

The acquisition of automatic response biases through categorization

YVONNE LIPPA

*Max Planck Institute for Psychological Research, Munich, Germany
and Indiana University, Bloomington, Indiana*

and

ROBERT L. GOLDSTONE

Indiana University, Bloomington, Indiana

Participants categorized left- and right-pointing line drawings presented together with pictures, whose pointing direction (to the left or right) was ambiguous (spatially ambiguous pictures; Experiments 1 and 2) or that pointed neither to the left nor to the right (spatially neutral pictures; Experiment 3). Subsequently, the spatially ambiguous and neutral pictures were used in a Simon task, wherein participants made left and right keypresses on the basis of the color of the pictures, while ignoring the object that they depicted. In all three experiments, performance was facilitated when the response required by the color matched the pointing direction of the line drawings with which the picture had been previously paired. Performance was impeded when the response required by the color did not match the pointing direction of the line drawings with which the picture had been previously paired. Thus, responses indirectly associated with pictures by category membership were automatically triggered even when the responses were inappropriate.

The present research explores the finding that a stimulus acquires the ability to elicit a response R simply because it has been grouped together with other stimuli that elicit response R. This phenomenon has been demonstrated in human learning studies (Astley & Wasserman, 1996; Dougher, Augustson, Markham, Greenway, & Wulfert, 1994; Eisman, 1955; Grice, Henriksen, & Speiss, 1972; Jeffrey, 1953; Malloy & Ellis, 1970; Wasserman & DeVolder, 1993), as well as in animal learning studies (Astley & Wasserman, 1998, 1999; Urcuioli, Zentall, & DeMarse, 1995; Urcuioli, Zentall, Jackson-Smith, & Steirn, 1989, Experiment 2; Wasserman, DeVolder, & Coppage, 1992; Zentall, Steirn, Sherburne, & Urcuioli, 1991). For the remainder of the paper, we will refer to these effects as response biases that are acquired through categorization.

Using a typical three-step procedure, one of the earlier demonstrations of the acquisition of response biases through categorization was provided by Jeffrey (1953). In the first step, the *categorization phase*, he trained preschoolers to classify the colors white, gray, and black into two categories by giving the verbal responses "black"

and "white." Some participants assigned white and gray to one response category and black to the other category, whereas other participants assigned black and gray to one and white to the other category. In the second, the *shaping phase*, the color gray was omitted, and the participants learned to make PUSH and PULL responses to the colors white and black, while still labeling them. In the final step, the *test phase*, the color gray was reintroduced, and the participants continued to make the PUSH and PULL responses of the shaping phase. Compared with a baseline ascertained prior to learning, the PUSH and PULL responses to the color gray were a function of category membership. If gray was originally grouped with white, it elicited the response assigned to white, and if it was grouped with black, it elicited the response assigned to black. Hence, despite the fact that the color gray was never assigned to a PUSH or PULL response, it elicited these responses because it inherited the response biases from its category members.

The purpose of the present research was to determine whether response biases that are acquired through categorization exert control over future behavior automatically or whether they are a result of one's learning experiences being intentionally applied to a new situation. In other words, does a learning experience exert influence because it is difficult to suppress or because it is explicitly used for strategic reasons? Explanatory accounts can be formulated for either notion. An *automatic response bias account* follows the framework of associative network models (e.g., Rescorla & Wagner, 1972; Rumelhart & McClelland, 1986) and attributes the influence of categorization to mediated

This research was supported in part by National Institutes of Health Grant MH56871 to R.G. We thank John Kruschke for making lab space and facilities available, Matthew R. Littell and Timothy J. Mahoney for running participants, and John Flowers, Bruce Milliken, Steven Soman, David Waller, and an anonymous reviewer for helpful comments on a previous version of the article. Correspondence concerning this article should be addressed to Y. Lippa, Department of Psychology, University of California, Santa Barbara, CA 93106-9660 (e-mail: lippa@psych.ucsb.edu).

or secondary generalization, a mechanism that was introduced by Hull (1939). In the categorization task, an association chain is established between the category members, their categorizing response, and the perceivable effects that follow the categorizing response. In the example of Jeffrey's (1953) study, the colors white and gray become associated with the verbal response "white" and its perceivable effects—for example, the sound of the word *white*. In the shaping phase, one category member (e.g., *white*) is associated with a new response (e.g., PULLING). This forms the link between *white* and PULLING, and because white evokes the sound of the word *white*, also the link between the sound of the word *white* and PULLING. The latter mediating link is the reason why other category members inherit the response bias from the color white. If in the test task, gray is presented, the associated response WHITE is activated as is its corresponding effect "sound of white," which, in turn, activates PULLING, thus creating the bias. In sum, according to an automatic response bias account, category learning establishes associative links between stimuli and responses that subsequently exert control over response selection processes. The formation of stimulus–response associations is subject to the person's voluntary behavior. The response bias resulting from the associations, however, is mandatory because, once formed, the associations transfer information automatically.

According to a strategic response bias account, behavior is biased in a test task because previously learned rules are applied once again (Brewer, 1974; Holyoak, Koh, & Nisbett, 1989). Such an account may also utilize the framework of an associative network, but instead of transferring information automatically, associative links are used to explicitly retrieve information from memory. For instance, after categorizing the three colors and then associating a PUSH and PULL response with the colors white and black, Jeffrey's (1953) preschoolers were asked to perform the same PUSH–PULL response task when the color gray was presented. Although animals might treat this test task as a new, independent task, humans do not. They recognize that the present stimuli and responses were part of the previous tasks, and they may even suspect that transfer between the tasks is being tested. Thus, in order to decide whether to PUSH or PULL when gray is presented, it is a reasonable strategy to recall that gray was previously grouped together with white and that, therefore, it should be treated as being white. The response in the test task is biased because it reflects what has been experienced before. Yet, this bias is strategic in nature because the person can choose not to take advantage of the previous learning experience but to apply a different strategy.

One way to provide evidence in favor of either the automatic response bias account or the strategic response bias account is to employ a test task in which performance is unlikely to be influenced by strategic behavior. If such a task provides no evidence for response biases, we can conclude that the strategic component is a necessary variable in biased behavior. If, in turn, response biases do occur,

there is evidence that performance can be determined by previous experience automatically.

The experiments mentioned above, which investigated the acquisition of response biases through categorization, are all ambiguous in regard to the test task requirements. This holds in particular for human learning studies. Jeffrey's (1953) preschoolers, for example, were instructed to continue categorizing the stimuli (naming the colors) in the test phase, while they produced a pushing or pulling response. Thus, there was a strong task demand to strategically use the labels. In the test task employed by Malloy and Ellis (1970), participants were exposed to all members of each category and were asked to identify the stimuli from the shaping phase by giving the name learned in the shaping phase. This procedure is quite elegant because the participants' attention was drawn to what had happened in the shaping phase and, thus, was pulled away from what happened in the categorization phase. However, the stimuli were variations of a prototype, and, therefore, it is not clear whether the obtained transfer effects resulted from perceptual similarities (primary generalization; Hull, 1939) or from categorization (secondary generalization; Hull, 1939). The most convincing evidence for the acquisition of automatic response through categorization comes from studies using conditioned responses. For example, after having participants categorize abstract figures, Dougher et al. (1994, Experiment 1) paired one category member with an electric shock. A subsequent test task using skin conductance as a measure demonstrated that the conditioning transferred to the other category members. Because a change in galvanic skin response is less subject to voluntary control, this finding provides some evidence that response biases acquired through categorization occur automatically. However, galvanic skin responses following electric shocks, or eye blink responses following air puffs (Grice et al., 1972), can be argued to be exceptional responses because of their role in avoiding pain and threat. They form a distinct class of responses, and it is still open to question to what extent the findings can be generalized to other more arbitrary responses.

Another shortcoming that holds for all of the studies mentioned above is that the stimulus aspect that acquired the ability to elicit a response was task relevant in the test phase. That is, the task that must be solved in the test phase required humans and animals to consider stimulus aspects that had been relevant beforehand. Thus, even if category learning forms associations between the critical stimulus and the response, evidence for response biases does not necessarily indicate that these links exert control automatically. After all, the organism is supposed to respond to the critical stimulus and might therefore intentionally use previously relevant associations, an argument in favor of the strategic response bias account.

A test task that overcomes these problems and that is a better test of whether a stimulus elicits a response automatically is one in which the critical stimulus is task irrelevant. Participants are exposed to the critical stimulus,

but the problem they must solve involves a second stimulus or stimulus attribute. That is, all task demands and strategies refer to the second stimulus aspect, so that if the critical stimulus aspect turns out to influence performance, it can only be attributed to mandatory processing. A task that meets these requirements is the *Simon task* (Simon & Rudell, 1967). This is a choice-reaction task, in which the response varies on one dimension (e.g., position: left vs. right) and the stimuli vary on two dimensions, one that is task relevant (e.g., color: red vs. blue) and one that is task irrelevant (e.g., position: left vs. right). The typical finding, termed the Simon effect, is that performance is facilitated and impeded when the irrelevant stimulus attribute matches and mismatches, respectively, the attribute of the required response.

In the present study, we used the Simon task as our test task. Specifically, in three experiments, participants completed first a categorization task, in which line drawings that exhibited spatially ambiguous features (Experiments 1 and 2) or no spatial features (Experiment 3) acquired the ability to elicit left and right responses, and then participants were transferred to a Simon task. We were interested in whether the preceding learning experience would generate Simon effects, indicating the automaticity of response biases acquired through categorization.

EXPERIMENT 1

Experiment 1 provided a first test as to whether response biases that are acquired through categorization are automatic in nature. We shortened the three-step procedure of classic secondary generalization studies by skipping the shaping phase. Instead of associating a new response with some of the category members, we used stimuli that already possessed the ability to elicit particular responses—namely, line drawings of left- and right-pointing arrows and fingers. In a categorization task, the left- and right-pointing stimuli were grouped together with stimuli that were expected to acquire response biases, specifically, line drawings of an eagle and a French horn.

In a subsequent Simon task, the participants gave left and right responses to the eagle and French horn, with shape of the stimuli being the irrelevant stimulus attribute, and color being the relevant stimulus attribute. If categorization causes stimuli to inherit the response biases of the other members of their category, and if this bias is automatic in nature, performance in the Simon task should be a function of the respective picture–color combination. Performance should be facilitated if the response required by the color matches the pointing direction of the stimuli with which the picture was previously paired (compatible condition). Performance should be impeded if the response required by the color does not match the pointing direction of the stimuli with which the picture was grouped together (incompatible condition). To ensure that the present Simon task produced Simon effects, the left- and right-pointing fingers and arrows also served as stimuli. Here,

trials were considered compatible (C) and incompatible (IC) if there was a correspondence and noncorrespondence, respectively, between the response required by the color and the pointing direction of the stimulus.

Besides the picture–color combination, we varied the time between the onset of the picture and the onset of the color. We manipulated stimulus onset asynchronies (SOAs) in order to account for the potentially dynamic nature of response biases. According to the automatic response bias account, these effects occur because responses are automatically activated through associative links. This activation is a time-consuming process, especially if not only one, but several associative links are involved. Thus, it could be that the pictorial information must precede the color information in order for it to gain influence. On the other hand, it has been shown that activation that is irrelevant to the task decays over time (Hommel, 1994). A particularly large SOA might therefore produce no response bias at all because facilitating or interfering information has already decayed. Because the particular processing times of our stimuli and colors were uncertain and, in addition, participants might differ in processing the stimulus information, we chose three different SOAs. Each originally black picture turned red or blue after 0, 100, or 300 msec. This range should allow us to account for the potential dynamic of response biases, as well as for individual differences.

Method

Participants

Thirty-two undergraduate students (16 females and 16 males) from Indiana University served as participants in order to fulfill a course requirement and were run in parallel sessions. Their ages ranged from 18 to 23. They were right handed by self-report and had normal or corrected-to-normal vision.

Apparatus and Stimuli

Stimulus presentation and data acquisition were controlled by the software package ERTS (Experimental Run Time System; Beringer, 1998) and Dell Dimension XPS P133s computers. Line drawings from the Snodgrass and Vanderwart (1980) picture corpus served as stimuli: left- and right-pointing versions of the finger and the arrow, the eagle, and the French horn (see Figure 1). The stimuli were presented on a white background on Sony Multiscan 15SFII monitors with a resolution of 800 × 600 pixels in black, red, or blue (Hue-Saturation-Value model; 0°, 0%, and 0% for black; 0°, 100%, and 100% for red; 240°, 100%, and 100% for blue), depending on the task and condition. The pictures differed in width and height, but all fit in a 6.0 × 5.3 cm fixation frame. The viewing distance was approximately 50 cm.

The left- and right-pointing pictures will be referred to as spatially biased stimuli because they already possessed the ability to elicit left and right responses. The eagle and the horn will be referred to as spatially ambiguous stimuli because they are not symmetric and exhibit spatially ambiguous attributes. When one focuses on the eagle's body and the horn's end, they can be considered to be pointing to the left, and when one focuses on the eagle's beak and the horn's mouthpiece, they can be considered to be pointing to the right.

Procedure

Within a 1-h session, the participants consecutively completed a categorization task and a Simon task.

Categorization task. The participants learned to assign two left-pointing stimuli and one spatially ambiguous picture to one category and two right-pointing stimuli and the other spatially ambiguous picture to a second category. Whether the eagle or the horn was paired with the left- or right-pointing stimuli was counterbalanced across participants. The stimuli appeared in the center of the computer screen. The participants indicated category membership by moving the mouse cursor on a top or bottom button (3.2 cm wide and 1.6 cm high, shown on the screen throughout the experiment) and then clicking the left mouse button. The top and bottom buttons were located 4.3 cm above and below the center of the screen and were always labeled Category 1 and Category 2, respectively. The category that contained the left-pointing stimuli was always assigned to the top button, and the category containing the right-pointing stimuli was always assigned to the bottom button. This up-left/down-right assignment is counter to the up-right/down-left preference reported in the literature (Weeks & Proctor, 1990), so that any effects we might find would not be attributed to a preexisting stimulus-response mapping preference. Training continued until each participant had completed 592 trials correctly (148 replications for the eagle and for the horn and 74 replications for each of the arrows and fingers). Trials with an incorrect response, or with a latency to click the mouse button equal to or less than 120 msec (mean anticipations: 0) or greater than 10 sec (mean overtime responses: 1) resulted in the stimulus being turned off and a brief beep. Those trials were repeated at a random position in a block. After completing the 592 learning trials, the participants could end the training session by fulfilling the criterion of 12 consecutively correct trials.

Simon task. The participants performed a choice reaction task, with the shape of the pictures being the irrelevant attribute and the color of the pictures being the relevant stimulus attribute. The task was to press as quickly and accurately as possible the left and right shift keys in response to the colors red and blue, irrespective of which picture was shown. The color-to-key mapping was counterbalanced across participants.

Each trial started by presenting a fixation frame (6.0×5.3 cm) in the center of the screen, which remained visible throughout the trial. After 500 msec, one randomly chosen picture was displayed until the participant responded or 1,000 msec had passed (missing). To take time differences in processing the shape and color information into account, we used three SOAs between the onset of the picture and its color: The picture could appear colored from the very beginning (SOA of 0 msec) or be presented in black and then turn red or blue 100 or 300 msec after its initial appearance. The eagle and the horn were each presented 16 times in red and 16 times in blue for each of the three SOA conditions, while each of the two arrows and fingers was presented 8 times. Overall, 384 trials were completed. Auditory feedback was given again on incorrect, missing (2.8%), and anticipation trials (latency equal to or less than 120 msec: 0.5%), which were then repeated at a random position later in the block. After the experiment, the participants were asked to report the direction in which they perceived the eagle and horn to be pointing.

Results

All participants completed the 592 correct trials in the categorization task. The mean reaction time (RT) for the group was 1,195 msec. On average, 25 errors were made. The learning criterion was fulfilled without any errors within 12 trials with a mean RT of 1,076 msec.

To determine performance in the Simon task, RTs and errors were calculated separately for the two stimulus sets (spatially biased stimuli vs. spatially ambiguous stimuli), the three SOA conditions, and the two compatibility conditions (compatible vs. incompatible). Separate analyses of variance (ANOVAs) were conducted on RTs and errors

for spatially biased and ambiguous stimuli, using the within-subjects variables of condition (compatible and incompatible) and SOA (0, 100, and 300 msec).

Spatially Biased Stimuli

There was an effect of SOA. The longer the delay between the picture and color onset, the faster [$F(2,62) = 44.93$, $MS_e = 870.37$, $p < .001$] and more error-prone [$F(2,62) = 4.33$, $MS_e = 15.64$, $p < .05$] the response. In addition, we obtained a Simon effect. Performance in the compatible condition (454 msec and 2.5% errors) was better than in the incompatible condition (490 msec and 6.2% errors) [$F(1,31) = 132.40$, $MS_e = 444.12$, $p < .001$ and $F(1,31) = 27.60$, $MS_e = 28.74$, $p < .001$ for RTs and errors, respectively]. This finding ensured the functionality of the task. The condition \times SOA interaction did not approach significance for RT or error data ($p > .20$ and $p > .10$, respectively).

To check whether the Simon effect was influenced by practice, we calculated the same ANOVA as described above with RTs and errors broken down in two halves (first half of correct trials in each condition vs. second half of correct trials in each condition). Performance in the first half is depicted in the left panel of Figure 1. The compatibility effect observed in RTs was the same in the first and second halves ($p > .15$). For errors, we found that the compatibility effect was slightly larger in the first half (4.4%) than in the second half (2.9%) [$F(1,31) = 2.99$, $MS_e = 20.28$, $p = .094$]. This is probably because the participants made more errors in the first half (5.0%) than in the second half (3.7%) [$F(1,31) = 4.63$, $MS_e = 35.73$, $p < .05$].

Spatially Ambiguous Stimuli

Overall performance varied again with SOA. The participants responded faster [$F(2,62) = 62.71$, $MS_e = 615.77$, $p < .001$] and made more errors [$F(2,62) = 3.08$, $MS_e = 11.71$, $p = .053$] the longer the delay between picture and color onset. The Simon effect failed to reach significance in error data (3.2% vs. 3.6% for C and IC, respectively; $F < 1$) but was significant in RTs [$F(1,31) = 4.50$, $MS_e = 393.17$, $p < .05$]. The participants responded faster in the compatible (459 msec) than in the incompatible condition (465 msec). As revealed by an ANOVA using the additional factor half (first half of correct trials in each condition vs. second half of correct trials in each condition), this compatibility effect occurred only in the first half (a 10-msec difference between C and IC) and not in the second half (1-msec difference) [$F(1,31) = 5.38$, $MS_e = 381.15$, $p < .05$]. Performance in the first half is depicted in the right panel of Figure 1. The Simon effect in the first half did not depend on whether, in the categorization task, the eagle (horn) was assigned to left- or right-pointing stimuli ($F < 1$). The absence of the Simon effect in the error data was true for the first as well as the second half ($F < 1$).

The judgments concerning the subjective pointing direction of the eagle and the horn were as follows: 7 participants gave no valid answer (4 participants in the condition where the eagle and horn were grouped with left- and right-pointing stimuli, respectively, and 3 with the

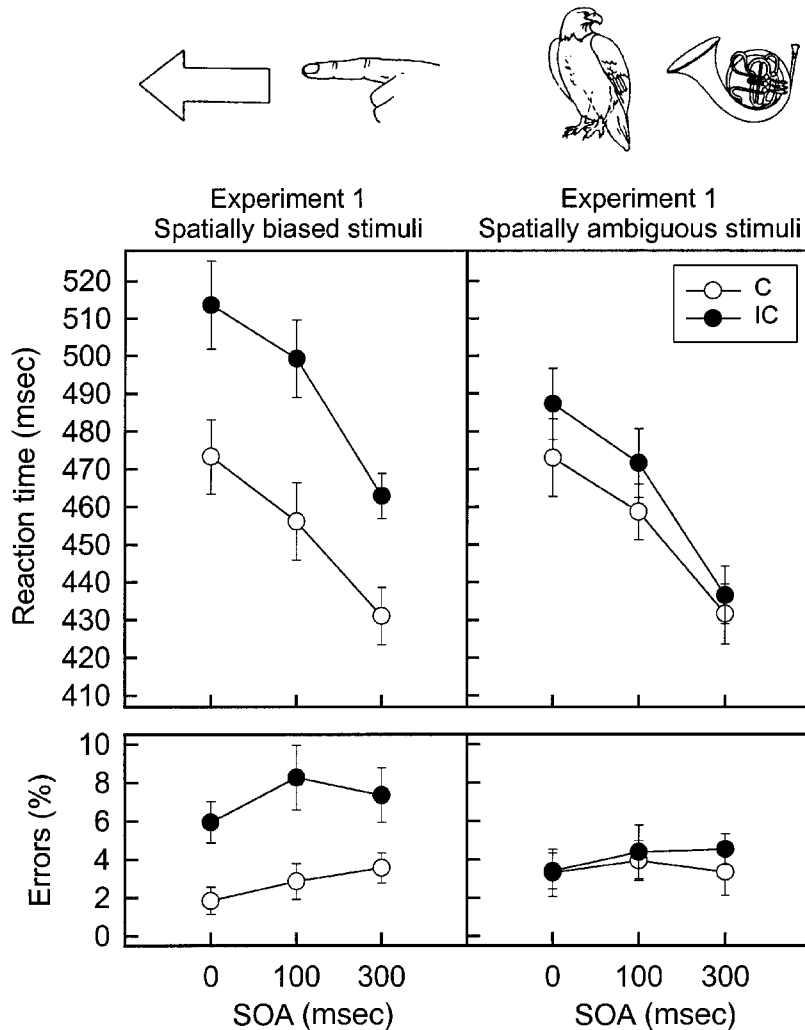


Figure 1. The Simon effect in the first half of trials for spatially biased (left panel) and spatially ambiguous (right panel) stimuli in Experiment 1. Reaction times and errors are presented as a function of condition (C = compatible; IC = incompatible) and SOA. The bars indicate standard errors of the means.

reversed mapping). Of the 25 participants that gave valid answers, 4 gave only part answers, describing only the eagle or only the horn. Thus, instead of 50 observations, there were only 46. If the spatially ambiguous pictures were paired with left-pointing stimuli, a left-pointing direction was indicated 16 times and a right-pointing direction 6 times. If the spatially ambiguous pictures were paired with right-pointing stimuli, a right pointing direction was indicated 17 times and a left-pointing direction 7 times. This is a significantly greater proportion than would be expected by chance [$\chi^2(1,46) = 8.712, p < .001$]. Thus, for the majority of participants, the pointing judgment for the spatially ambiguous stimuli was a function of category membership.

Discussion

In addition to a strong Simon effect for spatially biased stimuli, Experiment 1 provided two results. First, the cate-

gorization task turned out to be effective in creating response biases for spatially ambiguous stimuli. According to the verbal judgments at the end of the experiment, the majority of the participants indicated that the eagle and horn pointed to the right when they shared a category with right-pointing stimuli and that they pointed to the left when they shared a category with left-pointing stimuli. Of course, these judgments might have been influenced by strategies because the participants may have judged the stimuli in accordance with what they experienced in the categorization task. Hence, the explicit judgments provided evidence for the effectiveness of categorization in evoking response biases, but they did not address the question of whether the response biases occurred automatically.

More conclusive, however, was the second finding of Experiment 2, showing that spatially ambiguous pictures elicited a left or right response even when the shape of the

picture was irrelevant to the task at hand. The automatic response bias obtained in the Simon task occurred only in RTs and vanished over time. In the first half of trials, it amounted to 10 msec. This is comparable to the automatic response bias effect that occurs if the response is directly associated with a picture.¹ However, the lack of an effect in the second half of trials suggests that automatic response biases are prone to rapid extinction. From an associative network point of view, one may argue that the associations between the stimuli and the categorizing responses weakened rapidly or were rather transient from the beginning. In either case, information transfer becomes impaired.

The automatic response bias of spatially ambiguous pictures was not modulated by the amount of time that had elapsed between the onset of the picture and the onset of the color. In part, this may be due to the fact that with increasing SOAs, the participants responded faster and made more errors; for instance, a lower RT level might have minimized the difference between the compatible and the incompatible conditions and thus counteracted an increase of effect size due to the increase in SOAs. The lack of an SOA effect may in part also be due to individual differences. It could be that the participants differed too much in processing the stimulus information to identify a single optimum delay between the picture and the color onset.

EXPERIMENT 2

Experiment 1 provided evidence for the automatic nature of response biases that are acquired through categorization. However, since the Simon effect obtained for spatially ambiguous stimuli was relatively small and persisted only temporarily, Experiment 2 was carried out in order to replicate the effect.

We made several changes in the experimental design. First, we raised the number of categorization trials from 604 to 900 trials, which should strengthen the associations between the stimuli and the categorizing responses and thus increase the effect size. Second, instead of completing the categorization task first and then the Simon task, the participants alternated between both tasks. This manipulation should also increase the effect size by preventing the associations from decaying. In addition, it allowed us to monitor the development of the response biases because we obtained performance in the Simon task several times in the course of learning. Third, because SOA did not affect the Simon effect in Experiment 1, we employed only one SOA condition. We chose the SOA of 100 msec because it yielded numerically the largest effect. A fourth change involved the exclusion of the spatially biased stimuli from the Simon task. We wanted to ensure that the Simon effect for spatially ambiguous stimuli does not depend on the intermixing of all category members in the test task. In addition, we doubled the number of participants in order to increase the power to detect the presence or absence of a Simon effect for spatially ambiguous stimuli.

Method

Participants and Data Preparation

Sixty-seven undergraduate students from Indiana University served as participants in order to fulfill a course requirement and were run in parallel sessions. Their data were cleaned of trials that had an RT shorter than or equal to 120 msec or longer than 1,000 msec. This was true for 4% of the data.

Apparatus, Stimuli, and Procedure

Experiment 2 was a replication of Experiment 1 with the following exceptions. Stimulus presentation and data acquisition were controlled by Power Mac 7300 Apple computers. The stimuli were presented on Apple monitors with resolutions of 832×624 pixels. The pictures fit in a 15×15 cm area and were viewed from approximately 45 cm. The colors in the Simon task were red and green (Hue-Saturation-Value model for green: 120° , 100%, and 100%).

Within a 2-h session, the participants completed three blocks, each consisting of a categorization task (each 300 trials; randomly drawn from the six experimental stimuli) and a subsequent Simon task (each 80 trials; 20 for each of the four possible picture-color combinations). Since in Experiment 1, the compatibility effect occurred independently of assignment between the spatially biased and the spatially ambiguous stimuli in the *categorization task*, in Experiment 2, we employed only one assignment: The eagle was always paired with right-pointing stimuli, and the French horn with left-pointing stimuli. The assignment between the category and the top and bottom button on the screen was randomized. After the participants made their response, a check or an X appeared on the screen (depending on whether they were right or wrong, respectively) for 500 msec, and then the entire screen was erased. After an intertrial interval (ITI) of 500 msec, the next categorization trial started.

In the *Simon task*, the participants responded to the colors red and green by pressing the "q" key with their left index fingers and the "p" key with their right index fingers. The color-key assignment was randomized across participants. A trial started with the presentation of a picture in black that turned red or green after 96 msec. The colored picture remained on the screen for 500 msec after the participants made their responses. In addition, a check or an X appeared on the screen for 500 msec. The screen then went white, and after an ITI of 500 msec a new trial started.

The participants were given rest periods every 50 trials in the categorization task and every 40 trials in the Simon task. During these rest periods, the participants were told their average RT and percent correct for the preceding block of trials.

Results

All participants achieved more than 95% accuracy by the last categorization block. ANOVAs were conducted on RTs and errors, using as the within-subjects variables compatibility condition (compatible and incompatible) and block (testing after the first, second, or third block). As is shown in Figure 2 (left panel), there was an effect of practice. RTs were shorter in the second (469 msec) and third (466 msec) blocks than in the first block (498 msec) [$F(2, 132) = 14.67$, $MS_e = 3,784.00$, $p < .001$]. In addition, we obtained a highly significant Simon effect [$F(1, 66) = 29.87$, $MS_e = 564.35$, $p < .001$]. The participants responded faster on compatible (471 msec) than on incompatible (484 msec) trials. The size of the Simon effect varied slightly with block [$F(2, 132) = 2.59$, $MS_e = 332.40$, $p = .079$]. The 7-msec effect in the first block was marginally significant [$F(1, 66) = 3.36$, $MS_e = 531.54$, $p = .071$] whereas the 17-msec effect in the second block [$F(1, 66) =$

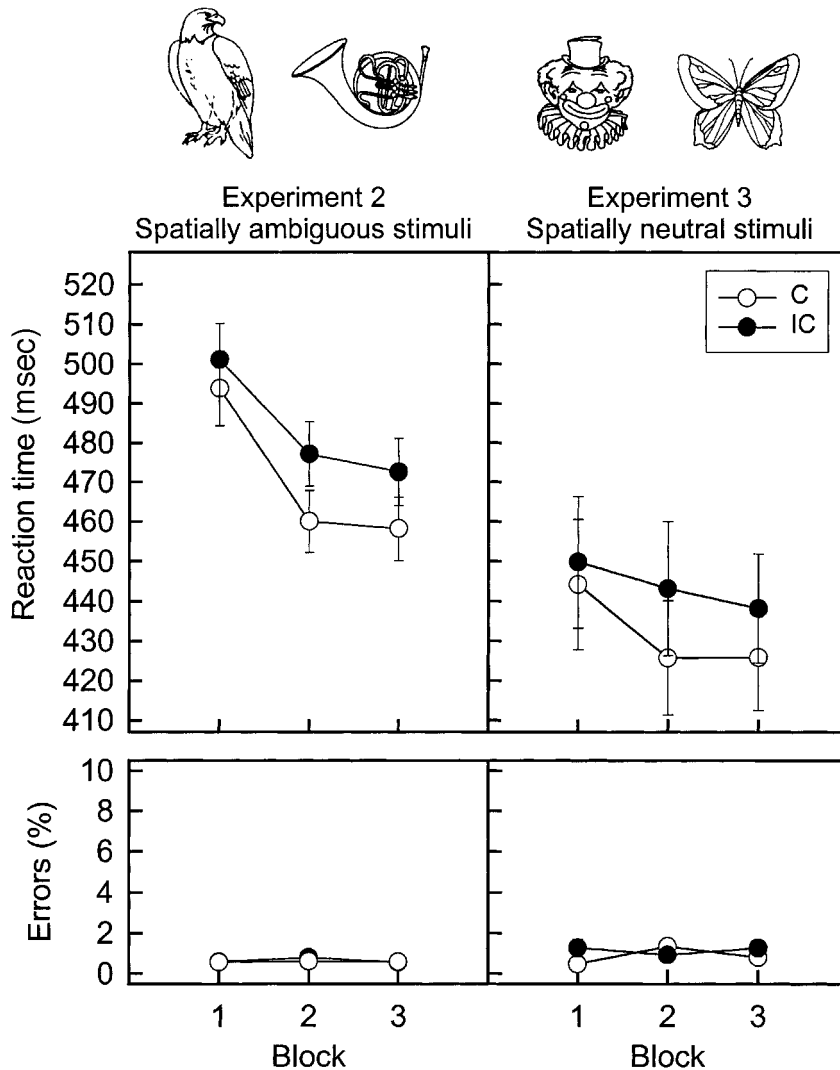


Figure 2. The Simon effect for spatially ambiguous stimuli in Experiment 2 (left panel) and for spatially neutral stimuli in Experiment 3 (right panel). Reaction times and errors are presented as a function of compatibility condition (C = compatible; IC = incompatible) and blocks of trials. The bars indicate standard errors of the means.

23.82, $MS_e = 412.25, p < .001$] and the 15-msec effect in the third block [$F(1,66) = 24.43, MS_e = 285.36, p < .001$] were highly significant. The effect in the first block differed from the effect in the second block [$F(1,66) = 4.19, MS_e = 769.92, p < .05$] but not from the effect in the third block [$F(1,66) = 2.54, MS_e = 669.31, p = .116$]. The effect sizes of the second and third blocks were virtually the same ($F < 1$). There were no effects in the error data ($F < 1$), probably because the participants made on average less than 1% errors.

Discussion

The major finding of Experiment 1 was replicated in Experiment 2: A Simon effect was present for spatially ambiguous stimuli. This indicates that these stimuli inherited

the ability to automatically trigger a left or right response from their spatially biased category members. Although Experiment 1 yielded a 10-msec effect for only the first half of trials (96 trials), in Experiment 2, we obtained a 13-msec effect across all trials (240 trials). Thus, interspersing the Simon task with categorization trials as well as increasing the number of categorization trials helped to produce more robust response biases.

Experiment 2 further revealed that it takes a certain amount of categorization experience to establish the response biases. Although after the first 300 categorization trials, the Simon effect amounted to only 7 msec, the effect was more than doubled after the second 300 categorization trials (17 msec). From an associative network point of view, this may reflect the strengthening of the as-

sociations between the stimuli and the categorizing responses with repeated experience. The size of the Simon effect remained about the same after the last 300 categorization trials. Thus, under the condition of limited learning experience provided in Experiment 2, the maximum size of a Simon effect might be about 15 msec.

EXPERIMENT 3

Experiments 1 and 2 provided evidence for the acquisition of automatic response biases through categorization for spatially ambiguous stimuli. The features of those stimuli could indicate a left as well as a right response (e.g., the eagle's beak pointing to the right, and the eagle's belly pointing to the left). Membership in a category of left-pointing or right-pointing stimuli resolved this ambiguity and created unidirectional, automatic response biases.

Experiment 3 tested whether the acquisition of automatic response biases through categorization could also be shown for stimuli that exhibit no features corresponding to a left or right response. Therefore, we replicated Experiment 2 by using spatially neutral stimuli—a clown and a butterfly (see Figure 2, right panel)—instead of the spatially ambiguous stimuli (i.e., the eagle and the French horn of the previous experiments). Because the former stimuli are symmetric along their vertical axis, they should be spatially neutral with respect to a left or right response. If the mere fact of sharing the same category creates automatic response biases, Experiment 3 should yield a Simon effect, too.

Method

Participants and Data Preparation

Twenty-two undergraduate students from Indiana University served as participants in order to fulfill a course requirement and were run in parallel sessions. The data were cleaned of trials that had an RT shorter than or equal to 120 msec or longer than 1,000 msec. This was true for 2.8% of the data.

Apparatus, Stimuli, and Procedure

Experiment 3 was a replication of Experiment 2 with the following exceptions. The butterfly and the clown (see Figure 2, right panel) from the Snodgrass and Vanderwart (1980) picture corpus served as the stimuli that were predicted to inherit the response biases from the left- and right-pointing arrows and fingers. We will refer to the butterfly and the clown as spatially neutral stimuli. The clown shared a category with the right-pointing fingers and arrows, and the butterfly shared a category with the left-pointing fingers and arrows. In each of the three categorization blocks, each of the six pictures was presented 50 times.

Results

All participants achieved more than 95% accuracy by the last categorization block. The data are depicted in Figure 2 (right panel). ANOVAs were conducted on error and RT data, using as the within-subjects variables compatibility condition (compatible and incompatible) and block (testing after the first, second, or third block). Performance did not vary with block ($p > .20$) but did with condition. The participants responded faster on compat-

ible (432 msec) than on incompatible (444 msec) trials, [$F(1,21) = 6.54$, $MS_e = 703.20$, $p < .05$]. As in Experiment 2, the Simon effect was smaller in the first block (6 msec) and larger in the second (17 msec) and third (12 msec) blocks. However, the interaction between condition and block was not reliable [$F(2,42) = 1.13$, $MS_e = 397.36$, $p = .326$]. For error data, the only significant effect was the interaction between block and condition [$F(2,42) = 4.12$, $MS_e = 1.42$, $p < .05$]. In the first block, we obtained a small Simon effect: Performance in the compatible condition was slightly less error prone than performance in the incompatible condition (a difference of 0.8%) [$F(1,21) = 3.87$, $MS_e = 1.83$, $p = .062$]. In the second and third blocks, however, performance in the compatible and incompatible conditions did not differ statistically [$F(1,21) = 1.29$, $MS_e = 1.56$, $p = .269$ for the second block, a difference of 0.4%, and $F(1,21) = 2.55$, $MS_e = 0.93$, $p = .125$ for the third block, a difference of 0.5%].

Discussion

Experiment 3 yielded a Simon effect for spatially neutral stimuli. Like the eagle and French horn in the previous experiments, the clown and butterfly inherited the ability from their spatially biased category members to automatically elicit a left and right response. The overall effect amounted to 12 msec and is thus comparable to the effect we obtained in Experiment 2. The findings of Experiment 3 thus show that even stimuli that exhibit no spatial cues cease to be spatially neutral and automatically trigger spatial responses when they are categorized with spatially biased stimuli.

GENERAL DISCUSSION

In the present study, we were interested in whether or not response biases that are acquired through categorization exert control over behavior in a subsequent test task automatically. We approached this question by testing whether line drawings, after they were categorized with pictures of left- and right-pointing fingers and arrows, elicit a left and right response in a Simon task wherein the shape of the line drawings was the task-irrelevant stimulus attribute and color was the task-relevant stimulus attribute. In Experiment 1, we found a reliable but quickly decaying Simon effect for line drawings that were spatially ambiguous (i.e., stimuli that originally indicated a left as well as a right response). Experiment 2 replicated Experiment 1 using an increased number of categorization trials and interspersing the Simon task with the categorization trials. This yielded a more robust Simon effect for spatially ambiguous stimuli. Experiment 3 replicated the procedure of Experiment 2 and produced a Simon effect for spatially neutral stimuli (i.e., stimuli that indicated originally neither a left nor a right response).

The robust presence of a Simon effect across the three experiments shows that response biases that were acquired through categorization needed to be suppressed in order not to perform the incorrect response. We conclude from

this evidence that categorization can create automatic response biases that influence behavior without having strategic intent. Naturally, this does not imply that response biases that are acquired through categorization are always automatic and never strategic in nature. If the task allows strategies to be applied, strategies probably remain a critical variable in determining behavior. Our study indicates that when it is unlikely that behavior is governed by strategies, response biases can still become effective.

Consistent with previous studies in animal learning (Astley & Wasserman, 1998, 1999; Wasserman et al., 1992), we have shown that for human participants the acquisition of response biases through categorization does not depend on the perceptual features of the experimental stimuli. In Experiment 3, the clown and the butterfly did not share spatial features with their category members or with the responses. Nevertheless, they ceased to be neutral after categorization and elicited left and right responses automatically. Future research might investigate whether the categorization effect is independent of characteristics of the categorizing response, as well. In the present experiments, both the categorizing responses as well as the acquired responses were spatial in nature. It would be interesting to know whether a verbal, nonspatial response in the categorization task would yield the same automatic response biases.

In the introduction, we discussed Hull's (1939) secondary generalization idea as a possible mechanism for the influence of categorization. According to this idea, newly acquired characteristics are passed from one category member to another because such characteristics are attached to the category and not to a single stimulus. Specifically, if a stimulus is categorized together with other stimuli and thereafter acquires a new property, this does not only affect the attributes of the stimulus but those of the category as well. Therefore, any other object that shares with the stimulus the same category will be influenced by the new experience.

As pointed out by Urcuioli and Lionello-DeNolf (2001), Hull's (1939) analysis implies that the occurrence of categorization-induced generalization depends on the order of events. If the acquisition of a new attribute *follows* the categorization task—and this is true for all of the previously cited studies—the analysis mentioned above applies, and generalization should occur. However, if the acquisition of a new attribute *precedes* the categorization task—and, logically, this is true for the three experiments we reported—no generalization should occur. If a stimulus first acquires a new attribute (e.g., the “leftness” to a scribble that we call a left-pointing arrow), this feature is uniquely attached to the stimulus and not to a category, because categorization (e.g., assigning the picture of an eagle and a left-pointing arrow to a top response) has not yet taken place. That is, the vehicle of generalization, the common category, is yet to be formed and cannot mediate the newly acquired property. This suggests that our results—as well as the results of other studies showing that

swapping the categorization and shaping task effectively induces generalization (Urcuioli & Lionello-DeNolf, 2001)—cannot be accounted for by Hull's secondary generalization mechanism. To explain these findings a different mechanism is needed.

An alternative mechanism of categorization-induced generalization has been discussed in animal learning research (Urcuioli & Lionello-DeNolf, 2001; Wasserman & DeVolder, 1993; Zentall, 1998). This mechanism holds that newly acquired characteristics are passed from one category member to another because category members cue each other and, thus, each other's properties. That is, irrespective of when stimuli are categorized in the chain of events, the common category will mediate between the category members and generate effects of generalization.² According to this explanation, the present experiments produced automatic response biases not because the eagle itself evoked the left response but because it prompted its category members, namely, left-pointing arrows and fingers, which in turn elicited a left response. For the time being, this is a speculative idea for a mechanism, and future research is needed to determine its validity.

In sum, the present study showed that response biases that are acquired through categorization are automatic in nature. This supplements previous findings that have shown that response biases that are acquired through direct stimulus–response mapping are automatic in nature (MacLeod & Dunbar, 1988; Proctor & Lu, 1999; see also note 1). More generally, the present research addresses the degree to which higher mental processes are deliberate (cf. Bargh & Ferguson, 2000; Hommel, 2000). By showing that (presumably) intentionally executed actions can be regulated by automatic processes, the present study suggests that willful behavior could be determined.

REFERENCES

- ASTLEY, S. L., & WASSERMAN, E. A. (1996). Mediating associations, essentialism, and nonsimilarity-based categorization. In T. R. Zentall & P. M. Smeets (Eds.), *Stimulus class formation in humans and animals* (pp. 111-133). Amsterdam: Elsevier.
- ASTLEY, S. L., & WASSERMAN, E. A. (1998). Novelty and functional equivalence in superordinate categorization by pigeons. *Animal Learning & Behavior*, *26*, 125-138.
- ASTLEY, S. L., & WASSERMAN, E. A. (1999). Superordinate category formation in pigeons: Association with a common delay or probability of food reinforcement makes perceptually dissimilar stimuli functionally equivalent. *Journal of Experimental Psychology: Animal Behavior Processes*, *25*, 415-432.
- BARGH, J. A., & FERGUSON, M. J. (2000). Beyond behaviorism: On the automaticity of higher mental processes. *Psychological Bulletin*, *126*, 925-945.
- BERINGER, J. (1998). *Experimental Run Time System* [Computer program]. Frankfurt: BeriSoft Corporation.
- BREWER, W. F. (1974). There is no convincing evidence for operant or classical conditioning in adult humans. In W. B. Weimer & D. S. Palermo (Eds.), *Cognition and the symbolic processes* (pp. 1-42). Hillsdale, NJ: Erlbaum.
- DOUGHER, M. J., AUGUSTSON, E., MARKHAM, M. R., GREENWAY, D. E., & WULFERT, E. (1994). The transfer of respondent eliciting and extinction functions through stimulus equivalence classes. *Journal of the Experimental Analysis of Behavior*, *62*, 331-351.
- EISMAN, B. S. (1955). Attitude formation: The development of a color-

- preference response through mediated generalization. *Journal of Abnormal Psychology*, **50**, 321-326.
- GRICE, G. R., HENRIKSEN, K., & SPEISS, J. M. (1972). Response mediated generalization in eyelid conditioning with reduced conflicting information. *Journal of Experimental Psychology*, **92**, 398-404.
- HALL, G. (1996). Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning. *Animal Learning & Behavior*, **24**, 233-255.
- HOLYOAK, K. J., KOH, K., & NISBETT, R. E. (1989). A theory of conditioning: Inductive learning with rule-based default hierarchies. *Psychological Review*, **96**, 315-340.
- HOMMEL, B. (1994). Spontaneous decay of response code activation. *Psychological Research/Psychologische Forschung*, **56**, 261-268.
- HOMMEL, B. (2000). The prepared reflex: Automaticity and control in stimulus-response translation. In S. Monsell & J. Driver (Eds.), *Attention and performance XVIII: Control of cognitive processes* (pp. 247-273). Cambridge, MA: MIT Press.
- HULL, C. L. (1939). The problem of stimulus equivalence in behavior theory. *Psychological Review*, **46**, 9-30.
- JEFFREY, W. E. (1953). The effects of verbal and nonverbal responses in mediating an instrumental act. *Journal of Experimental Psychology*, **45**, 327-333.
- MACLEOD, C. M., & DUNBAR, K. (1988). Training and Stroop-like interference: Evidence for a continuum of automaticity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **14**, 126-135.
- MALLOY, T. E., & ELLIS, H. C. (1970). Attention and cue-producing responses in response-mediated stimulus generalization. *Journal of Experimental Psychology*, **83**, 191-200.
- PROCTOR, R. W., & LU, C. H. (1999). Processing irrelevant location information: Practice and transfer effects in choice-reaction tasks. *Memory & Cognition*, **27**, 63-77.
- RESCORLA, R. A., & WAGNER, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and non-reinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: Current research and theory* (pp. 64-99). New York: Appleton-Century-Crofts.
- RUMELHART, D. E., & MCCLELLAND, J. L. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 1). Cambridge, MA: MIT Press.
- SIMON, J. R., & RUDELL, A. P. (1967). Auditory S-R compatibility: The effect of an irrelevant cue on information processing. *Journal of Applied Psychology*, **51**, 300-304.
- SNODGRASS, J. G., & VANDERWART, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory*, **6**, 174-215.
- URCUIOLI, P. J., & LIONELLO-DE NOLF, K. (2001). Some tests of the anticipatory mediated generalization model of acquired sample equivalence in pigeons' many-to-one matching. *Animal Learning & Behavior*, **29**, 265-280.
- URCUIOLI, P. J., ZENTALL, T. R., & DEMARSE, T. (1995). Transfer to derived sample-comparison relations by pigeons following many-to-one versus one-to-many matching with identical training relations. *Quarterly Journal of Experimental Psychology*, **48B**, 158-178.
- URCUIOLI, P. J., ZENTALL, T. R., JACKSON-SMITH, P., & STEIRN, J. N. (1989). Evidence for common coding in many-to-one-matching: Retention, intertrial interference, and transfer. *Journal of Experimental Psychology: Animal Behavior Processes*, **15**, 264-273.
- WASSERMAN, E. A., & DEVOLDER, C. L. (1993). Similarity and non-similarity-based conceptualization in children and pigeons. *Psychological Record*, **43**, 779-793.
- WASSERMAN, E. A., DEVOLDER, C. L., & COPPAGE, D. J. (1992). Non-similarity-based conceptualization in pigeons via secondary or mediated generalization. *Psychological Science*, **3**, 374-379.
- WEEKS, D. J., & PROCTOR, R. W. (1990). Salient-features coding in the translation between orthogonal stimulus and response dimensions. *Journal of Experimental Psychology: General*, **119**, 355-366.
- ZENTALL, T. R. (1998). Symbolic representation in animals: Emergent stimulus relations in conditional discrimination learning. *Animal Learning & Behavior*, **26**, 363-377.
- ZENTALL, T. R., STEIRN, J. N., SHERBURNE, L. M., & URCUIOLI, P. J. (1991). Common coding in pigeons assessed through partial versus total reversals of many-to-one conditional and simple discriminations. *Journal of Experimental Psychology: Animal Learning Processes*, **17**, 194-201.

NOTES

1. In a pilot study, we asked 24 participants to associate pictures that were symmetric along the vertical axis (e.g., a butterfly or a clown) with left and right responses. Two pictures went with a left and two other pictures went with a right keypress. After 225 correct assignments between each picture and its response, the participants performed a Simon task, with the shape of the picture being the irrelevant attribute and the color of the picture being the relevant attribute (cf. Experiment 1). A picture-color combination was considered compatible if the color required the same response that was assigned to the picture in the picture-response mapping task. If the color required the opposite response, it was considered an incompatible trial. The participants responded 10 msec faster and made 1.7% fewer errors in the compatible condition than in the incompatible condition [$t(23) = 2.77, p < .05$, two-tailed, and $t(23) = 2.70, p < .05$, two-tailed, for RTs and errors, respectively].
2. A mechanism that underlies this form of mediation would necessitate bidirectional associations between the category members and the categorizing response. Because in the present experiment, stimulus presentation and response execution in the categorization task overlapped, this could have been the case (cf. Hall, 1996).

(Manuscript received February 1, 2001;
revision accepted for publication May 24, 2001.)