to learn. The more difficult types require attention to all three dimensions in order to be mastered. The increased difficulty of Type VI can be attributed to its lack of within-category similarity structure.

**Contemporary Replications of Shepard et al. (1961)**

Given the small number of participants and the focus on cumulative errors without tracking time course (in addition to aspects of
outmoded methodology), Nosofsky, Gluck, et al. (1994) conducted a replication and extension of Shepard et al. (1961). The stimulus dimension values were shading (solid interior lines or dotted interior lines), shape (square or triangle), and size (large or small). Rather than training each participant on an extended series of distinct classification schemes with the same materials, Nosofsky, Gluck, et al. tested each learner on two randomly selected category structures. The original SHJ pattern (I < II < III, IV, V < VI) was replicated with a somewhat larger sample size (40 participants for each category structure; 20 if only considering performance on the first of the two learning tasks). Specifically, Nosofsky, Gluck, et al. reported significantly fewer errors in Type II than in Types III, IV, and V based on an “average” $t$ test (p. 356). The traditional ordering held consistently from initial learning through asymptote (see Figure 2 based on data from Nosofsky, Gluck, et al., 1994). A useful implication is that the SHJ paradigm can be suitably employed with a shorter training session (though Nosofsky, Gluck, et al., 1994, made no claims in this regard).

Smith et al. (2004) used the SHJ paradigm for a cross-species investigation of rhesus monkey versus human category learning. Their methodology differed from Nosofsky, Gluck, et al. (1994); Brief time-out periods were enforced after incorrect responses; the dimension values for shading were white versus dark gray; each learner ($N = 47$) was tested on all six types in counterbalanced order; and learners were given a motivational framework based on points and cash prizes. For the human learners, a significant advantage was found for Type II relative to a combination of Types III, IV, and V—though the size of the advantage was distinctly less than in Nosofsky, Gluck, et al. The Type II advantage was once again consistent across the learning session.

Impact of the Traditional SHJ Ordering on the Psychology of Category Learning

The signature characteristic of the traditional SHJ ordering that compels so much scientific attention is the relative ease of learning Type II and, to a lesser extent, the relative difficulty of learning Type IV. As an instantiation of the XOR logical rule, Type II is non-linearly separable. It is also lacking in ecological validity or manifest intuitiveness as a categorization basis (e.g., white squares and black triangles vs. white triangles and black squares). By contrast, Type IV is linearly separable and embodies a rudimentary form of the family resemblance organization prevalent in natural categories (Rosch & Mervis, 1975). At the theoretical level, the traditional Type II advantage over Type IV (supported by findings such as in Medin & Schwanenflugel, 1981) has contributed to increased acceptance of exemplar-based models (along with certain other models discussed below) and a decline of the prototype view among researchers who focus on the phenomenology and modeling of artificial classification learning (see the following reviews: Murphy, 2002; Ross, Taylor, Middleton, & Nokes, 2008).

The finding that Type II is more easily learned than Type IV has served as a critical model-fitting challenge. Nosofsky, Gluck, et al. (1994) showed that attention learning covering map (ALCOVE; Kruschke, 1992)—an adaptive network model implementing the exemplar-based approach (Medin & Schaffer, 1978; Nosofsky, 1986)—produced a better account of the SHJ ordering than its major competitors at the time: the rational model (Anderson, 1991) and configural cue model (Gluck & Bower, 1988). The best fits of the rational model and configural cue model each showed nearly equivalent learning curves for Types II and IV, did not capture the speed of Type I acquisition, and incorrectly predicted Type III to be the second easiest category structure. Prototype-based models (Hampton, 2006; Homa, 1984; Posner & Keele, 1968; Rosch & Mervis, 1975; Smith & Minda, 1998) predict that linearly separable category structures are easier to learn, so even equal performance between Types II and IV (let alone a Type II advantage) rules out prototype models that predict a clear Type IV advantage.

Nosofsky, Gluck, et al. (1994) attributed the success of ALCOVE largely to its use of dimensional selective attention. They showed (also in Kruschke, 1992) that ALCOVE does not predict faster Type II learning if dimensional selective attention is turned off (i.e., free parameter set to zero). Further, Nosofsky and Palmeri (1996) tested the SHJ types using integral dimension stimuli (for which it is difficult to pick out the distinct properties; see Garner, 1974) and found that human learning performance was well fit by ALCOVE without selective attention. A reversal was observed such that Type II was more difficult to learn than Type IV when the three perceptual dimensions of the stimuli were not subject to straightforward extraction or analysis as separable qualities. The theoretical position associated with the exemplar view is that models must be able to account for either fast or slow Type II learning, and the key causal factor is the extent to which selective

![Figure 2](image-url) Proportion correct across the first eight blocks of learning for Type II and Type IV from Nosofsky, Gluck, et al. (1994).
attention is invoked by the experimental conditions. That is, when stimuli are comprised of separable dimensions, then Type II is learned faster than Type IV; when integral dimension stimuli are used, a Type IV advantage emerges.

The traditional SHJ ordering continues to serve as a benchmark for models of category learning—helping to establish the explanatory power of three mechanic models of category learning. The rule-plus-exception (RULEX) model (Nosofsky, Palmeri, & McKinley, 1994) of two-choice classification learning tasks accounts for the SHJ ordering in terms of a process of serial evaluation of logical rules (from simple to complex) plus memorization of exemplars when needed as exceptions to rules. SUSTAIN (Supervised and Unsupervised STRatified Adaptive Incremental Network; Love et al., 2004) is an adaptive network model (like ALCOVE) that is capable of forming clusters (i.e., sub-prototypes or abstractions over sets of similar exemplars) in addition to storing individual exemplars. SUSTAIN captures a traditional Type II advantage because the category structure requires the formation of fewer clusters than modal solutions to SHJ Types III–VI. The ability of ALCOVE and SUSTAIN to predict a Type II advantage has been the cornerstone of a widely-held theoretical view that category learning is best explained in terms of attention-weighted similarity to reference points.

The divergent autoencoder (DIVA) model (Kurtz, 2007, 2012) is an alternative account of category learning based on traditional connectionist learning via back-propagation (Rumelhart, Hinton, & Williams, 1986) with an autoassociative architecture. The psychological theory underlying DIVA is that classification tasks are solved by error-driven learning of the following: (1) a encoding scheme (incoming weights to hidden layer) to re-represent all items in the training set in a PCA-like feature space, and (2) a decoding scheme for each category (channels of weights connecting the hidden layer to banks of output units) that predict values in the original input space based on the encoding. The learning process coordinates the recoding and decoding schemes so that the members of each category are reconstructed along their category channel with minimal distortion. Along other channel(s), the decoding scheme is incompatible since it has been optimized for members of another category—therefore, the predicted feature values will not match the original input. An example is assigned to a particular category if its features are consistent with learned expectations for that category, that is, if the recoding/decoding process leads to feature predictions that match the observed features. Type II is learned relatively easily by DIVA because the model is sensitive to the within-category correlation inherent in the XOR rule and is insensitive to the non-linearly separable nature of the class discrimination. While DIVA correctly captures the traditional SHJ ordering, the predicted Type II advantage is small (Kurtz, 2007). A recent extension of DIVA (Kurtz, 2012) uses a novel form of focusing to account for the data of Nosofsky, Gluck, et al. (1994).

The Uncelebrated Cases of No Type II Advantage

A detailed review of the literature on the SHJ paradigm reveals a surprise: A Type II advantage often does not occur. Lewandowsky (2011) tested acquisition of the SHJ six types in a study of working memory capacity as an individual difference variable in human category learning. The materials were standard geometric stimuli with the shading dimension instantiated as red versus unfilled interior. Each learner experienced a modular sequence of the SHJ types designed to account for possible order effects. The results showed no significant difference or statistical trend between Types II and IV.

To test the claim that the traditional dimensions of variation in category learning materials (shape, size, shading) are systematically non-independent, Love and Markman (2003) compared learning of different subtypes of Type II and a baseline of Type IV. Prior SHJ replications had collapsed across the subtypes that derive from the assignment of physical stimulus dimensions to the logical category structure (i.e., the Type II structure can be instantiated with any two of the dimensions serving as the basis for the XOR rule). Love and Markman (2003) used traditional geometric stimuli with values for the shading dimension of red versus blue interior fill color. Each participant (N = 26 per condition) was assigned to one of the following category structures: shape-irrelevant XOR, shape-relevant XOR, or Type IV. The Type II structure was easier to learn if the critical dimensions were size and texture (i.e., shape irrelevant), as opposed to when shape was relevant. A Type II advantage relative to Type IV was found only for the shape-irrelevant subtype of Type II, not the subtypes for which shape was relevant. Mathy and Bradmetz (2011) also found that subtype differences for Type II learning were prevalent, but that they were non-systematic. It appears that subtle interplay between stimulus materials and dimension assignment may be a factor, or that Type II performance is notably variable when looked at in terms of subgroups of learners.

What do these results signify for our understanding of the SHJ ordering? A close look at the methodological fine points revealed an intriguing dichotomy. Nosofsky, Gluck, et al. (1994) gave learners “explicit instructions that the relevant rule and dimensions for the second problem were chosen independently of those that were relevant in the first problem” (p. 355). The use of the word “rule” invites an approach that promotes acquisition of category structures based on fully predictive logical rules (i.e., Types I and II). In addition, the use of the term “relevant” implies that some dimensions matter and others do not, and it further suggests that the learning task be seen as a matter of determining which dimensions are relevant. This invites the learner to engage in hypothesis testing or selective attention above and beyond any natural inclination. Smith et al. (2004), in the other extant replication that yielded the traditional SHJ ordering, published their word-for-word instructions to human subjects that included the phrases “once you learn the right categories or rules they will work to classify the shapes for the whole task” (p. 402) and “the rules and relevant dimensions will be decided independently for the new task” (p. 403). By contrast, in Lewandowsky (2011) and Love and Markman (2003), there was no evidence of the terms “rules” or “relevant dimensions” being used in task instructions. This is a potentially critical distinction: Only the experiments that we know to have employed such focusing language in the instructions showed a traditional Type II advantage.

We next consider SHJ studies with design elements that extend beyond the standard paradigm. While it could be argued that the SHJ ordering is a phenomenon restricted to the use of the standard geometric materials and the standard classification learning task, this is not consistent with the idea that the ordering represents a major insight into the psychology of category learning or that it is...
a definitive benchmark for evaluating models. Feldman (2000, 2006) used a design in which the SHJ learning problems were framed in terms of distinguishing positive from negative examples based on a fixed-duration, array-based presentation of all labeled examples. The stimuli were amoeba-like drawings that varied in three dimensions out of a possible four: shape of nuclei, size of nuclei, shading of nuclei, or number of nuclei. All participants were tested on each of the six types. There was no Type II advantage (in fact, Type II was not the second easiest to learn). Lafond et al. (2007) conducted extensive further analysis of Feldman's (2000) data and found no reliable differences among Types II, III, IV, and V.

Love (2002) investigated category acquisition under supervised and unsupervised learning modes using a novel set of stimulus materials (described in detail below) that were normed for equal dimension salience and dimension independence. To achieve the desired comparison, a non-traditional form of supervised classification learning was used in which the variable representing category membership in the SHJ task was instantiated as a stimulus feature to be predicted (rather than as a classification label to be chosen). At test, category knowledge was evaluated using forced-choice recognition judgments. Each participant was trained on one of the SHJ Types (I, II, IV, and VI) in either a supervised, intentional unsupervised, or incidental unsupervised mode. Despite finding no significant difference in accuracy between Types II and IV, Love did not interpret the results as contrary to the traditional SHJ ordering because a greater proportion of learners reached criterion in Type II than in Type IV. This pattern of no difference in overall accuracy, but greater ease of reaching essentially perfect performance in Type II, can be understood in terms of Type II learning being more all-or-none (see Smith et al., 2004)—or possibly just being easier to verbalize, retain, and consistently apply—than Type IV.

Rehder and Hoffman (2005) tested learners on SHJ Types I, II, IV, and VI to evaluate the role of selective attention in category learning using eye-tracking methodology. This study differed from traditional classification learning designs in that the stimulus dimensions (keyboard symbols: x, o, ?, =, ¥ presumably of equivalent salience) were displayed in a spatially-separated equilateral triangle arrangement to facilitate the recording of eye movements. Each participant was trained on one of the four selected SHJ types. No significant difference was observed in accuracy between Type II and Type IV. Similarly, Love and Markman (2003) reported experiments using text-based stimuli that showed no reliable differences in accuracy between Types II and IV.

To summarize, there are two published failures to find a traditional Type II advantage using the SHJ paradigm with fully standard materials and procedures (Lewandowsky, 2011; Love & Markman, 2003). We have identified an instructional detail as a possible explanation. In addition, across a range of studies testing the SHJ ordering with non-standard design elements, a Type II advantage has never been observed. It is difficult to make causal attributions from these studies because they tend to modify the stimulus materials and learning conditions at the same time (and we do not have access to the detailed instructions). However, there appears to be good reason to dispense with any notion of a general Type II advantage with separable stimuli. A more nuanced position is that models need to be able to show fast Type II learning because it can occur—therefore, models that are incapable of fast Type II learning fail to capture part of the overall phenomenology. Rather than trying to determine a “true” or universal SHJ ordering, we need to understand why Type II is sometimes advantaged, even if it is generally not.

**Overview of Experiments**

Our studies are designed to find out when and why a Type II advantage occurs. The dependent measure is mean classification accuracy during training (though we also report the proportion of participants reaching criterion). Love and Markman (2003) have suggested that the proportion of subjects reaching criterion may measure mastery more effectively than proportion correct. This is partially motivated by the SHJ types differing in terms of the efficacy of applying a suboptimal, one-dimensional logical rule (a unidimensional rule supports 75% accuracy in Type IV, but only 50% accuracy in Type II). We consider how learners advance toward high levels of accuracy to be the element of primary psychological interest, rather than the point at which they achieve flawless performance. While this issue might be open to debate, it is clearly the case that the traditional SHJ ordering is articulated with respect to accuracy and that models are tested with respect to this measure.

We begin with an attempt to replicate the finding of a Type II advantage using a traditional version of the SHJ paradigm that specifically includes instructions with rule language. The prediction is successful replication of the Type II advantage based on the use of standard geometric materials, a standard classification learning task, and rule language. In the subsequent set of experiments, we compare learning of Type II and Type IV using neutral instructions (by which we mean no focusing language). This is expected to lead to attenuated Type II learning and no advantage over Type IV. To complete this phase of the investigation, we directly manipulate the inclusion of instructional rule language within a single experimental design. The experimental series continues with studies using non-standard stimuli to assess specific conditions and mechanisms that determine the rate of acquisition of Type II. These studies advance an argument that it is the extent to which experimental conditions promote the discovery of verbalizable rules or regularities that determines whether Type II learning is likely to be faster, slower, or even with Type IV.

**Experiment 1**

In studies of category learning, the task instructions given to participants tend to be fairly “by the book” and receive little scrutiny. In the published reports that comprise the positive evidence for a Type II advantage (Nosofsky, Gluck et al., 1994; Shepard et al., 1961; Smith et al., 2004), the task is presented to participants explicitly in terms of the acquisition of rules and relevant dimensions (note that logical rules can be seen as equivalent to a restricted similarity computation based on matching values on relevant dimensions). Such language in the task instructions could clearly facilitate Type II learning since the development and application of a logical rule (or a mechanism that effectively captures such a rule) over the two relevant dimensions is critical. By contrast, the intermediate SHJ types including Type IV do not have a single cue or cue combination that is a perfect predictor of category membership (though these structures are
amenable to a rule-plus-exception solution). In this experiment, we compare Type II and Type IV learning under traditional experimental conditions with the specific inclusion of rule language.

Method

Participants. A total of 169 students from Binghamton University participated in this experiment for partial fulfillment of course credit. Of the participants, 119 were randomly assigned to learn the Type II problem, and 50 were randomly assigned to learn the Type IV problem. The larger sample for Type II is because we elected to test all three subtypes of Type II and one arbitrarily selected subtype of Type IV (for which no systematic differences have been attributed to dimension assignment).

Materials. Training items were standard geometric stimuli for the SHJ paradigm (see Table 1). The shape dimension was square or triangle (as in Nosofsky, Gluck, et al., 1994), shading was black or white, and size values were 28 mm² or 56 mm². Stimuli were displayed on a computer screen situated in the center of a light gray display rectangle on a blue background. We avoided order or practice effects by randomly assigning each participant to learn only Type II or Type IV.

Procedure. For each trial, a stimulus item was displayed accompanied by two buttons labeled “Alpha” and “Beta.” Participants were asked to select the appropriate category via mouse click. After providing a response, the learner received evaluative feedback with the correct category label displayed next to the example (continuing to the next trial was self-paced). Presentation order was based on a random ordering of each pass through the eight training items. The maximum length of the training session was set at eight presentations of each item for a total of 64 learning trials. The successful SHJ replications (Nosofsky, Gluck, et al., 1994; Smith et al., 2004) both showed a clear Type II advantage over this span. Participation was complete when the learner reached a criterion of two consecutive perfect passes through the eight items or completed all 64 trials. The instructions to each participant before the learning task stated the following:

In this experiment, you will be shown examples of geometric images. Your job is to learn a rule that allows you to tell whether each example belongs in the Alpha or Beta category. As you are shown each example, you will be asked to make a category judgment and then you will receive feedback. At first you will have to guess, but you will gain experience as you go along. Try your best to gain mastery of the Alpha and Beta categories.

Results and Discussion

For this experiment (and all subsequent experiments in which a criterion is used), our primary dependent measure of classification accuracy was evaluated under the assumption that learners who reached criterion would have continued to produce errorless responding (as in Nosofsky, Gluck, et al., 1994). A 2 (problem type) × 8 (learning block) analysis of variance (ANOVA) revealed that the mean proportion correct was greater for Type II learners (M = .794, SD = .161) than Type IV learners (M = .734, SD = .112), F(1, 167) = 5.678, p = .018, η² = .033. There was no reliable interaction. Learning curves for all experiments are provided in Figure 3, and histograms of individual performance are shown in Figure 4. As in previous reports, we found that a greater proportion of Type II learners (.630) successfully reached criterion compared to Type IV learners (.320), χ²(1) = 93.074, p < .001.

A one-way ANOVA comparing the Type II subtypes revealed no significant differences between shape-irrelevant (M = .777, SD = .154), size-irrelevant (M = .835, SD = .164), and color-irrelevant (M = .771, SD = .161), F(2, 116) = 1.911, p = .153, η² = .032. However, an ANOVA comparing the three subtypes of Type II versus Type IV was significant, F(3, 165) = 3.456, p = .018, η² = .059, and a Dunnett’s post hoc test showed that only performance on the size-irrelevant subtype was significantly better than Type IV (p = .005; all other ps > .3). These results are broadly consistent with prior evidence of a Type II advantage being subtype-dependent, but they do not show the advantage for shape-irrelevant XOR observed by Love and Markman (2003).

These results replicate the finding of an overall Type II advantage with the use of rule instructions, although the effect is driven by just one of the Type II subtypes (size-irrelevant). It is possible that the two existing SHJ replications (Nosofsky, Gluck, et al., 1994; Smith et al., 2004) were also driven by a single subtype, but this does not seem highly likely, at least in the former case, given the magnitude of the difference. The results of this experiment effectively demonstrate and reinforce prior findings of a Type II advantage with rule language in the instructions.

Experiment 2

The next question is as follows: What happens without instructions that use rule language? As reviewed above, published tests of the SHJ ordering that do not provide details about the task instructions do not show a Type II advantage. A unifying interpretation of the existing data is that rule instructions promote fast Type II learning and contribute to a Type II advantage. Our key prediction is that a Type II advantage will not occur with neutral instructions. On account of predicting a null effect, that is, no difference between Types II and IV, we sought to maximize experimental power using a simple one-factor design in order to find an effect if one exists.

We calculated the number of subjects needed to obtain a 95% chance of finding an effect of the magnitude seen in Nosofsky, Gluck, et al. (1994). We targeted this amount of power because (with an alpha level of .05) Type I and Type II errors are equal, that is, it is just as unlikely to erroneously find the effect when it is not truly present as it is to erroneously fail to find the effect that is truly present. In Nosofsky, Gluck, et al., the reported difference between the mean proportion correct for Type II (0.832) and Type IV (0.722) after 64 trials (matched to our experiment) is 0.120. Because standard deviations were not published, we estimated using standard deviations from our SHJ studies (0.163 for Type II and 0.110 for Type IV). Accordingly, the effect size in Nosofsky, Gluck, et al. is d = 0.791, that is, a large effect. To achieve 0.95 power, 43 participants are needed in each condition of our experiment. For a moderate

---

1 To validate this assumption, in a similar experiment conducted without a learning criterion, we found that 89% of learners produced completely errorless responding after meeting the requirements of a slightly less demanding criterion than the one used in the present experiment; further, the occasional errors did not occur disproportionately for Type II or Type IV.
effect \((d = 0.5)\), 105 participants are needed in each group to have a 95\% chance of finding an effect if it exists.

Method

Participants. A total of 322 students from Binghamton University participated in this experiment for partial fulfillment of course credit. In considerable excess of the numbers needed according to the power calculation above, 133 participants were randomly assigned to the Type II category structure, and 189 participants were randomly assigned to the Type IV category structure. The larger sample for Type IV is due to testing all four possible subtypes.

Materials. The stimuli and category structures were the same as in Experiment 1 except that all four subtypes of Type IV were used.

Results and Discussion

A 2 (problem type) \(\times\) 8 (learning blocks) ANOVA revealed no main effect of problem type, \(F(1, 320) = 2.544, p = .112, \eta^2 = .008\), indicating that the mean proportion correct did not differ across conditions.

Table 1

The Two Possible Values for Each Perceptual Stimulus Dimension in Each Experiment

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Feature 1</th>
<th>Feature 2</th>
<th>Feature 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1, 2, 4</td>
<td>Size: 28 mm(^2)</td>
<td>Square</td>
<td>White fill</td>
</tr>
<tr>
<td>3, 5</td>
<td>Size: 38 mm(^2)</td>
<td>Circle</td>
<td>White fill</td>
</tr>
<tr>
<td>6, 7</td>
<td>White border</td>
<td>Diagonal line</td>
<td>Interior dots</td>
</tr>
<tr>
<td>8</td>
<td>Size: 28 mm(^2)</td>
<td>Diamond</td>
<td>Speckled fill</td>
</tr>
</tbody>
</table>

Figure 3. Proportion correct for Type II and Type IV across learning for all experiments. Error bars indicate the standard error of the mean.

Procedure. The procedure was the same as in Experiment 1 except the instructions did not include rule language:

In this experiment, you will be shown examples of geometric images. Your job is to learn to tell whether each example belongs to the Alpha or Beta category. As you are shown each example, you will be asked to make a category judgment and then you will receive feedback. At first you will have to guess, but you will gain experience as you go along. Try your best to gain mastery of the Alpha and Beta categories.
significantly between Type II ($M = .722, SD = .199$) and Type IV ($M = .694, SD = .124$). This failure to reject the null hypothesis of no difference between Types II and IV is based on a much larger sample than ever before used in the literature on the SHJ paradigm. A significantly larger proportion of participants in the Type II group (.504) reached criterion relative to the Type IV group (.228), $\chi^2(1) = 26.487, p < .001$. It may well be that Type II is more likely to be quickly mastered, yet not generally easier to learn. We believe this can be understood in terms of greater extremes (i.e., non-normality or bimodality) in Type II learning—such that more learners excel, but also more struggle. Although analyses such as ANOVAs assume normal distributions, the tests are robust to violations of normality, especially when the sample sizes are as large as they are in the present research (Harwell, Rubinstein, Hayes, & Olds, 1992). As such, the bimodality of the distributions does not impact the appropriateness of the statistical tests reported here. We explore the potential role of bimodality in the General Discussion.

A one-way ANOVA on mean accuracy across the Type II subtypes revealed a significant difference, $F(2, 131) = 6.470, p = .002$, and Tukey’s post hoc tests showed the difference was driven by better performance in the size-irrelevant subtype. Participants
who learned this subtype showed significantly higher accuracy ($M = .798$, $SD = .189$) than participants learning the color-irrelevant subtype ($M = .654$, $SD = .177$), $p = .001$, and a trend toward better performance than those learning the shape-irrelevant subtype ($M = .714$, $SD = .204$), $p = .09$. An ANOVA comparing all three subtypes of Type II to the average of all Type IV learners was also significant, $F(3, 319) = 7.312$, $p < .001$. Dunnett’s post hoc test showed that only the size-irrelevant subtype resulted in performance that was significantly better ($p < .001$) than Type IV (other $p s > .2$). These results are consistent with the subtype effects in Experiment 1 and suggest that faster acquisition with the size-irrelevant Type II subtype is reliable for these materials.

The key outcome of this experiment is an observed non-difference between Type II and Type IV with considerable experimental power. The results show a possible hint of better Type II performance later in learning (see Figure 3), so it would be reassuring to assess learning over a longer duration. Another possible concern is that there are a number of minor methodological details in the present design that differ from Nosofsky, Gluck, et al. (1994). For these reasons, as well as the need for healthy skepticism about a null effect, the next experiment is a replication using slightly different methodology within the standard paradigm.

**Experiment 3**

The design of this experiment is in close accord with the methodological details of the SHJ replication by Nosofsky, Gluck, et al. (1994)—which has been considered the definitive demonstration of the SHJ ordering and the basis for fitting formal models. As in Nosofsky, Gluck, et al., participants were tested on two problem types. Since we were not evaluating all six SHJ types, every participant was tested on Types II and IV. This raises potential concerns about contamination or ordering effects, but it offers the advantage of eliminating sample variability by testing the same set of learners on both category structures. The stimuli were standard geometric materials following Shepard et al.’s (1961) original work. This study employed the same randomization scheme for item presentation order (described below) as in Nosofsky, Gluck et al. The key methodological deviation from Nosofsky, Gluck, et al. was in the instructions, where we used the term “strategies” instead of “rules” and did not mention “relevant” dimensions.

**Method**

**Participants.** A total of 111 undergraduate students from St. Mary’s College of Maryland completed this experiment for partial fulfillment of a course requirement.

**Materials.** The dimensions of variation matched Nosofsky, Gluck, et al. (1994), though we followed Shepard et al.’s (1961) original study for the dimension values: shape (circle or triangle), size (38 mm² or 76 mm²), and shading (white or black). Stimuli were presented on a computer screen against a gray background. Each participant learned both the Type II and Type IV problems with order randomized. The assignment of subtype was determined randomly for each participant.

**Procedure.** Before the training task, the dimensions of variation among the stimuli were illustrated to participants, and they practiced the response procedure. Participants were instructed that they would be tested on the first of two categorization problems in which they were to learn to categorize the objects into two different categories. Before the second problem, participants were given explicit instructions that, although the physical dimensions were the same, the structure of the problem was different and any strategy that was suitable for the first problem might not be suitable for the second. The wording of these instructions differed critically from Nosofsky, Gluck, et al. (1994), in that our language did not include any reference to rules or relevance.

In the first block of trials, the eight stimuli were presented in a random sequence of eight trials and then were presented again in a new random sequence of eight trials. For all remaining blocks, each of the eight stimuli appeared twice within a random sequence of 16 trials. For each learning trial, the stimulus appeared on the screen, the participant responded by pressing either “A” or “B” on the keyboard, and corrective feedback was displayed below the stimulus for 1 s followed by a 0.5-s inter-trial interval. This process was iterated until the participant correctly classified all of the items in one block of 16 trials or completed the eighth and final block. This was less strict than the two perfect blocks required by Nosofsky, Gluck, et al. (1994). The stopping point for learners who did not reach criterion was set at eight blocks (96 trials)—longer than Experiment 1, but shorter than Nosofsky, Gluck, et al.²

**Results and Discussion**

In direct contrast with the results of Nosofsky, Gluck, et al. (1994), a 2 (problem type) × 8 (learning blocks) ANOVA revealed no evidence of a difference in accuracy between Type II ($M = .743$, $SD = .165$) and Type IV ($M = .743$, $SD = .122$), $F(1, 110) = 0.002$, $p = .966$, $\eta^2 < .001$. We did find a significant interaction between problem type and learning block, $F(7, 770) = 4.215$, $p < .001$, $\eta^2 = .037$. This interaction appears to be driven by an initial advantage (Blocks 1–2) for Type IV combined with a slight Type II advantage during the second half of learning (which dissipates by Block 8). To address this issue, we performed a secondary 2 (problem type) × 6 (learning blocks) ANOVA restricted to Blocks 3–8. Again, the main effect of problem type was not significant, $F(1, 110) = 0.571$, $p = .452$, $\eta^2 = .005$, and the main effect of learning blocks was significant, $F(5, 550) = 51.256$, $p < .001$, $\eta^2 = .318$. Notably, there is not a significant interaction between problem type and learning block when the analysis is restricted to the final six learning blocks, $F(5, 550) = 1.791$, $p = .113$, $\eta^2 = .016$. This confirms that the observed interaction is driven by an advantage for Type IV in the first two learning blocks.

Participants learning the Type II structure were significantly more likely to reach criterion, $\chi^2(1) = 7.152$, $p = .007$ (61% of Type II learners reached criterion compared to 43% of Type IV learners). Since this experiment is based on a shorter maximum learning period than Nosofsky, Gluck, et al. (1994), we looked specifically at their results over the same duration (keeping in mind their use of a stricter learning criterion). Figure 2 shows Nosofsky, Gluck, et al.’s data in the same presentation format used

² We note that for all of the SHJ types (other than Type VI), classification performance was quite good (above 90% correct) by the end of the eighth block in Nosofsky, Gluck, et al. (1994).
for our own results. As can be seen, they found a rapidly emerging Type II advantage that dissipated over time (perhaps due to a ceiling effect), while we found no Type II advantage during the same window.

Since each learner was trained on both problem types, there exists the possibility of order effects. To assess the role that order effects may have played in mean performance on the two problem types, a 2 (problem type) × 2 (testing order) × 8 (learning blocks) ANOVA was performed on the accuracy data. There was no main effect of problem type or testing order. The interaction between problem type and testing order fell narrowly short of significance, \(F(1, 109) = 3.891, p = .051, \eta^2 = .034\). To help interpret this marginal interaction, we performed separate 2 (testing order) × 8 (learning blocks) ANOVAs for the Type II and Type IV problems. Mean accuracy was significantly higher when Type II was tested second as opposed to first, \(F(1, 109) = 3.988, p = .048, \eta^2 = .035\). In addition, there was a significant interaction of testing order and learning block, \(F(7, 763) = 2.381, p = .021, \eta^2 = .021\), which suggests that the advantage for the second problem was greatest early in learning. There was not a significant difference in accuracy for Type IV based on testing order, and there was also not an interaction between testing order and learning block. According to these results, experiencing the Type II problem after Type IV may make Type II easier to learn, while learning Type IV after Type II shows no benefit. Therefore, experimental designs that train learners on multiple problems may favor Type II.

Using a very close match to the experimental procedure of Nosofsky, Gluck, et al. (1994), we found a compelling non-difference between Types II and IV without rule instructions. In order to further test our claims and to address an open question having to do with the randomization scheme used in SHJ studies, we conducted another test of Type II versus Type IV using neutral instructions (no rule language).

**Experiment 4**

Shepard et al.’s (1961) original study and Nosofsky, Gluck, et al.’s (1994) replication (and our Experiment 3) used a distinctive randomization scheme for the ordering of item presentation during learning. The first and second sets of eight trials are based on a random ordering of the training set. After these first 16 trials, the ordering is based on randomized sets of 16 trials. There is some (though not a perfect) tendency in the prior literature for a Type II advantage to be associated with this by-8-8-16’s randomization scheme. In the by-8-8-16’s scheme, the order within each block of 16 is based on randomly selecting items from a set of 16 (two of each), while the order for blocks of eight is based on randomly selecting from a set of eight (one of each). When there are multiples of each item within the set drawn from, this tends to lead to a more “lumpy” ordering in terms of the spacing of the presentations of each item. In an ordinary by-8’s randomization scheme, an item repetition (i.e., when the very same item is presented more than once in immediate succession, thereby rendering the second of the two trials trivially easy if the learner notices the repetition and recalls the correct answer) can only occur if a block happens to start with the same item with which the last block ended (a one in eight chance). Three-in-a-row repetitions never occur, and there cannot be extended periods during which a particular item is not experienced. By contrast, a by-8-8-16’s randomization scheme tends to produce more item repetitions, more near repetitions (with an intervening item), potential three-in-a-row repetitions, and substantial gaps between exposures (maximum of 28 trials). The acquisition of particular category structures may be impacted by the likelihood of such distributional anomalies.

Our interest in presentation order was furthered by a small-scale pilot study conducted by one of the authors (Kenneth J. Kurtz) in another line of research. In that investigation, Types II and IV were tested using a full-order randomization scheme (by-64) in which all 64 trials were randomly ordered without any constraints to promote regular distribution of the appearances of each item. A Type II advantage similar in magnitude to Nosofsky, Gluck, et al. (1994) was observed (without rule language). Therefore, while the causal mechanism is not clear, increased distributional anomalies in presentation order could favor Type II learning. In order to fully and directly test this possibility, we used a large sample size and manipulated the randomization of item presentation including all three randomization schemes of interest: by-8’s, by-8-8-16’s, and by-64.

This experiment provides a further test of the reliability of the failure to observe a significant difference between Types II and IV without rule language (Experiments 2 and 3). We again predict no difference in accuracy between the two problem types, though the pilot study raises the possibility of faster Type II learning in association with increased distributional anomalies in presentation order.

**Method**

**Participants.** A total of 227 students at Binghamton University participated in this experiment for partial fulfillment of course credit. By random assignment, 114 participants were trained on subtypes of the Type II category structure, and 151 participants were trained on the subtypes of the Type IV structure.

**Materials.** The stimuli and category structures were the same as in Experiment 2.

**Procedure.** The instructions given to participants did not include rule language, and the learning session was conducted as in Experiment 2. However, the randomization scheme for presentation order was manipulated as a between-subjects variable. Stimulus presentation order was randomized either by-8’s (as in Experiments 1 and 2), by-8-8-16’s (as in Nosofsky, Gluck, et al., 1994; Shepard et al., 1961; and Experiment 3), or by-64 (as in the pilot). In addition, there was no performance-based criterion for ending the learning session (the by-64 randomization scheme offers no meaningful breakdown of the learning phase into passes through the training set).

**Results and Discussion**

A 2 (problem type) × 3 (randomization scheme) ANOVA was conducted to determine whether the learning of Types II and IV was influenced by randomization scheme (see Table 2 for means and standard deviations across randomization schemes and problem types). No significant main effects were found due to problem type or randomization scheme, and there was not a significant interaction. Collapsing across randomization schemes, there was no significant difference between the learning of Type II (\(M = .735, SD = .170\)) and Type IV (\(M = .722, SD = .106\), \(p > .4\). A
comparison of Type II subtypes revealed a significant difference, \( F(2, 111) = 17.211, p < .001, \eta^2 = .237. \) Post hoc tests showed again that performance on the size-irrelevant subtype \( (M = .845, SD = .163) \) was significantly better than performance on the shape-irrelevant \( (M = .665, SD = .137) \) and color-irrelevant \( (M = .682, SD = .149) \) subtypes \( (ps < .001) \). Based on these results, there is no case to be made for randomization scheme as an explanatory factor in Type II versus Type IV learning. This raises the question of what to make of the Type II advantage in the pilot study. Given that randomization schemes are (of course) a matter of random variation, it is possible for a particular scheme to produce less or more distributional anomalies in a particular sample. We computed the mean number of item repetitions in each data set and found no extreme variation from expected values. The best explanation we can put forth is the difference in sample size. It is naturally the case that smaller-scale studies are more likely to reflect biased samples of an underlying distribution. In fact, it is not out of the question to view the prior evidence for a Type II advantage (Nosofsky, Gluck, et al., 1994; Shepard et al., 1961; Smith et al., 2004) in this light, that is, as outliers attributable to small sample size. We take up the sampling issue more fully in the General Discussion.

This experiment provides further evidence that a Type II advantage does not occur without using rule language and speaks clearly against randomization scheme as an explanatory factor in the SHJ ordering. Even so, the observation that randomization scheme influences the learning setting in terms of the likelihood of distributional anomalies is a methodological point that deserves consideration in the design of category learning experiments and as a topic for further research.

### General Discussion

While we have demonstrated repeated failures to observe a Type II advantage without rule instructions, we now take a different experimental tack in order to directly test the claim that Type II learning is advantaged with rule instructions, but not without. We predict a significant interaction within a single experimental design that would pinpoint the critical role of instructions in determining the relative ease of learning.

### Method

**Participants.** A total of 90 undergraduate students from St. Mary’s College of Maryland completed this experiment for partial fulfillment of a course requirement. Participants were randomly assigned to either the rule condition \( (N = 44) \) or the no-rule condition \( (N = 46) \) for learning both Types II and IV.

**Materials.** The stimuli were the same as those used in Experiment 3.

**Procedure.** The procedure followed that of Experiment 3 without initial practice with the responding procedure. The critical change is the inclusion of the rule language condition in which participants were instructed to “learn the relevant rule and dimensions that allow you to categorize the objects into two different categories.” This language was presented before the first problem and again before the second problem.

### Results and Discussion

To examine the course of learning and any interaction with the sequence of learning the problem types, a 2 (problem type) \( \times 2 \) (rule condition) \( \times 2 \) (sequence) \( \times 8 \) (learning blocks) ANOVA was performed on the block-by-block accuracy data. In terms of main effects, the mean accuracy for Type II \( (M = .756, SD = .152) \) was significantly higher than Type IV \( (M = .719, SD = .207) \), \( F(1, 86) = 5.748, p = .019, \eta^2 = .063. \) The mean accuracy for the rule condition \( (M = .721, SD = .457) \) was not significantly different than the no-rule condition \( (M = .754, SD = .203) \), \( F(1, 86) = 1.934, p = .168, \eta^2 = .022; \) however, the main effect of block was significant, \( F(7, 602) = 131.289, p < .001, \eta^2 = .604. \) Importantly, there was a significant interaction, as predicted, between rule condition and problem type, \( F(1, 86) = 4.249, p = .042, \eta^2 = .047. \) The interaction between problem type and block was significant, \( F(7, 602) = 4.666, p < .001, \eta^2 = .051, \) as well as the interaction between condition and block, \( F(7, 602) = 2.930, p = .005, \eta^2 = .033. \) All other main effects and interactions were not significant \( (ps > .15). \)

Further analyses showed that significantly more participants reached criterion for Type II \( (567) \) than for Type IV \( (311), \) \( \chi^2(1) = 17.612, p < .001. \) Comparing Type II learners across conditions, significantly more participants learned Type II to criterion in the rule condition \( (682) \) than in the no-rule condition \( (457), \) \( \chi^2(1) = 4.649, p = .031. \) For those that learned Type II to criterion, the proportion correct was not significantly different for the rule condition \( (M = .859, SD = .079) \) and the no-rule condition \( (M = .863, SD = .103) \), \( p = .880. \) The proportion of participants that learned Type IV to criterion was not significantly different for the rule condition \( (.227) \) and the no-rule condition \( (.391), \) \( \chi^2(1) = 2.823, p = .093. \) For Type IV learners, the proportion correct for the rule condition \( (M = .8703, SD = .052) \) was not significantly different from the proportion correct for the no-rule condition \( (M = .837, SD = .061) \), \( p = .155. \)

While our previous studies offer a series of compelling null results and cross-experiment contrasts, the present results provide a definitive demonstration of advantaged Type II learning only with rule instructions. Given the consistency between the SHJ orderings reported in the reviewed literature, those in Experiments 1–4, and the present results, the evidence firmly indicates that the Type II advantage is conditional on specialized instructions. The impact of instructions is not without precedent in the categorization literature (e.g., Medin & Smith, 1981; see also Brooks, 1978; Smith & Kemler-Nelson, 1984). Such findings are certainly grounds for careful consideration in the preparation of experiments. In the present case, we are led to revise a foundational finding in the field.
Experiment 6

Up to this point, we have collected evidence with standard geometric stimuli for the claim that the Type II advantage depends on the use of rule language in the instructions. However, there remains a strong suggestion that stimulus-based dependencies are mediating our evaluation of the underlying category structures. Love (2002) tested the SHJ ordering with a set of materials that were normed for equal salience and independence, but the materials were used in conjunction with an unusual form of supervised learning in which the class labels were instantiated as values to be predicted on a fourth binary dimension. This transforms the nature of the task from classification to inference and undermines the special status of category-level organization. Our next study is designed to replicate and extend Love’s study by looking at the SHJ ordering using the carefully controlled materials in a traditional classification task. As a further extension, we have repeatedly observed a lack of rapid acquisition of a somewhat complex logical rule (XOR)—what about the traditional finding of near-immediate acquisition of a simple rule? In this design, we include the SHJ Type I (unidimensional rule) category structure.

This experiment addresses whether the observed non-difference between Types II and IV generalizes to stimulus materials that are not based on the standard trio of shape, size, and shading but instead are normed for dimensional salience and independence. Our prediction is that there will be no difference between Type II and Type IV. We also predict that the factors mitigating rapid Type II acquisition will not affect Type I learning that should be learned with only a handful of errors. Essentially, a unidimensional rule should remain easy to acquire across variations in the learning task.

Method

Participants. A total of 149 Binghamton University students participated in this experiment for partial fulfillment of course credit. Of the participants, 51 were randomly assigned to Type I, 51 were randomly assigned to Type II, and 47 were randomly assigned to Type IV.

Materials. Stimuli were taken from Love’s (2002) materials made available for download. In order to create a set of eight training items, three dimensions were selected from the available total of five (the remaining two dimensions were held constant). The training items were all small blue squares—so the dimensions of variation in traditional geometric stimuli (shape, size, and shading) were all held constant. The eight items varied on three binary dimensions: the color of a border surrounding the blue square (yellow or white), interior dots (present or absent), and interior diagonal-line (present or absent). The features were all reasonably easy to notice, distinguish, and verbalize. Since the dimensions of the stimuli were tested for independence and were calibrated for equal salience, only one subtype of each category structure was used. For the Type I problem, the relevant dimension was the presence (or absence) of interior dots. For the Type II problem, one category consisted of examples with interior dots and diagonal lines or neither of these features, while the contrast category had interior dots and no diagonal lines or diagonal lines and no interior dots. The dimension of border color was irrelevant. For the Type IV category structure, the prototype for one category was as follows: yellow border, diagonal line, and no interior dots; and the corresponding prototype consisted of a white border, no diagonal line, and interior dots.

Procedure. The instructions did not involve rule language, and the learning session was conducted as in Experiment 2.

Results and Discussion

As predicted, there was no significant difference between Type II \( (M = .704, SD = .184) \) and Type IV \( (M = .705, SD = .103) \) learners, \( F(1, 96) = 0.001, p = .976 \), and there was no interaction between learning and problem type. This replicates with controlled materials the failure to find a reliable difference between Types II and IV in the experiments above and in Love (2002). As in the previous experiments, the proportion of learners who reached criterion in Type II (353) was significantly greater than the proportion reaching criterion in Type IV (128), \( \chi^2(1) = 6.586, p = .010 \). Type I accuracy \( (M = .944, SD = .065) \) was as high as expected, with an average of 3.6 total errors made during learning (obviously more accurate than Types II and IV). Our major finding is that there is no difference between Types II and Type IV generalizes to stimulus materials that are not reinforced with different materials: When the unpredictable impact of non-independence or differential salience among dimensions is controlled, we observe nearly identical classification accuracy in the learning of Type II and Type IV.

Simulation study. An important goal of the present investigation is to provide a new and improved basis for evaluating and comparing models of category learning. As an example of the type of work we have in mind, we conducted a simulation study to test whether ALCOVE can fit the findings of the present experiment. Modeling these data in particular is a useful contribution because the stimuli are normed for non-independence and equal salience. It has been established that ALCOVE is capable of predicting a Type II advantage with selective attention or relatively slower Type II learning without selective attention. A notable aspect of the published modeling results on the SHJ ordering without selective attention (Kruschke, 1992, Figure 5; see also Nosofsky & Palmeri, 1996, Figure 3) is that Type I learning slows dramatically. In fitting ALCOVE to our data, we expected that the free parameters could be optimized to capture the non-difference between Type II and Type IV, but that ALCOVE would not be able to predict this non-difference along with the near-immediate acquisition of Type I that was seen (as usual) in the human data.

ALCOVE was applied to the current set of human accuracy data using the identical model configuration as in Kruschke (1992). First, ALCOVE was fit to the observed accuracy data for Types I, II, and IV, simultaneously, by searching for the values of the response mapping constant (\( \Phi \)), sensitivity (\( c \)), the attention learning rate (\( \lambda \)), and the learning rate for association weights (\( \lambda_w \)) that minimized the root-mean-squared deviation (RMSD) between observed and predicted values. A detailed description of the psychological meaning of the parameters is available in Kruschke’s study, but we point out that the attention learning rate (\( \lambda_\alpha \)) determines the rate at which observers learn to selectively attend to relevant dimensions. This analysis resulted in an RMSD of 0.1429 across all three categorization types. The best fitting parameters were \( \Phi = 2.4415, c = 2.9097, \lambda_w = 1.227, \) and \( \lambda_\alpha = 1.382 \). Figure 5 displays ALCOVE’s predictions based on these parameter values. ALCOVE accurately predicts the rapid learning we observed for Type I; however, this comes at the expense of incorrectly predicting an early and large advantage for Type II over Type IV. ALCOVE is not able to simultaneously predict
similar performance on Types II and IV along with near-immediate Type I acquisition.

As an additional test of ALCOVE's ability to account for the present results, the model was tested on only the observed data for Types II and IV, and the best fitting parameters were used to generate a corresponding prediction for Type I. The purpose of this was to confirm that ALCOVE can predict a non-difference between Types II and IV and to determine the consequences of the resulting parameter settings on Type I learning. This analysis resulted in an RMSD of .045389. The best fitting parameter values were $\Phi = 0.6771$, $c = 5.6942$, $\lambda w = .4654$, and $\lambda a = .0021$. ALCOVE successfully accounts for the non-difference between Types II and IV; however, this causes the model to drastically underpredict the speed with which learners acquire Type I (see Figure 6). The non-difference prediction is achieved in large part by extreme reduction of the attention learning rate to nearly zero. Without substantial attention learning, ALCOVE is not able to allocate the bulk of the attention weight to the single dimension that is relevant to the Type I categorization problem. Our results closely mirror Kruschke's (1992) demonstration of ALCOVE's predictions when attention learning is turned off. In sum, ALCOVE in its standard form faces a serious shortcoming—the model cannot predict the conjunction of (1) similar time course of learning for Types II and IV and (2) near-immediate acquisition for Type I as observed in our human learning data.

**Experiment 7**

The results of Experiments 1–6, in conjunction with previous studies, suggest that instructions including rule language may be necessary to produce a Type II advantage. However, we also found that the Type II advantage in Experiment 1 was driven by the

![Figure 5](image-url)

*Figure 5.* Observed behavioral data from Experiment 6 and attention learning covering map (ALCOVE) simulation fit to Type I, Type II, and Type IV. ALCOVE predicts slower Type IV learning than was observed.

![Figure 6](image-url)

*Figure 6.* Observed behavioral data from Experiment 6 and attention learning covering map (ALCOVE) simulation fit to Types II and IV only. ALCOVE predicts slower Type I learning than was observed.
size-irrelevant subtype. An open question is whether rule language alone is enough to produce a Type II advantage or whether it requires the use of rule language in conjunction with some property inherent in the combination of the physical properties of our materials and the size-irrelevant subtype. In this experiment, we combine rule language with the controlled stimuli of Love (2002). This design also helps in the interpretation of Experiment 6 because the observed non-difference between Types II and IV could be attributed to either the materials or the lack of rule language in the instructions. We predict no difference between Type II and Type IV without rule language (as in Experiment 6). When rule language is included, this should lead to faster Type II learning, but it is unclear whether this is sufficient to produce a Type II advantage with normed stimuli.

Method

Participants. A total of 162 undergraduate students from St. Mary’s College of Maryland completed this experiment for partial fulfillment of a course requirement. Participants were randomly assigned to either the rule condition (N = 81) or the no-rule condition (N = 81).

Materials. The materials were the same as those used in Experiment 6.

Procedure. The procedure was the same as in Experiment 3; however, we included conditions with rule language in which participants were instructed to use “a rule that allows you to categorize the objects into two different categories.” The rule language was presented in the practice session and prior to the first and second problems.

Results and Discussion

To examine the course of learning and any role of the learning sequence of the problem types, a 2 (problem type) × 2 (rule condition) × 2 (sequence) × 8 (learning blocks) ANOVA was performed on the block-by-block accuracy data. In a surprising reversal of the traditional Type II advantage, mean accuracy for Type II (M = .667, SD = .134) was significantly lower than Type IV (M = .749, SD = .117), F(1, 158) = 47.767, p < .001, η² = .232. The overall difference in accuracy for the rule condition (M = .702, SD = .107) and the no-rule condition (M = .715, SD = .093) was not significant, F(1, 158) = 0.756, p = .386, η² = .005; however, the main effect of block was significant, F(7, 1106) = 185.220, p < .001, η² = .540. The interaction between rule condition and problem type was not significant, F(1, 160) = 0.072, p = .789, η² < .001. There was also an interaction between problem type and learning block that approached significance, F(1, 158) = 3.512, p = .063, η² = .022, reflecting a greater difference between Type II and Type IV in the earlier blocks than in the later blocks. All other main effects and interactions were not significant (ps > .1).

The proportion of participants that reached criterion was not significantly different for Type II (.301) compared to Type IV (.370), χ²(1) = 1.673, p = .196. To further examine differences in the proportion correct across learning and the proportion of participants that reached the learning criterion, we compared the rule condition and the no rule condition separately for each problem type. For Type II learners, the proportion correct was not significantly different between those in the rule condition (M = .662, SD = .138) and those in the no rule condition (M = .672, SD = .131), p = .664. The proportion of participants that learned Type II to criterion was not significantly different for the rule condition (.296) compared to the no rule condition (.309), χ²(1) = 0.029, p = .864. Likewise, for Type IV learning, proportion correct did not significantly differ between rule (M = .741, SD = .124) and no rule (M = .757, SD = .108) conditions (p = .379), and there was not a significant difference in the proportion of participants that reached criterion in the rule condition (.407) compared to the no-rule condition (.333), χ²(1) = 0.953, p = .329.

The pattern of results across our studies indicates that rule language alone is not sufficient to produce a Type II advantage. Moreover, the sum of available evidence suggests that a Type II advantage occurs only when the XOR relationship is mapped in particular ways onto particular physical materials along with rule instructions. While there may be other ways to get a Type II advantage, one has never been observed outside of these very particular circumstances.

The present results show not only a lack of a Type II advantage, but a reversal based on significantly better learning of Type IV. While this result, just like findings of a Type II advantage, may turn out to be more of an exception than the rule, it is notable to observe a full reversal of the traditional ordering. The notion of a universal SHU ordering clearly must be set aside, and accounting for the factors mediating the Type II–Type IV relationship is the critical challenge for models and theories. If the present results are not due to sampling and individual differences, what else might explain the poor learning here of Type II relative to Type IV? Why do rule instructions work in conjunction with standard geometric materials to produce a Type II advantage, yet have no impact with normed materials?

Selective attention does not provide a basis for explaining this divergence. Another possibility is that the key issue is verbalizability of the features of the stimuli. This is consistent with prior findings that monkeys and children, as well as adult humans facing integral dimension stimuli, do not learn Type II easily (Minda et al., 2008; Nosofsky & Palmeri, 1996; Smith et al., 2004). Focusing instructions encourage the learner to formulate explicit hypothetical rules or regularities underlying category membership. While Love’s (2002) materials are perfectly separable, they are presumably harder to talk or think about than the highly familiar and overlearned dimension values based on shape, size, and shading. Focusing instructions might not have their usual influence if the dimensions of the stimuli are cumbersome to focus upon. We developed a new set of stimuli specifically for testing Type II and Type IV with less verbalizable geometric materials. The predicted result is slow Type II learning—possibly to the point of a Type IV advantage.

Experiment 8

In Experiment 1, and in the rule instructions condition of Experiment 5, a Type II advantage occurred when learners were instructed to form a rule. With neutral instructions, we repeatedly found no difference between Type II and Type IV. Using Love’s (2002) materials, Type II learning was nearly the same (Experiment 6) or reliably worse (Experiment 7) than Type IV—possibly due to reduced verbalizability of the materials. In this experiment,
we made it hard for learners to easily posit, evaluate, accept, and apply a verbal rule or regularity. Rather than using an instructional manipulation or a dual-task methodology to interfere with verbal processing (e.g., Maddox, Ashby, Ing, & Pickering, 2004; Zeithamova & Maddox, 2006), we adopted another approach. Based on our experience studying the acquisition of XOR categories, it seems that a common way to acquire and apply the XOR rule is by developing a verbal mantra of sorts that one can refine and recite as needed during learning. Our hypothesis is that stimulus features that are substantially harder to talk about might undermine Type II learning. To be clear, these are still fully separable, noticeable, analyzable, and distinguishable dimensions—it is just that they do not readily translate into short, easy, clear, and obvious descriptors.

Instead of learning a category based on “white squares and black triangles,” learners in the present study face a category based on “speckled rectangles and checkerboard diamonds” (see Figure 7). These stimulus features are not fundamentally resistant to verbalization (as is the case with integral dimension stimuli). The difference here reflects whether the verbal coding of the perceptual content is overt, highly familiar, easily brought to mind, or—in more phonological terms—compact, fluent, and high frequency. We envision the learner having no trouble discerning which dimension value is displayed, but facing a less fluid process of constructing (and performing further operations on) a verbal coding.

Our prediction is that an SHJ experiment with stimulus features that are specifically harder to verbally code will lead to difficulty in Type II learning, but will leave Type IV unaffected. Without entering the debate on explicit versus implicit, declarative versus procedural, one system versus two, and so forth (see Ashby, Alfonso-Reese, Turken, & Waldron, 1998; Erickson & Kruschke, 1998; Nosofsky & Johansen, 2000; Reber & Squire, 1994; Stanton & Nosofsky, 2007), we seek to determine whether observations of a Type II advantage derive not only from an explicit invitation to form rules but also from the presence of easily coded features that function highly effectively as elements over which to articulate rules or rule-like regularities.

**Method**

**Participants.** A total of 215 Binghamton University students participated in this experiment for course credit. Of the participants, 122 were randomly assigned to the Type II category structure, and 93 were randomly assigned to the Type IV category structure. Dimension assignment was counterbalanced across the three Type II subtypes and four Type IV subtypes.

**Materials.** The stimuli consisted of geometric shapes varying along the standard geometric dimensions, but with less overt and “harder to talk about” differences in the dimension values. Instead of squares and triangles, the shapes were diamonds and rectangles. Instead of clear differences in the interior fill color, there were differences in interior texture (speckled or checkerboard pattern). The relative difference in size was reduced (28 mm2 vs. 48 mm2) relative to our prior studies.

**Procedure.** The instructions did not include rule language, and the procedure was the same as in Experiment 2.

**Results and Discussion**

In a nearly significant reversal, a 2 (problem type) × 8 (learning block) ANOVA revealed a trend toward a main effect of problem type, \( F(1, 213) = 3.381, p = .067, \eta^2 = .016 \). The performance of Type II learners (\( M = .649, SD = .160 \)) was marginally worse than that of Type IV learners (\( M = .683, SD = .093 \)). There was also an interaction between problem type and learning block, \( F(7, 1491) = 3.40, p = .001, \eta^2 = .016 \). The significant interaction is driven by an advantage for Type IV over Type II in the first half of the learning blocks. In an analysis restricted to the first four blocks of learning, the Type IV advantage is highly significant, \( F(1, 81) = 8.446, p = .004, \eta^2 = .045 \) (restricting analysis to the last four blocks of learning results in no significant difference between the problem types; \( p = .825 \)). The time course interaction suggests that Type II learners may be able to compensate or benefit from practice over time with the low-verbalizable stimulus features. The near-advantage of Type IV occurs in conjunction with the usual finding that the proportion of participants reaching the learning criterion was significantly greater for Type II (.320) than Type IV (.254), \( \chi^2(1) = 22.924, p < .001 \). Even when Type IV is generally better learned, Type II is more likely to show fast mastery.

A one-way ANOVA addressing subtype differences within Type II shows a significant difference, \( F(2, 119) = 3.802, p = .025, \eta^2 = .060 \), and a Tukey’s post hoc test revealed a marginal difference favoring the size-irrelevant subtype (\( M = .704, SD = .186 \)) over the shape-irrelevant subtype (\( M = .609, SD = .145 \)), \( p = .054 \), as well as a trend favoring the color-irrelevant subtype (\( M = .677, SD = .156 \)) over the shape-irrelevant subtype, \( p = .083 \). An ANOVA comparing the three Type II subtypes with Type IV was significant, \( F(3, 179) = 3.669, p = .013 \), and Dunnett’s post hoc test showed that the shape-irrelevant subtype of Type II was significantly worse than Type IV (\( p = .028 \)).

As predicted, Type II learning was adversely affected relative to Type IV when the differences in item shape, size, and shading were more difficult to verbalize. Participants appear to have had the most trouble formulating a category based on “slightly larger speckled things and slightly smaller checkerboarded things.”

---

**Figure 7.** a. Examples of stimuli used in Experiments 1, 2, and 4. b. Examples of stimuli used in Experiment 8.
General Discussion

These eight experiments, in conjunction with the review of the SHJ literature, provide clear evidence against a generalized Type II advantage. We provide (see Figure 8) an illustrated overview that summarizes key findings across the series of experiments. Experiments 2–4 represent three separate comparisons of Type II and Type IV that do not show evidence of a difference in mean accuracy during learning. We also find that a Type II advantage can occur, although it appears to require specialized circumstances: explicit instructions to the learner encouraging rule formation and particular choices of physical stimuli and dimension assignment (subtype). Experiments 6–8 extend our evidence to normed stimuli and suggest that verbalizability, rather than selective attention, may be the best explanation of the observed range of Type II performance. Instructions and materials that do not facilitate verbalization lead to slower Type II learning and can produce a reliable reversal in the form of a Type IV advantage. Further research is needed, but our findings lay the foundation for a revision of our understanding of the SHJ ordering and its place in revealing the nature of human category learning.

Several empirical studies in the literature (Love, 2002; Minda et al., 2008; Nosofsky & Palmeri, 1996; Smith et al., 2004) show that particular types of learners or learning settings produce a reversal of the traditional Type II advantage such that Type II is actually harder to learn. We highlight verbalizability and the degree of favorability for rule-like processing as a mediating factor. Love (2002) found that Type II was more difficult to learn under incidental (pleasantness ratings) than intentional (instructions to memorize items) unsupervised learning conditions. Love interpreted this finding in terms of intentional learning being more rule-driven. This fits well with our explanatory framework. Nosofsky and Palmeri (1996) used non-separable stimuli in an SHJ framework and found that Type II learning was nearly as hard as Type VI. While they interpreted this phenomenon in terms of reduced ability to allocate selective attention to integral dimensions, we point out that reduced ability to verbalize such dimensions is also a promising explanation. As such, the use of integral dimensions is an extreme version of our use of low-verbalizable dimensions. Finally, some types of learners, namely rhesus monkeys (Smith et al., 2004) and young children (Minda et al., 2008), show pronounced difficulty in Type II learning. While a number of theoretical interpretations can be made, the idea that monkeys and children have greater difficulty in verbalizing dimensions and/or decreased tendency to seek logical rules is compelling. In sum, we see general coherence and consistency between our present results and the existing literature on difficulty in Type II learning.

One limitation of the present research is our use of training sessions shorter than those in the well-known demonstrations of a traditional Type II advantage. We take the view that (1) the most interesting part of the learning curve is captured by our studies, (2) a Type II advantage—when found—has appeared early in training, and (3) existing learning curves offer little to suggest the likelihood of a late-emerging Type II advantage. Nonetheless, present evidence is limited to the period of learning during which mean accuracy reaches the 80%–90% level.

Aggregate Versus Individual Differences in Types II and IV

Two statistical signatures are consistently seen in conjunction with the observed lack of difference in mean accuracy between Type II and Type IV. The first is a difference in variability: The standard deviation for Type II is consistently higher—meaning that there is greater individual variation in performance. The second is that more learners reach the learning criterion in Type II even when overall performance is not better. This suggests that a subset of Type II learners make few errors on their way to rapid mastery—since aggregate performance is close to equivalent, it follows that there must also be a subset of Type II learners who make many errors and minimal progress in acquiring the classification.

Nosofsky, Gluck, et al. (1994) noted the potential issue of individual differences and stated that the distributions across learners (histograms) “appeared to be symmetric and unimodal” (p. 367). To our knowledge, none of the other existing SHJ experiments have been analyzed at the individual level. Figure 4 shows histograms for each of our experiments. By visual inspection, the Type II distributions for our basic SHJ experiments with geometric stimuli without rule language appear to be bimodal, while the Type IV distributions appear unimodal. Apparently, most Type II learners do not conform to the aggregate—instead, the average performance reflects a combination of fast, successful learners who make few errors, along with learners who show nearly chance-level performance across the learning session, that is, “failure to launch.” If this is the case, then theoretical models need to account for these two profiles in Type II learning (rather than for the aggregate). Importantly, the variation across SHJ experiments in terms of aggregate results can be understood in terms of the number of learners in a sample that fall into each of the two profiles. A small sample is more likely to diverge from the distribution of the population, so experiments with less power are more likely to show particularly high (or low) Type II aggregate performance. The potential for an exaggerated difference between sample and population is higher with a bimodal distribution than a normal distribution because of greater variability.

We conducted a set of analyses to validate the visual evidence of bimodality. The Shapiro–Wilk test of normality confirmed that the distribution of Type II learning accuracy was not normally distributed in any of the experiments ($ps < .001$; see

![Figure 8. An illustrated overview of key methodological differences and results across experiments. Filled-in stars represent the presence of a Type II advantage over Type IV based on average proportion correct. Empty stars represent a lack of Type II advantage. EXP = Experiment.](image-url)
Table 3). In all cases, this lack of normality was due to a significantly smaller degree of kurtosis than would be present in normally distributed data (zs < -2). In contrast, the distribution of Type IV learning accuracy was never significantly different from normal (all ps > .05). While this is strong evidence that the Type II distribution is not normally distributed and that Type IV is, we sought to quantitatively assess the bimodality of the distributions. We developed a metric of bimodality based on the total frequency of observations in the top and bottom ranges of accuracy relative to the frequency of observations in the middle range. The ranges were defined simply by grouping the bins into thirds (i.e., the middle range contained values ranging from .60 to .80). A normal distribution is one that is dominated by observations in the middle range, while the bimodal distribution of interest is one in which observations outside of the middle range in both directions are more likely. The bimodality metric is therefore computed as the product of the frequency of high and low accuracy scores divided by the square of the frequency of middle accuracy scores. As can be seen in Table 3, the bimodality calculation consistently yielded scores much higher for Type II than Type IV. While the bimodality metric was reliably close to zero (indicating a normal distribution) for Type IV learning, the bimodality metric indicated a product of frequency for extreme scores ranging upward of twice as large as the frequency of central scores for Type II. It is notable that for the two experiments in which a reversal and a near reversal of the Type II advantage was observed, the bimodality metric for Type II learners was markedly lower.

These analyses suggest two important theoretical considerations. The first is that Type II learning appears to have a bimodal nature based on relatively high likelihoods of both especially good and especially poor learning performance—accordingly, average performance does not match up well with actual individual learning performance. Second, this bimodal distribution can help explain variability across experiments in Type II performance. Smaller samples are more likely to misrepresent the population distribution than in normal distributions, and specific learning conditions can shift the balance between the two extreme learning profiles. The bimodality of Type II can be understood in terms of the category structure having a more all-or-none character as a logical rule based on XOR. Some learners may take an approach to the task that is more likely to reveal the XOR relationship, while others take an approach that is detrimental. This dynamic is only at play for the Type II category structure. The learning conditions, such as instructions and materials, appear to mediate these likelihoods.

Revising the SHJ Ordering and Assessing the Theoretical Implications

What do these findings mean for the SHJ ordering as a foundational empirical finding in the field of categorization and as a benchmark for theoretical models? As for the idea that Type II is generally easier to learn than Type IV, this is not a supportable claim. Both Type II and Type IV advantages can occur, and the modal finding (in this experimental series and the wider literature) is a non-difference. We have proposed that Type II learning shows two dominant profiles: One is rapid acquisition, and the other is relatively labored acquisition. These two profiles in combination readily lead to aggregate performance that is not reliably different from Type IV learning. However, factors having to do with the conditions of learning and the sampling of learners can shift the observed distribution so that Type II learning, in the aggregate, is facilitated or attenuated relative to Type IV.

What are the theoretical implications of this revised interpretation of human performance on the SHJ task? The initial impact of the SHJ ordering was to refute the pure stimulus generalization account of human category learning. This implication remains intact since stimulus generalization predicts Type II to be generally harder to learn than Types III–V and also predicts greater learning difficulty for Types I–V (especially Type I) than is actually observed (Shepard et al., 1961). In addition, there is no basis for explaining the bimodality or sensitivity to learning conditions of Type II learning.

Mechanistic models. First, we consider the competitors to the exemplar view that yielded poorer fits according to Nosofsky, Gluck, et al. (1994): Does the configural cue or rational model benefit from the new interpretation of Type II in the SHJ ordering? It appears that the configural cue and rational models should not be considered wrong for predicting no difference in aggregate learning between Types II and IV. However, few individual Type II learners actually look like the aggregate. In addition, it is a problem for the models if they can never predict rapid Type II learning because it is part of the overall phenomenology. It also remains the
case that abstraction-based approaches such as prototype models are fundamentally committed to predicting a general Type IV advantage over Type II that does not match the existing range of human data.

Models based on attentionally-mediated similarity to reference points, including ALCOVE (and the closely related generalized context model [GCM]; Nosofsky, 1984) and SUSTAIN, give an impressive account of the traditional SHJ ordering with a Type II advantage. As discussed, ALCOVE can predict a range of Type II/Type IV relationships depending primarily on the value of the free parameter controlling the use of selective attention. Based on a model analysis technique called parameter space partitioning, Pitt et al. (2006) characterized the global performance of ALCOVE across parameter space on the SHJ ordering. The authors found a set of 11 “universal” orderings produced by ALCOVE that capture general properties of the model’s behavior. Across parameter space, ALCOVE was generally committed to Type I being easily learned, to Type VI being difficult, and to Types III–V being equivalent, but the model was capable of extremely far-ranging predictions about Type II as either faster or slower than Type I, as well as faster or equal to Type IV. None of the universal patterns showed a Type IV advantage over Type II.

Kruschke (1992) showed that ALCOVE does not predict the traditional Type II advantage when selective attention is turned off: Types II and V were harder to learn than Types III and IV. With moderate attention learning and the values held constant on the other parameters, ALCOVE predicts a small, late-emerging Type II advantage. Nosofsky and Palmeri (1996) simulated human performance on the SHJ types with integral dimensions by turning off selective attention and optimizing the quantitative fit to human data using ALCOVE’s remaining parameters. This also resulted in Type II and Type V being harder to learn than Types III and IV; in addition, there was an overall slowing of learning across the problem types (by a factor of approximately two compared to Nosofsky, Gluck, et al., 1994). This is consistent with exemplar models having a pure stimulus generalization foundation that is supplemented by selective attention.

Based on this evidence, the bimodal Type II distribution in human learners could potentially be explained in terms of a group of learners using selective attention and a group of learners who do not. Accordingly, ALCOVE could generate a set of individual learning outcomes that show a strong Type II advantage (as in Nosofsky, Gluck, et al., 1994) and also a set of individual learning outcomes without selective attention that show a Type II disadvantage. We are interested to see whether parameter settings in ALCOVE can produce the two observed learning profiles and whether such settings can be plausibly linked to a psychological basis.

A challenge for ALCOVE is to explain non-rapid Type II learning without upsetting corresponding predictions for the other SHJ types. Our simulation results make clear that ALCOVE cannot predict near-immediate success in Type I learning under the parameter settings that slow down Type II learning—though exactly this pattern was seen in human learners. Additionally, the parameter settings that slow down Type II learning seem to have the same impact on Type V, but the equivalence of Types III–V is a robust characteristic of the SHJ ordering.

Love et al. (2004) have explained SUSTAIN’s account of the traditional SHJ ordering primarily in terms of the number of clusters formed by SUSTAIN in its modal solution for each category structure. Type II is generally solved by SUSTAIN with four clusters (one centered on each of the critical correlated features: [00-, 11-] in one category and [01-, 10-] in the other category), while Type IV is most often solved with six clusters (00-, 010, 100, 11-, 101, 011). SUSTAIN occasionally solves Type IV using just two clusters: one for each prototype. Contrary to the human data, SUSTAIN therefore predicts bimodality for Type IV, not Type II.

DIVA (Kurtz, 2007) captures the traditional qualitative ordering of the SHJ types, but it does not predict the extent of the Type II advantage observed by Nosofsky, Gluck, et al. (1994). In the current form of the model (Kurtz, 2012), a focusing (attentional) capability is introduced. As with other models, DIVA may be able to account for the observed human learning data by combining observations with and without focusing into a single distribution. Specifically, after DIVA generates activations at the output layer in the usual manner, the choice rule selectively weights the reconstructive error in accord with the output activations for that dimension on the category channels. As such, the reconstructive error for dimensions that are highly relevant to the category distinction play a larger role in determining membership probability, and the reconstructive error for irrelevant dimensions has a reduced impact. This increases the speed with which DIVA learns SHJ Types I and II but has little impact on category structures that require all three dimensions (including Type IV). Unlike ALCOVE, each network run with DIVA is unique due to the influence of the small random initial connection strengths. Consistent with human learning, preliminary tests show that DIVA shows higher variability across network runs for Type II than for Type IV. Full simulation studies are needed, but DIVA shows promise as a means of accounting for the revised SHJ ordering.

What are the implications of these findings for models that specify a separate rule-based learning module? The RULEX model (Nosofsky, Palmeri, & McKinley, 1994) was effectively fit to the traditional SHJ ordering and also showed some success as an account of individual differences in human category learning. However, if Type II is not generally easier to learn than Type IV, this is a problem of the highest order for a model based on the design principle of first testing all logical rules (in order of complexity) and then resorting to rules supported by memorization of exceptions.

While we do not intend to take a position on the separate versus single system debate, it is appropriate to make a few observations. First, the sensitivity of Type II learning to rule-promoting or rule-inhibiting learning conditions suggests that rules or something rule-describable are part of the toolkit used by learners in the artificial classification learning paradigm. Second, we would like to see the predictions of formal models that are based on separate explicit and implicit systems (e.g., COVIS [competition between verbal and implicit systems]; Ashby et al., 1998) for the SHJ ordering. Third, we note that rules can be effectively viewed as a type of abstraction that involves focusing on selected dimensions. As such, the theoretical divide may have more to do with the sensitivity to verbalizability than anything else. An intriguing direction we hope to pursue in future research is to evaluate the idea that internal verbalization of the basis for categorization may be bottom-up (based on monitoring of abstracted statistical regu-
larities) rather than top-down (based on hypothesis testing in a search space).

**Mathematical models.** In recent years, the SHJ ordering has also served as a test case for a number of computational-level models that operate at the level of mathematical formalism rather than psychological mechanism. What are the implications of the revised SHJ benchmark in this domain? Feldman (2000) proposed an algebraic complexity account in terms of Boolean compressibility that predicted a Type II advantage (the complexity of Type II was 4, while Types III–V had a complexity value of 6). This prediction matches the traditional Type II advantage, but not the revised view. Interestingly, a number of independent researchers (Lafond et al., 2007; Mathy & Bradmetz, 2004; Vigo, 2006) reported that Feldman failed to find the actual optimal compressions for Types III and IV (which have, respectively, complexity values of 4 and 5). This suggests that Type II is actually less (or not) distinct from Types III–V based on a Boolean complexity analysis. Feldman (2006) developed a new formulation of algebraic complexity that similarly predicts Type II to group with Types III–V. Therefore, there does appear to be a good correspondence between the Type II non-advantage observed in most SHJ studies and the formal complexity of the category structures. Vigo (2009) presented an alternative formalization based on invariance instead of complexity that also predicts a minimal difference between Types II and IV. To summarize, as opposed to the predictions of these models missing the mark by failing to predict the traditional Type II advantage, they instead appear to be consistent with human aggregate performance (though they do not speak to the issues of bimodality or sensitivity in Type II learning).

Goodman et al. (2008) proposed a rational rules model that combines aspects of logical rule induction, rational (Bayesian) modeling, and Boolean complexity. The rational rules model predicts a Type II advantage except under a special parameter setting in which it predicts a reversal (Type II harder to learn than Types III–V). This suggests a potential explanation of the bimodality and sensitivity of Type II. The key parameter, $b$, refers to the a priori likelihood that any given example is an outlier and determines whether the model tends to form simple or complex rules. The model therefore predicts that Type II is harder to acquire when the learner or the learning setting favors unidimensional rules—this makes sense because such rules provide no traction on XOR. It is worth further exploration to see whether the bimodality and sensitivity of Type II learning can be understood in terms of Type II being easy to learn unless one is strongly committed to finding a simple rule.

In sum, a number of modeling approaches show the potential to account for the revised ordering. We are particularly interested to see what fits can be achieved using established mechanistic models or improvements upon them. While ALCOVE appeared to offer a strong account of Type II variability based on selective attention, our behavioral results and simulation study contradict this theoretical account in its current form. A further implication of the present findings is that models should be evaluated for their fit to performance distributions to strike a useful compromise between fitting aggregate and individual difference data (see Ashby, Maddox, & Lee, 1994; Estes, 1956; Myung, Kim, & Pitt, 2000). Specifically, when there is a set of qualitatively distinct profiles underlying aggregate performance, it is more important to show whether (and how) models can capture these profiles than it is to show a fit to aggregate performance that is not psychologically consequential.

**Summary**

We conclude by emphasizing that the present investigation supports a revision of the SHJ ordering in which (1) Type II is not generally easier to learn than Type IV, and (2) the ease of Type II acquisition is critically mediated by properties of the task setting and the individual learner. Under our best efforts to test participants in the most basic experimental circumstances, there was no evidence of an aggregate difference in ease of learning. The results of our studies, combined with evidence from the existing literature, show that learners who approach the Type II task with an orientation toward the development of a verbalizable rule-like basis for classification show faster acquisition, while learners without such an orientation show slower acquisition. The common interpretation of the SHJ ordering as reflecting stimulus generalization supported by dimensional selective attention is not a satisfactory account. In light of these findings, Shepard et al.’s (1961) six types may produce yet another wave of theoretical impact as researchers evaluate and extend formal models to account for the varieties of Type II category learning.

**References**

Summarizing Monte Carlo results in methodological research: The one- and two-factor fixed effects ANOVA cases. Journal of Educational Statistics, 17, 315–339.


Received October 26, 2011
Revision received March 1, 2012
Accepted April 16, 2012