***** Article in press in *Psychological Research* *****

LEARNING AND TRANSFER OF CATEGORY KNOWLEDGE IN AN INDIRECT CATEGORIZATION TASK

Sebastien Helie & F. Gregory Ashby

University of California, Santa Barbara

Running head: Learning and transfer in indirect category learning

For correspondence,

Sebastien Helie
Department of Psychological & Brain Sciences
University of California, Santa Barbara
Santa Barbara, CA 93106-9660

Phone: (805) 284-9474
Fax : (805) 893-4303
E-mail: helie@psych.ucsb.edu

Word count (main text only): 6,281

Version RR1, last modified May 24th, 2011.

Abstract

Knowledge representations acquired during category learning experiments are 'tuned' to the task goal. A useful paradigm to study category representations is indirect category learning. In the present article, we propose a new indirect categorization task called the "Same" – "Different" categorization task. The same-different categorization task is a regular same-different task, but the question asked to the participants is about the stimulus category membership instead of stimulus identity. Experiment 1 explores the possibility of indirectly learning rule-based and information-integration category structures using the new paradigm. The results suggest that there is little learning about the category structures resulting from an indirect categorization task unless the categories can be separated by a one-dimensional rule. Experiment 2 explores whether a category representation learned indirectly can be used in a direct classification task (and vice-versa). The results suggest that previous categorical knowledge acquired during a direct classification task can be expressed in the same-different categorization task only when the categories can be separated by a rule that is easily verbalized. Implications of these results for categorization research are discussed.

Keywords: indirect category learning, categorization, same-different task, COVIS.

Learning and transfer of category knowledge in an indirect categorization task

Learning about categories is an important cognitive endeavor, but it is rarely an end: category learning is useful in that it dictates how to interact with or use specific objects (Markman & Ross, 2003). As a result, Markman and Ross (2003) argued that the knowledge representations acquired during category learning are 'tuned' to the task goal. A useful paradigm to test this hypothesis is indirect category learning (Brooks, Squire-Graydon, & Wood, 2007; Minda & Miles, 2010; Minda & Ross, 2004). In an indirect category-learning task, participants are not asked to make a classification decision, but learning the categories will improve their performance (Minda & Ross, 2004). For instance, Minda and Ross (2004) asked participants to decide how much food was needed to feed a set of artificial creatures. The creatures belonged to two separate categories, and category membership was a predictor of food consumption (the other factor was the size of the creature). One group of participants was not told about the categories and only received corrective feedback on the amount of food selected (the *indirect* group), whereas the other group had to make a categorical judgment on each trial (followed by categorization feedback) before deciding how much food to feed the creature (the *direct* group). The results showed that participants in these two conditions learned different category representations. Briefly, the performance of participants who learned the categories indirectly was suggestive of a similarity-based representation whereas the performance of participants who learned the categories directly appeared to be rule-driven. Brooks et al. (2007) found similar results in a different indirect category-learning task also involving a limited set of artificial creatures.

The goal of this article is twofold. First, previous indirect category learning research has focused on complex conditions in which a limited set of stimuli were presented, and the categories could be learned using either a similarity-based or rule-based strategy. Here, we try to better control the participant response strategy by using a new indirect category-learning task called the "Same"-"Different" categorization task. This paradigm is intuitively simpler and relies on a well-known paradigm (i.e., the same-different task; e.g., Bamber, 1969; Krueger, 1978; Thomas, 1996). Its simplicity should make it easier to determine whether errors are due to the categorization process, or to processes unique to the goal-directed task. In addition, we used the randomization technique from Ashby and Gott (1988) to generate a large number of stimuli, and the possibility of indirectly learning rule-based and information-integration category structures is explored. Second, because the category representations that are learned directly and indirectly differ, a follow-up question is whether the representations learned in one context can be used in another context. Specifically, can a category representation learned indirectly be used in a direct classification task (and vice-versa)? This paper is an initial attempt at answering both of these questions using a new indirect categorization task, the same – different categorization task.

Direct classification

In a typical classification task, a stimulus is presented on each trial and the participant depresses a response key indicating the category membership of the stimulus. Feedback is then provided, and a new trial is initiated (Ashby & Maddox, 2005). For instance, the stimuli might be circular sine-wave gratings (e.g., Gabor disks) varying in bar width and bar orientation. To visualize the category structures, each stimulus can be represented graphically by a point in a two-dimensional space like those shown in Figure 1 (with each axis representing a different stimulus dimension). Many studies have reported striking differences in how people learn with *rule-based* (the top panels and the bottom-right panel) versus *information-integration* (the bottom-left panel) category structures (e.g., Ashby, Ell, & Waldron, 2003; Ashby, Maddox, & Bohil, 2002; Maddox, Ashby, & Bohil, 2003; Maddox & Ing, 2005; Waldron & Ashby, 2001). In rule-based tasks, the optimal categorization strategy can be learned using an explicit reasoning process and is often easy to describe verbally (Ashby, Alfonso-Reese, Turken, & Waldron, 1998). For instance, the top-left panel of Figure 1 shows the simplest and most widely studied rule-based category structures. Note that the optimal one-dimensional rule here is "respond A if the bars are thick and B if they are thin." There are similar verbal rules for the two category structures shown in the right column of Figure 1. In contrast, the optimal categorization strategy for the information-integration category structures shown in the bottom-left panel of Figure 1 is difficult or impossible to describe verbally. Accuracy is maximized only if information from two or more stimulus dimensions is integrated at some pre-decisional stage (Ashby & Gott, 1988).

Insert Figure 1 about here

Previous research has found many qualitative dissociations between the learning performance of rule-based and information-integration category structures (for a review, see Maddox & Ashby, 2004). However, human participants can reliably learn all these different category structures with a direct classification task (for a review, see Ashby & Maddox, 2005).

Indirect category learning: The same-different categorization task

The main characteristic of indirect category learning tasks is that categorization is only part of the process that is required to achieve another goal. Furthermore, the response and feedback are directly related to the goal; not to categorization. Here we propose a new task that fits these criteria.

The same-different categorization task enhances the direct classification paradigm by forcing participants to learn the category structures indirectly. Precisely, the participants are shown two stimuli simultaneously and their task is to depress one button if the two stimuli belong to the same category and another button if the two stimuli are from different categories. Essentially, the same-different categorization task is a regular same-different task (e.g., Bamber, 1969; Krueger, 1978; Thomas, 1996), but the question asked to the participants is about category membership instead of stimulus identity. The participants need to learn the categories in order to maximize accuracy, but no stimulus is ever associated with a particular response key, and direct categorization feedback is never provided.

Possible response strategies

The same-different categorization task can be performed in many different ways. First, the participants could separately categorize each stimulus on the screen and then compare their category labels. If both stimuli are assigned to the same category, the participant responds “Same”. Otherwise, the participant responds “Different”. This strategy could lead to perfect accuracy if the categories are learned correctly. However, it is also possible that participants will not categorize the stimuli separately – either because they are unable to learn the adequate category representations or because they simply opt for another, possibly simpler, strategy. Two strategies that do not require separately categorizing each stimulus are similarity-based strategies and guessing strategies. One way for participants to use a similarity-based strategy is to compare the distance between the two stimuli in perceptual space to a threshold. Distances smaller than the threshold would elicit a "Same" response, whereas distances greater than the threshold elicit a "Different" response. This strategy is likely to be suboptimal because with many categories, there will be stimuli that are more similar to some stimulus in the contrasting category than to other stimuli in the same category. Another strategy, which is obviously suboptimal, but exceedingly simple, is to just guess. Our analysis includes fitting computational models that will attempt to identify which of these strategies was used by each participant.

Relation to previous indirect category learning tasks

The same-different categorization task is similar to the indirect category-learning procedures used by Brooks et al. (2007) and Minda and Ross (2004) in the sense that participants do not make categorization responses, and the feedback is not about categorization. We believe that this is the essence of indirect category learning. Even so, the same-different categorization task does differ from these previous tasks on other potentially important aspects. First, we explicitly told participants that the stimuli belong to two different categories. This is similar to Brooks et al. (2007), but different from Minda and Ross, who did not tell their indirect-learning participants that the stimuli belonged to categories. Hence, in addition to being indirect, learning of the categories in the Minda and Ross condition was incidental. Second, the goal-directed task is much simpler here than in previous experiments. In Minda and Ross (2004), participants were required to estimate the amount of food required to feed a creature (Minda & Ross, 2004), whereas Brooks et al. (2007) asked participants to determine the number of moves on a chessboard required to reach a target. In our task, participants were simply asked whether the two stimuli belonged to the same or different categories. Because this task does not require any extra computation (e.g., estimation), it should be easier to attribute errors to failures of categorization. In any case, these differences are not critical to indirect category learning and should not change the conclusion that a different category representation is learned and used when the goal-directed task is changed.

Overview of the experiments and hypotheses

Experiment 1 is a first test of the same-different categorization task. Specifically, the goal of this experiment is to introduce the new paradigm and to test whether participants can learn rule-based and information-integration category structures by making same-different judgments about category membership. According to the COVIS theory of categorization (Ashby et al., 1998), participants should categorize the stimuli in the rule-based conditions using an explicit hypothesis-testing strategy that is flexible and does not depend on the consistency of the stimulus-response association; what is important is the consistency of the stimulus-category association. In contrast, participants should categorize the stimuli in the information-integration condition using an implicit procedural-learning system that heavily depends on the consistency of the stimulus-response mapping (Ashby et al., 2003). In the same-different categorization task, there is no systematic stimulus-response association: a particular stimulus can be associated with the “Same” or “Different” response depending on what stimulus is simultaneously displayed. Hence, COVIS predicts better same-different categorization performance in the rule-based conditions than in the information-integration condition.

Experiment 2 tests whether the representations learned in a direct classification task can be transferred to the same-different categorization task, and if the category structures learned in the same-different task can be transferred to a direct category-learning task. According to COVIS, rule-based categorization is a two-stage process and the category representation is separate from both the stimulus and response representations. Hence, knowledge should at least partially transfer from one task to the other. In contrast, information-integration categorization is learned using a procedural-based system and, given the importance of the consistent stimulus-response mapping, it seems likely that no intermediate category representation is present. If this is the case, the learning in information-integration conditions should be task specific and no transfer should be observed.

Experiment 1

This experiment introduces the same-different categorization task and tests whether participants can learn both rule-based and information-integration category structures by making same-different judgments about category membership.

Method

Participant

Fifty-nine undergraduate students at University of California Santa Barbara were recruited to participate in Experiment 1. Twenty participants were trained in the 1D-Width condition, 19 were trained in the 1D-Orientation condition, and the remaining 20 participants were trained in the information-integration condition. Each participant was given credit for participation as partial course requirement.

Stimuli and Apparatus

The stimuli were circular sine-wave gratings of constant contrast and size (for an example stimulus, see Helie, Waldschmidt, & Ashby, 2010) presented on a 21-inch monitor (1,280 × 1,024 resolution). Each stimulus was defined by a set of points (x_1, x_2) sampled from a 100 × 100 stimulus space and converted to a disk using the following equations: *width* = $x_1/30+0.25$ cpd, and *orientation* = $9x_2/10+20°$. This yielded stimuli that varied in orientation from 20° to 110° and in bar width between 0.25 and 3.58 cpd. The stimuli were generated with Matlab using Brainard's (1997) Psychophysics Toolbox and occupied an approximate visual angle of 5°. In each trial, two stimuli were presented simultaneously. The stimuli were centered vertically and the mid-point between the two stimuli was centered horizontally. The horizontal distance between the two stimuli was approximately 5° of visual angle. The category structures are shown in Figure 1.

For the 1D-Width condition (top-left), category "A" stimuli were generated using a multivariate normal distribution with the following parameters (Ashby & Gott, 1988): $\mu_a = \{40, 50\}$; $\Sigma_a = \{10, 0; 0, 280\}$. The same sampling method was used to generate category "B" stimuli: $\mu_b = \{60, 50\}$; $\Sigma_b = \Sigma_a$. Stimuli in the 1D-Orientation condition were obtained by rotating the 1D-Width stimuli by 90° counterclockwise (top-right), and the stimuli in the information-integration condition were obtained by rotating the 1D-Width stimuli by 45° counterclockwise (bottom-left). Note that perfect accuracy was possible in all three conditions.

Stimulus presentation, feedback, response recording, and response time (RT) measurement were acquired and controlled using Matlab on a Macintosh computer. Responses were given on a standard Macintosh keyboard: the “d” key for a “Same” response and the “k” key for a “Different” response (sticker labeled “A” and “B” respectively). Visual feedback was given for a correct (green checkmark) or incorrect (red “X”) response. If a response was too late (more than 5 seconds), participants saw the words “Too Slow”. If a participant hit a wrong key, the words “Wrong Key” were displayed.

Procedure

The experiment lasted about 60 minutes and was composed of 12 blocks of 50 trials (for a total of 600 trials). Participants were told that the stimuli could be separated into two categories and that their task was to decide whether the two stimuli presented on the screen in each trial were drawn from the same or different categories. A trial went as follows: a fixation point (crosshair) appeared on the screen for 1,500 ms and was followed by the stimuli. After the participants made a response, correct or incorrect visual feedback was given for 2,000 ms, with the stimuli remaining on the screen for the entire duration of feedback. The participants were allowed to take a break between blocks if they wished.

Result

Accuracy

The mean accuracy per block for each condition is shown in Figure 2. A Condition (3, between) × Block (12, within) ANOVA shows statistically significant effects of Block [$F(11, 616) = 2.09$, $p < .05$] and Condition [$F(2, 56) = 63.9$, $p < .001$]. The mean accuracies in Block 1 were 86.3% (1D-Width), 74.8% (1D-Orientation), and 54.1% (information-integration). Accuracies increased to 86.3% (1D-Width), 82.5% (1D-Orientation), and 57.0% (information-integration) in Block 12. Separate within-subject *t*-tests comparing the performances in the first and last blocks in each condition suggest that participants in the 1D-Orientation condition were the only ones who statistically improved with practice [an improvement of 7.7%; $t(18) = 2.59$, $p < .05$]; the performance of participants in the other conditions improved by less than 3%, which was not statistically significant [both *t*s(19) < 0.21, *n.s.*]. The absence of improvement in the 1D-Width condition is likely to be a ceiling effect, because an overall accuracy of 86.3% suggests that each stimulus was categorize with an accuracy of 93.2% (see Eq. 2 in the *Transfer accuracy* section of *Experiment* 2). In addition, *Posthoc Tukey HSD*s show that, overall, the one-dimensional rule groups were not statistically different, and that the participants in the two rule groups were more accurate than the participants in the information-integration group (both $p < .001$). Together, these analyses suggest that participants can do the task (and learn how to do the task) in the rule-based conditions but not in the information-integration condition. The interaction did not reach statistical significance [$F(22, 616) = 1.09$, *n.s.*].

Insert Figure 2 about here

Model-based analyses

The accuracy-based analyses suggest that participants could achieve good performance in the two rule-based conditions but not in the information-integration condition. Yet, it is important to know whether each participant eventually adopted a decision strategy of the optimal type. For instance, the inability to perform in the information-integration condition can stem from an incorrect categorization strategy or a difficulty in consistently applying the appropriate strategy. To answer this question, we fit four different types of decision-bound models (e.g., Maddox & Ashby, 1993) to the data from each individual participant: rule-based, information-integration, similarity-based, and guessing models. The rule-based models assumed either a single vertical or a horizontal bound, or that participants used a conjunction rule. The information-integration model assumed that the decision bound was a single line of arbitrary slope and intercept. For all the above models, it was assumed that the participants individually categorized the stimuli and then compared the outcomes.

However, it is also possible that the participants did not individually categorize the stimuli. As their name implies, the guessing models assumed that participants randomly chose a response on each trial, without considering the individual category membership of the stimuli. Finally, the similarity model calculated a weighted exponential distance between the stimuli and assumed that participants responded “Same” for small distances and “Different” for large distances. The similarity model had two free parameters, one to differentially weight the stimulus dimensions and another to describe the slope of the exponential distance. Like the guessing model, the similarity model also does not assume separate classification of the stimuli.

The results from the model-based analyses are shown in Table 1. As can be seen, most participants in the rule-based conditions appeared to be responding optimally. Furthermore, note that the responses of these participants were more likely to be best-fit by an optimal model later in training. A one-tail binomial test showed that the difference in proportion of best-fitting optimal models between early and late performance was statistically significant for the 1D-Width condition (p < .05) but not for the 1D-Orientation condition. Even so, the performance of most participants in both rule conditions was best fit by an optimal categorization model by the end of training. In the information-integration condition, only one participant used an optimal strategy at the beginning of the experiment, and no participant was using an optimal categorization strategy by the end of the experiment. A one-tail binomial test showed that this decrease in the proportion of optimal best-fitting models was not significant. It should be noted that none of the participants in any block was best fit by a similarity model. Participants not best fit by an optimal categorization model were best fit by a guessing model.

Insert Table 1 about here

Discussion

The results show a strong effect of category structure on performance in an indirect category-learning task. First, participants in rule-based conditions were fairly accurate in making same-different judgments, and their responses were mostly consistent with an optimal categorization strategy, suggesting that they learned the correct category representations. In contrast, participants in the information-integration condition performed only slightly better than chance throughout the whole experiment, and the model-fitting analyses suggest that they were not basing their same-different responses on accurate category representations. Yet, previous research has shown that participants can reliably learn all three category structures in direct classification (for reviews, see Ashby & Maddox, 2005; Maddox & Ashby, 2004). The failure of the information-integration participants is consistent with at least two different hypotheses. One possibility is that these participants learned little or nothing about the information-integration categories, but an alternative is that there was some category learning, but the participants were unable to apply that knowledge to the same-different judgment required in the task. Experiment 2 tests between these two possibilities.

Experiment 2

Experiment 2 tests whether participants can transfer knowledge gained during direct classification to the same-different task and vice versa. For example, if the information-integration participants in Experiment 1 learned the underlying categories but were unable to apply that knowledge to the same-different judgment then performance on a subsequent classification task with the same categories should be enhanced relative to a control group that did not have the prior same-different training. Experiment 2 will allow us to test for this possibility.

Performance on the one-dimensional rule-based category structures from Experiment 1 did not statistically differ and these one-dimensional conditions were easier than the information-integration condition. For this reason, the 1D-orientation condition was replaced by a more difficult rule-based condition (i.e., a conjunction rule).

Method

Participant

Ninety undergraduate students at University of California Santa Barbara were recruited to participate in Experiment 2. Thirty participants were trained with the 1D-Width category structure (Figure 1, top-left), 30 participants were trained with the information-integration category structure (Figure 1, bottom-left), and the remaining 30 participants were trained with the conjunction rule (Figure 1, bottom-right). Each participant was given credit for participation as partial course requirement. None of the participants took part in Experiment 1.

Stimuli and Apparatus

The stimuli were the same as in Experiment 1. The category structures for the 1D-Width and information-integration conditions were the same as in Experiment 1 (Figure 1, left column). In the conjunction group, category "A" stimuli were generated from two multivariate normal distributions with the following parameters (Ashby & Gott, 1988): μ_{a1} = {30, 50}; Σ_{a1} = {10, 0; 0, 150} and μ_{a2} = {50, 70}; Σ_{a2} = {150, 0; 0, 10}. A similar sampling method was used to generate category "B" stimuli: μ_{b1} = {50, 30}; μ_{b2} = {70, 50}; $\Sigma_{b1} = \Sigma_{a1}$; and $\Sigma_{b2} = \Sigma_{a2}$ (Figure 1, bottom-right).

Stimulus presentation, feedback, response recording, and RT measurement were acquired and controlled using Matlab on a Macintosh computer. For the same-different categorization task, the material was the same as in Experiment 1. For the direct classification task, a single stimulus was displayed at the center of the screen in each trial. Responses were given on a standard Macintosh keyboard: the "d" key for an "A" categorization and the "k" key for a "B" categorization (sticker-labeled as either "A" or "B"). Auditory feedback was given for a correct (high pitched tone) or incorrect (low pitched tone) response. If a response was too late (more than 5 seconds), participants saw the words "Too Slow". If a participant hit a wrong key, s/he heard a distinct beep and saw the words "Wrong Key".

Procedure

The experiment lasted for two sessions scheduled during the same week. Each session was composed of 12 blocks of 50 trials (for a total of 600 trials). In Session 1, half the participants in each condition were trained in the direct classification task while the remaining participants were trained in the same-different categorization task. The task practiced during Session 1 was called the *training task*. In Session 2, the first two blocks (100 trials) were done in the training task. The remaining 10 blocks (500 trials) were done in the other (unpracticed) task with the same category structures. This second task was called the *transfer task*. The participants were told that the same categories were used in the training and transfer tasks.

The procedure for the same-different categorization task was identical to that of Experiment 1. For the direct classification task, participants were told they were taking part in a categorization experiment and that they had to assign each stimulus into either an “A” or a “B” category. A trial went as follows: a fixation point (crosshair) appeared on the screen for 1,500 ms and was followed by the stimulus, which remained on the screen until the participant made a response; correct or incorrect auditory feedback was given for 1,000 ms; “wrong key” or “too slow” feedback was given for 2,000 ms. The participants were allowed to take a break between blocks if they wished.

Results

The data from the second session of two participants that were trained in the same-different categorization task with the information-integration category structures were missing (and not included in the analyses).

Training Accuracy

The mean accuracy per block is shown in Figure 3. A Condition (6, between) × Block (14, within) ANOVA on the training task accuracies shows a significant effect of Block [$F(13, 1053) = 7.19$, $p < .001$], Condition [$F(5, 81) = 51.38$, $p < .001$], and their interaction [$F(65, 1053) = 2.07$, $p < .001$]. The mean accuracies in Blocks 1 and 14 are shown in Table 2. The interaction shows that participants in the three conditions whose training task was same-different categorization did not improve their performance with practice [all $Fs(13, 182) < 1.54$, *n.s.*].[1] The participants trained with the information-integration and conjunction category structures were unable to perform the task and had very low accuracies throughout. In contrast, the participants trained with the 1D-Width category structures were proficient at performing the task from Block 1, but did not improve.

Insert Figure 3 about here

In contrast, the participants trained with the classification task improved their performance with practice in all conditions [all Fs(13, 182) > 2.52, $p < .01$]. However, participants in the 1D-Width condition were more accurate than participants in the other two conditions throughout the experiment. To summarize, the participants were unable to perform the same-different categorization task after 14 blocks of practice (700 trials) with any category structure except the 1D-Width. However, the participants were able to perform direct classification after 14 blocks of practice (700 trials) with all three category structures.

[1] Because of the missing data, the number of degrees of freedom of the error term in the information-integration condition was 143.

Insert Table 2 about here

Transfer accuracy

A first measure of transfer is to ask whether accuracy was the same in Blocks 14 and 15. No difference implies perfect transfer. The accuracy differences are shown in Table 2 (Block 15-14). Paired-sample *t*-tests were performed in each group. The accuracies in all groups trained on categorization declined when they switched to the same-different task [all ts(14) < -2.94, p < .05]. For the participants trained with the same-different task, accuracy did not change when the task switched to direct classification (see Table 2).

A second measure of transfer is based on the assumption that participants in the same-different task independently categorized the stimuli and then optimally compared the classification responses.[2] Note that a participant using this strategy would make a correct same-different response if both stimuli were categorized correctly or if both were categorized incorrectly. Let p equal the probability that a single stimulus is categorized correctly and let Q equal the probability of a correct same-different response. Then

$$Q = p^2 + (1 - p)^2. \qquad (1)$$

Thus, if participants are responding "Same" or "Different" by first categorizing each stimulus, then the accuracy in the first transfer block (i.e., Q) should be related to the accuracy in the last training block (i.e., p) via Eq. 1. Likewise, the predicted accuracy in

[2] An assumption made for all the model-based analyses except the similarity model (which was never the best-fitting model in our analyses) and the guessing models (which was only used when participants could not learn the task).

the first transfer block (Block 15) of participants who trained on the same-different task and then transferred to direct classification should be related via

$$p = \frac{1+Q}{2} \tag{2}$$

(i.e., solve Eq. 1 for *p*). These predicted accuracies are shown in Table 2 (in parentheses in Block 15). A separate χ^2 test was performed with each training task to compare the predicted and observed accuracy distributions. For participants trained with direct classification, the predicted and observed accuracies were similar [$\chi^2(2) = 0.36$, *n.s.*], suggesting that participants could use the knowledge acquired during a direct classification task to perform the same-different categorization task. This is in line with the decreased performance when the participants switched to the same-different categorization task. In contrast, participants trained with the same-different categorization task did not seem to use their category knowledge when transferred to the direct classification task [$\chi^2(2) = 17.05$, $p < .001$]. In all three conditions, the participant performance should have improved substantially when transferred to the direct classification task (because only one categorization judgment is required). The previous analyses showed that their performances did not statistically change when transferred to the direct classification task.

A third measure of transfer is to ask whether accuracy in the first transfer block (Block 15) is the same as the first-block performance of other participants who were originally trained with the same task and the same category structure (e.g., compare the first transfer block of Cat – II with the first training block of SD - II). If there was transfer, the first block of the transfer task should have been better than the first block of the corresponding training task. These accuracy differences are shown in Table 2

(Transfer gain). Independent sample *t*-tests were performed in each condition. Only the group performing direct classification with the conjunction rule showed evidence of transfer [$t(28) = 4.69$, $p < .001$]. The same-different categorization task was easier after previous training in direct classification with the conjunction rule. All other conditions showed no sign of positive or negative transfer [all $|ts(28)| < 1.51$, *n.s.*].[3]

Model-based analyses

The model-based analyses are shown in Table 3. Focus first on the participants trained with the classification task. As can be seen, the responses of most participants trained with rule-based category structures were best fit by an optimal model at the end of training. However, only one third of the information-integration participants were best fit by an optimal model at the end of training (most of the participants in this condition approximated an information-integration strategy by using a conjunction rule). Interestingly, most of the rule-based participants adequately transferred their category knowledge to the same-different categorization task. However, none of the information-integration participants transferred a categorization strategy that was best described by an optimal model. As in Experiment 1, most of the information-integration participants guessed when transferred to the same-different categorization task (and more training led to more guessing).

Insert Table 3 about here

[3] The independent *t*-test for the classification with information-integration condition had only 26 degrees of freedom (because of the missing data).

For the participants trained with the same-different task, only the participants in the 1D-Width condition produced a response pattern that could be described by an optimal strategy. Most of these participants were able to transfer their category knowledge when the task changed. In contrast, participants trained with the information-integration or the conjunction rule category structures were not using an optimal categorization strategy (most of these participants were either guessing or using a one-dimensional rule). This is not surprising since the best performance that these participants were able to achieve was 53.2% and 56.8% (respectively; see Table 2). However, these participants were able to learn an appropriate categorization strategy when they switched to the classification task. Note that, as in Experiment 1, the performance of none of the participants in any of the conditions was best fit by a similarity model.

To summarize, participants trained with the 1D-Width category structure could learn, use, and transfer their category knowledge regardless of the training/transfer task. In contrast, participants trained with the information-integration category structure could only learn and use an optimal strategy in the direct classification task, and no transfer of category knowledge was observed. Finally, participants trained with the conjunction rule could learn, use, and transfer their category knowledge when trained with the direct classification task. However, an optimal decision strategy could not be learned/transferred when trained with the same-different categorization task. Still, an optimal strategy could be used in the same-different categorization task if it was previously learned in a direct classification task.

Discussion

Experiment 2 shows two important results. First, the training results suggest that task difficulty may be the main factor in indirect category learning; not category structure. On the one hand, the 1D-Width condition was the easiest, and could be learned equally well with direct or indirect categorization tasks. On the other hand, the conjunction rule and the information-integration category structures were more difficult and could only be learned directly. As in Experiment 1, representations of difficult category structures could not be learned indirectly; only representations of easy category structures could be learned indirectly. However, representations could be learned directly regardless of the category structure.

Second, the transfer results suggest that rule-based strategies that were learned directly using classification could be transferred to the same-different categorization task, but not information-integration strategies that were learned directly using the same classification task. This result is counter-intuitive because it suggests that participants in the information-integration condition know the categories of the individual stimuli but that they cannot make a sameness judgment.[4] Yet, this result is consistent with previous studies of rule-based and information-integration category structures (e.g., Ashby et al., 2003; Waldron & Ashby, 2001), which suggest that the former are more abstract (and general), whereas the latter are more procedural (and specific). It was also predicted by the COVIS theory of categorization (Ashby et al., 1998), which suggests that

[4] This result cannot be explained by task difficulty alone, because the performances on the last block of direct classification training in the information-integration and conjunction conditions were not statistically different [Block 14; $t(26) = 1.57$, *n.s.*].

information-integration category structures are learned using a procedural-based system that relies heavily on a consistent stimulus-response mapping. The consistent stimulus-response mapping present in the direct categorization task is broken when the participants are transferred to the same-different categorization task.

General Discussion

Two experiments tested the ability of human participants to learn the same category structures using either direct or indirect categorization tasks. The results showed that participants learned little about the category structures while they were making same-different category judgments unless the categories were separated by a one-dimensional category bound. Experiment 2 also showed that previous categorical knowledge acquired during a direct classification task can be expressed in the same-different categorization task, but only when the categories can be separated by a rule that is easily verbalized. These results suggest that there may be a limit to what can be learned indirectly about the world. It may be that much categorical knowledge can only be acquired directly. Furthermore, our results also suggest that much of this knowledge may be inaccessible to abstract reasoning. These two findings are consistent with the predictions made by COVIS (Ashby et al., 1998), which suggests that information-integration category structures are learned using a procedural-based system that has no separate category representation. In contrast, rule-based category structures are learned with an explicit hypothesis-testing system that has a separate category representation that allows for transfer of category knowledge.

Indirect category learning and task difficulty

In the present experiments, only one-dimensional rule-based category structures could be learned while participants were making same-different category judgments. This learning was confirmed by the high accuracy of participants and by model fits suggesting that participants' behavior was consistent with the use of optimal one-dimensional strategies. It is well documented that one-dimensional categorization rules are usually easy to learn (for a review, see Ashby & Maddox, 2005). Even so, within the context of the same-different literature the success of participants in the one-dimensional conditions is somewhat surprising because previous research has shown that indirect category-learning tasks often focus participants on similarity relationships and away from rules (Brooks et al., 2007; Minda & Ross, 2004). If only similarity had been considered, the one-dimensional category structures and information-integration category structures should be equally difficult, because they are rotations of one another and learning the categories indirectly should have reduced the advantage provided by the rule-based system. Also, none of the participants in the present experiments was best fit by a similarity model.

So, what is different in the present experiment? One possibility is that our experiments used a larger number of stimuli, which discouraged simple memorization strategies. Brooks et al. (2007) used a dozen different stimuli, which could have been memorized during the experiment. However, Minda and Ross (2004) used 30 different stimuli, which makes it less likely that participants would memorize the stimuli. Another possible difference is that Minda and Ross did not tell participants in the indirect learning condition that there were categories. Hence, in addition to being indirect, learning of the categories was incidental in that condition. This may have reduced the probability of participants learning the task using a hypothesis-testing strategy. Finally, it is possible that this difference is based on the categorization paradigm used. The present experiments used well-studied category structures that have been shown to rely mainly on different categorization systems (for a review, see Maddox & Ashby, 2004). In contrast, the tasks used by Brooks et al. (2007) and Minda and Ross (2004) could be learned using a number of different strategies. The category structures and instructions used in our experiments may have biased participants into using a rule-based strategy.

The effect of category structures on category representations

The results from Experiment 2 show that (1) category representations that are learned indirectly are difficult to transfer to a direct classification task and, (2) category representations learned during a direct classification task can only be transferred to an indirect category-learning task if the strategy is rule-based. These results suggest that, consistent with Brooks et al. (2007) and Minda and Ross (2004), the category representations learned during an indirect category-learning task are different from the category representations learned during a direct classification task. This is also consistent with Markman and Ross (2003), who argued for task-specific learning of category representations. The results presented here show that the specificity/generality of the knowledge learned depends on the category structures/categorization strategy used. While none of the category structures used could be transferred from an indirect task to a direct task, rule-based category structures learned directly could be transferred to an indirect category-learning task. This is similar to evidence reviewed by Markman and Ross (2003), which suggests that the representations learned during inference can transfer to a classification task (but not the other way around). However, information-integration category structures learned directly could not be transferred to an indirect category-learning task. This distinction may be dependent on the learning system used to process the task, and it would be interesting to address the issue of inference learning with information-integration category structures.

Future work

Category learning in real-world situations is often goal-driven, and there is mounting evidence that the results obtained with indirect category-learning tasks differ from those obtained in typical feedback-based classification experiments (Brooks et al., 2007; Markman & Ross, 2003; Minda & Ross, 2004). Hence, it is reasonable to question how many of the published results from classification experiments will generalize to indirect goal-driven categorization. Future work should focus on proposing new indirect category-learning paradigms to see if some regularity emerges in the results obtained with different indirect category-learning tasks. Also, the link between inference and indirect category learning should be further explored to see if the dissociations found with multiple category systems are also present with inference.

Acknowledgments

This research was supported in part by the U.S. Army Research Office through the Institute for Collaborative Biotechnologies under grant W911NF-07-1-0072 and by National Institute of Health Grant R01 MH3760-2. The authors would like to thank Kyle Liban, Alecia M. Moser, and Sarah Kaufman for their help with running the participants. The authors would also like to thank two anonymous reviewers for useful comments on an earlier version of this manuscript. Requests for reprints should be addressed to Sebastien Helie, Department of Psychological & Brain Sciences, University of California, Santa Barbara, CA 93106-9660, or using e-mail: helie@psych.ucsb.edu.

References

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review*, *105*, 442-481.

Ashby, F. G., Ell, S. W., & Waldron, E. M. (2003). Procedural learning in classification. *Memory & Cognition, 31*, 1114-1125.

Ashby, F. G. & Gott, R. E. (1988). Decision rules in the perception and categorization of multidimensional stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*, 33-53.

Ashby, F. G. & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology, 56,* 149-178.

Ashby, F. G., Maddox, W. T., & Bohil, C. J. (2002). Observational versus feedback training in rule-based and information-integration category learning. *Memory & Cognition, 30,* 666-677.

Bamber, D. (1969). Reaction times and error rates for “same”-“different” judgments of multidimensional stimuli. *Perception & Psychophysics*, *6*, 169-174.

Brainard, D. H. (1997). Psychophysics software for use with MATLAB. *Spatial Vision, 10 ,* 443-436.

Brooks, L. R., Squire-Graydon, R., & Wood, T. J. (2007). Diversion of attention in everyday concept learning: Identification in the service of use. *Memory & Cognition*, *35*, 1-14.

Helie, S., Waldschmidt, J.G., & Ashby, F.G. (2010). Automaticity in rule-based and information-integration categorization. *Attention, Perception, & Psychophysics, 72*, 1013-1031.

Krueger, L. E. (1978). A theory of perceptual matching. *Psychological Review*, *85*, 278-304.

Maddox, W. T., & Ashby, F. G. (1993). Comparing decision bound and exemplar models of categorization. *Perception & Psychophysics*, *53*, 49-70.

Maddox, W. T., & Ashby, F. G. (2004). Dissociating explicit and procedural-learning based systems of perceptual category learning. *Behavioural Processes*, *66*, 309-332.

Maddox, W. T., Ashby, F. G., & Bohil, C. J. (2003). Delayed feedback effects on rule-based and information-integration category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *29*, 650-662.

Maddox, W. T., & Ing, A. D. (2005). Delayed feedback disrupts the procedural-learning system but not the hypothesis-testing system in perceptual category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 100–107.

Markman, A. B., & Ross, B. H. (2003). Category use and category learning. *Psychological Bulletin*, *129*, 592-613.

Miller, E. K., & Buschman, T. J. (2008). Rules through recursion: How interactions between the frontal cortex and basal ganglia may build abstract, complex rules from concrete, simple ones. In S. A. Bunge, & J. D. Wallis (Eds.) *Neuroscience of Rule-Guided Behavior* (pp. 419-440). New York: Oxford University Press.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167-202.

Minda, J. P., & Ross, B. H. (2004). Learning categories by making predictions: An investigation of indirect category learning. *Memory & Cognition*, *32*, 1355-1368.

Minda, J. P., & Miles, S. J. (2010). The influence of verbal and nonverbal processing on category learning. *The Psychology of Learning and Motivation*, *52*, 117-162.

Thomas, R. D. (1996). Separability and independence of dimensions within the Same-Different judgment task. *Journal of Mathematical Psychology*, *40*, 318-341.

Waldron, E. M., & Ashby, F. G. (2001). The effects of concurrent task interference on category learning: Evidence for multiple category learning systems. *Psychonomic Bulletin & Review*, *8*, 168-176.

Table 1. Percentage of optimal best-fitting models in Experiment 1

	First 100 trials	Last 100 trials
1D-Width	75	95*
1D-Orientation	80	85
Information-integration	5	0

Note. The optimal strategy in the 1D-Width condition is one-dimensional on the horizontal axis, the optimal strategy in the 1D-Orientation condition is one-dimensional on the vertical axis, and the optimal strategy in the information-integration condition is the general linear classifier. * $p < .05$

Table 2. Mean accuracy in Experiment 2

Training task	Block 1	Block 14	Block 15	Block 15-14	Transfer gain
Cat – 1D	0.852	0.924	0.824 (0.859)	-0.100*	-0.057
Cat – II	0.655	0.759	0.604 (0.634)	-0.155***	0.040
Cat – Conj.	0.639	0.824	0.691 (0.710)	-0.133***	0.145***
SD – 1D	0.881	0.920	0.884 (0.960)	-0.036	0.032
SD – II	0.564	0.532	0.614 (0.766)	0.060	-0.041
SD – Conj.	0.545	0.568	0.652 (0.784)	0.084	0.013

Note. Cat = Classification; SD = Same-different categorization; 1D = 1D-Width; II = information-integration; Conj. = Conjunction. Block 1 is the first block of training, Block 14 is the last block of training, and Block 15 is the first block of transfer. Numbers in parentheses are predictions made by an optimal decision model (see main text). Block 15-14 is a within-subject design. Transfer gain is a between design comparing the first transfer block of each condition with the first training block of the matching condition. For instance, the first transfer block of Cat – 1D was identical to the first training block of SD – 1D. Hence, the transfer gain for Cat – 1D is the difference between these performances (i.e., 0.824 – 0.881 = -0.057). * p < .05; *** p < .001.

Table 3. Percentage of optimal best-fitting models in Experiment 2

	Training		Transfer		
Training task	First 100 trials	Last 100 trials	First 100 trials	Last 100 trials	Strategy transfer
Cat – 1D	86.7	80	73.3	80	66.7 (60)
Cat – Conj.	40	66.7*	60	40	60 (53.3)
Cat – II	20	33.3	0	0	26.7 (0)
SD – 1D	86.7	100	80	86.7	80 (73.3)
SD – Conj.	6.7	0	40	73.3**	26.7 (0)
SD – II	0	0	13.3	40*	33.3 (0)

Note. The optimal strategy in the 1D-Width condition is one-dimensional on the horizontal axis, the optimal strategy in the conjunction condition is the intersection of separate one-dimensional rules on the horizontal and vertical axes, and the optimal strategy in the information-integration condition is the general linear classifier. Strategy transfer represents the percentage of participants who were best fit by the same model in the last 100 trials of training and the first 100 trials of transfer. Numbers in parentheses represent the percentage of optimal strategy transfer. Binomial test of proportions were separately performed for each condition to assess model learning during training and transfer. * $p < .05$; ** $p < .01$.

Figure captions

Figure 1. Category structures used in the experiments. The top panels and the bottom-right panel are rule-based conditions (top-left = 1D-Width; top-right = 1D-Orientation; bottom-right = Conjunction). The bottom-left panel is an information-integration category structure. For sine-wave gratings, the horizontal axis (x_1) in each panel represents the bars width and the vertical axis (x_2) in each panel represents the bars orientation. The optimal bound in the top-left panel is $x_1 = 50$. The optimal bound in the top-right panel is $x_2 = 50$. The optimal bound in the bottom-left panel is $x_2 = x_1$. The optimal bounds in the bottom-right panel are $x_1 = 40$ and $x_2 = 60$.

Figure 2. Mean proportion correct per block in Experiment 1. The error bars are standard errors.

Figure 3. Mean proportion correct per block in Experiment 2.

Figure 1

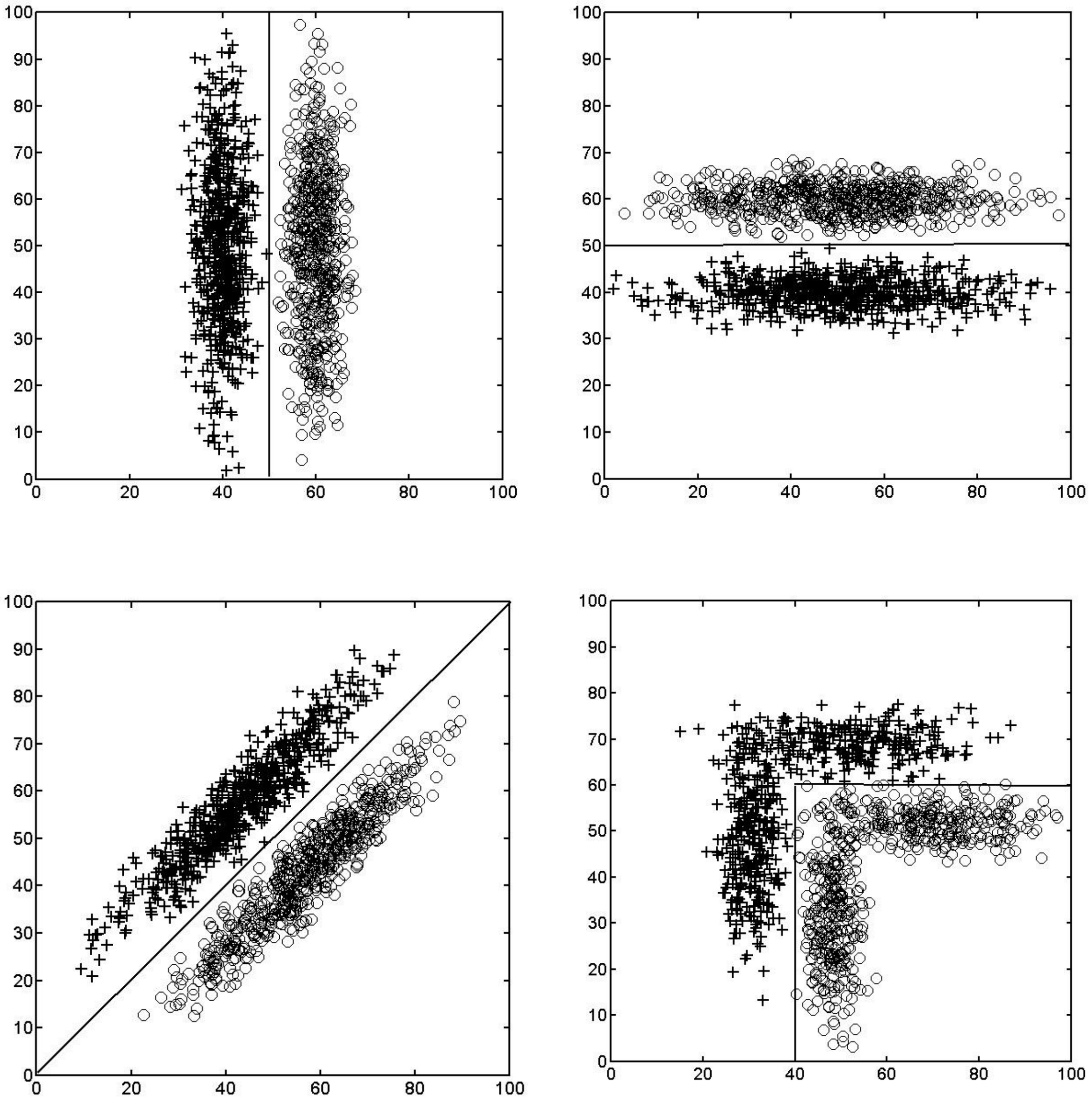
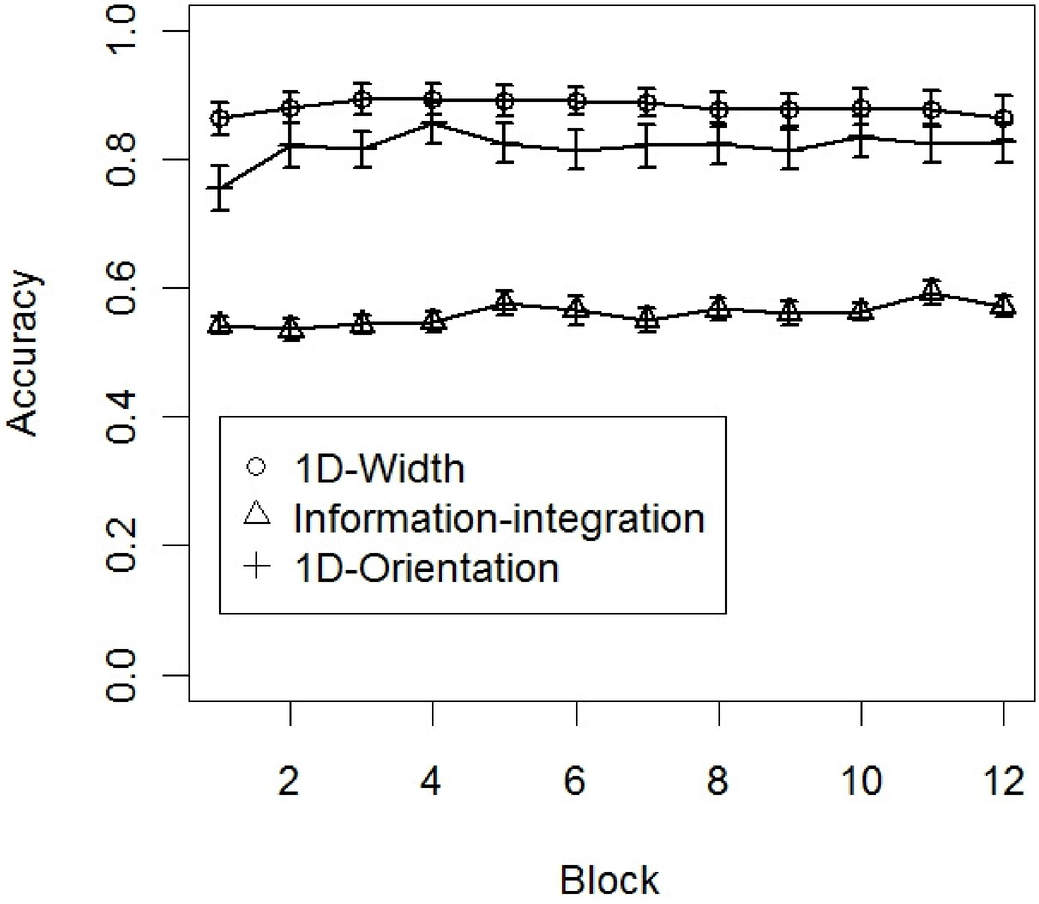
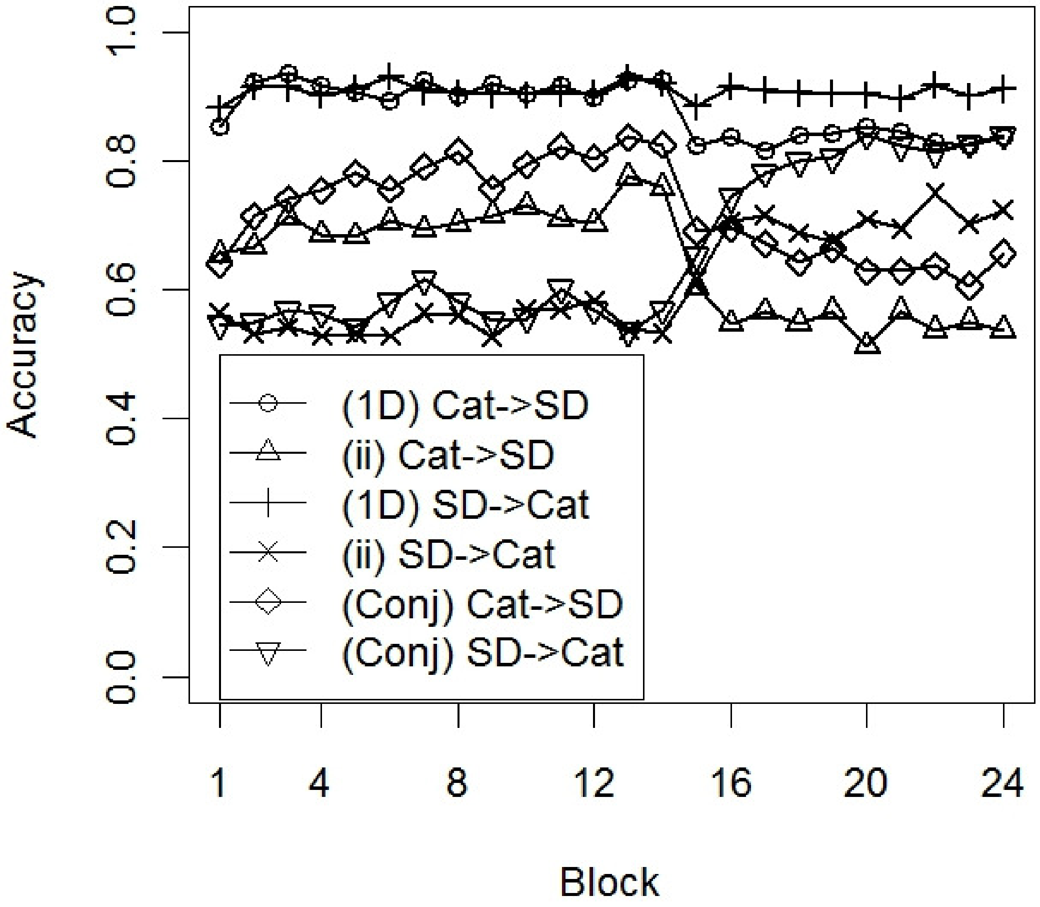


Figure 2


Accuracy
1.0
0.8
0.6
0.4
0.2
0.0
Block
2
4
6
8
10
12
○ 1D-Width
△ Information-integration
+ 1D-Orientation


Accuracy
0.0
0.2
0.4
0.6
0.8
1.0
Block
1
4
8
12
16
20
24
○ (1D) Cat->SD
△ (ii) Cat->SD
+ (1D) SD->Cat
× (ii) SD->Cat
◇ (Conj) Cat->SD
▽ (Conj) SD->Cat