

Learning and transfer of response–effect relations



Quarterly Journal of Experimental Psychology
1–18

© Experimental Psychology Society 2024



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/17470218241284259

qjep.sagepub.com



Markus Janczyk¹, Lea Eichfelder¹, Heinrich R Liesefeld¹
and Volker H Franz²

Abstract

Acting means changing the environment according to one's own goals, and this often requires bodily movements as responses. How these responses are selected is a central question in contemporary cognitive psychology. The ideomotor principle offers a simple answer based on two assumptions: An agent first learns an association between a response and its effects. Later, this association can be used in a reverse way: When the agent wants to achieve a desired effect and activates its representation, the associated response representation becomes activated as well. This reversed use of the learned association is considered the means to select the required response. In three experiments, we addressed two questions related to the first assumption: First, we tested whether effect representations generalise to more abstract conceptual knowledge. This is important, because outside the laboratory and in novel situations, effects are variable and not always exactly identical, such that generalisation is necessary for successful actions. Second, the nature of the response–effect relation has been debated recently, and more data are necessary to put theorising on firm empirical ground. Results of our experiments suggest that (a) abstraction to conceptual knowledge seems to occur only under very restricted situations, and (b) it seems that no (implicit) associations between responses and effects are learned, but rather (explicit) propositional knowledge in the form of rules.

Keywords

Action; ideomotor principle; learning; associations; propositional knowledge

Received: 23 January 2024; revised: 11 August 2024; accepted: 13 August 2024

Motor behaviour is an essential part of human life. In particular, moving the body is involved in acting, that is, in achieving goals or intended states. Actions might be a (rather) simple touch of a button to switch on a light, but also a slight push on a steering device to bring a landing aeroplane into the final position for touch-down. How motor movements in such actions are selected, addressed, and executed is a central question of cognitive psychology and an elegant answer is provided by the *ideomotor principle*: Actions are selected by anticipating their goal states, that is, those changes in the environment, the respective motor movement would bring about.

The present article focuses on one particular aspect of this idea, namely the prerequisite of having learned the respective relation. Thus, the mechanism we are dealing with is learning the relation between a motor and a (perceptual) effect representation, or in short, learning a response–effect (R-E) relation. We first ask whether any generalisation and/or abstraction of the involved effect

representation to conceptual knowledge occurs. Such representations are of particular interest as they might be re-used when an agent is confronted with a new situation. Second, our data also inform a recent discussion on what is actually learned in R-E learning experiments, that is, on the representational nature of the learned relation.

The next section will introduce the ideomotor principle and the respective experimental evidence in more detail, including the recently emerged debate about the representational nature of the learned relation (see Kaup et al., 2024, for a recent discussion of this issue across several fields of

¹Department of Psychology, University of Bremen, Bremen, Germany

²Department of Computer Science, University of Tübingen, Tübingen, Germany

Corresponding author:

Markus Janczyk, Department of Psychology, University of Bremen, Hochschulring 18, Bremen 28359, Germany.

Email: janczyk@uni-bremen.de

psychological research). The following section then discusses the available evidence regarding effect generalisation. While it seems clear from the literature that stimuli do generalise to some extent, the matter is less clear for effects as investigated here. Finally, we provide an overview of the present experiments and the corresponding predictions.

The ideomotor principle and RE learning

General idea of the ideomotor principle dates back to philosophical writings of the 19th century (e.g., Carpenter, 1852; Harleß, 1861; Herbart, 1825; see Stock & Stock, 2004, and Pfister & Janczyk, 2012, for more information on the history of those ideas). Having largely been dismissed from the agenda of psychological research (see Thorndike, 1913), ideomotor principle-inspired research has experienced a resurgence of interest since a publication of Greenwald (1970a) and its incorporation into the Theory of Event Coding (TEC; Hommel et al., 2001; for reviews, see Badets et al., 2016; Janczyk et al., 2023; Shin et al., 2010). Put simply, the core assumptions of the ideomotor principle are the following: (a) As a result of repeated co-occurrence of responses and their ensuing effects, humans first learn their relations in terms of bidirectional R-E associations. (2) Later, the response cannot be activated directly. Rather, anticipating one's goal, that is, retrieving and activating the effect representation, automatically co-activates the associated motor representation of the response. We focus here on the first assumption and it is particularly important to note that the bidirectionality is indeed a core assumption of the ideomotor principle. For example, Elsner and Hommel (2001) stated that "given the temporal overlap of the activation of the motor and the sensory pattern, the corresponding codes are integrated (i.e., linked with each other) so that activating one pattern on a later occasion will lead to activating the other one, too" (p. 230). They continue with noting that this "does not depend on attention being focused on the response-effect (R-E) relationship or an explicit intention to learn about it" (p. 230). Hence, although certainly debatable, this assumption is central for the reported experiments.¹

Systematic investigations of R-E learning heavily relied on an approach introduced by Greenwald (1970b) and later popularised by Elsner and Hommel (2001). In these experiments, participants first undergo an *acquisition phase* during which they repeatedly experience their responses to result in particular effects. In the Elsner and Hommel experiments, for example, participants provided about 200 left or right key press responses in a free-choice task (i.e., they were to freely choose which key they wanted to press on each trial; Berlyne, 1957; Janczyk et al., 2020; Naefgen et al., 2018).² Crucially, each key press predictably resulted in a low- or high-pitch tone as its effect (e.g., left key → low-pitch and right key → high-pitch).

The presumably established R-E associations are then probed in two versions of a subsequent test phase in which

the previous effects are now presented as stimuli. In the first version, a *forced-choice test phase*, participants are instructed to respond in a specific and instructed way. For example, one group is supposed to respond in an acquisition-congruent manner (low-pitch → left key and high-pitch → right key), whereas another group is supposed to respond in an acquisition-incongruent manner (high-pitch → left key and low-pitch → right key). Under the assumption that experiencing the former effect automatically activates the associated response, a response conflict is expected in the incongruent mapping thus lengthening response times (RTs). This result has been reported several times (e.g., Eder & Dignath, 2017; Elsner & Hommel, 2001, Exp. 1; Watson et al., 2015). In the second version, a *free-choice test phase*, participants are supposed to freely choose between pressing the left or right key upon presentation of one of the two stimuli (sometimes no-go trials are intermixed, indicated by an additional stimulus that was not presented in the acquisition phase). The assumption here is that experiencing a former effect gives its associated response a "head-start" and thus (acquisition-)congruent choices should occur more often than it would be expected if participants merely choose randomly. A response bias, that is, a percentage of (acquisition-)congruent choices larger than 50% has also been reported several times (e.g., Elsner & Hommel, 2001, Exp. 2–4; Janczyk et al., 2023, Exp. 3; Pfister et al., 2011). This observation was also made under dual-task conditions (see Exp. 4 in Elsner & Hommel, 2001) and it was accordingly argued that the "response bias is automatic" (Elsner & Hommel, 2001, p. 238).

In sum, the assumption of learned R-E associations has received empirical support. Yet, and this will be important for the present experiments, using a free-choice test phase has recently been criticised and the evidence gathered by this approach has been questioned (Custers, 2023; Sun et al., 2020). More precisely, it has been argued that the free-choice task invites deliberate reasoning about how to respond and the observed response bias might have other causes than an automatic activation of congruent responses. To this end, Sun et al. analysed the distribution of the percentage of congruent choices across participants in their Experiment 1 and observed bimodality: Most participants seemed to respond more or less randomly (resulting in a pronounced first peak at around 50%), while some participants provided mostly congruent responses (resulting in a second, less pronounced, peak at around 90%–100%). A similar result has been reported for the control (category) group in the work by Eichfelder et al. (2023). Thus, the typically reported response bias might be the result of averaging across participants who show qualitatively different types of behaviour.

One interpretation, advanced by Sun et al. (2020) and Custers (2023), is that participants do not learn R-E associations proper, but rather extract propositional knowledge in the form of rules that some, but not all, participants later

exploit strategically during the test phase, thus producing the peak at 90% to 100%. In the following, we will thus use the term *R-E relation* to remain neutral with regard to the nature of the learned knowledge: propositional knowledge versus associative knowledge.

Generalisation and abstraction in R-E learning

Conceivably, effects resulting from responses in natural behaviour outside the laboratory are not always exactly identical, vary on some dimension(s), might only be (perceptually) similar to earlier effects, or be exemplars of a broader category. Thus, to deal with new situations where the exact effect representation is not yet learned, it seems helpful if learned R-E relations extend to other, related effect representations as well. While *stimulus generalisation* appears as an established phenomenon (going back to early work of Guttman & Kalish, 1956), the empirical evidence for *effect generalisation* is—at best—mixed.

Hommel et al. (2003) reported the first series of three experiments addressing this question. In Experiment 1, two groups of participants underwent a typical acquisition phase. The category group produced the visually presented category words “furniture” and “animal” (originally in Spanish language) as action effects. In contrast, the exemplar group produced the visually presented words “chair” and “dog” (also originally in Spanish language), corresponding to exemplars of the categories used for the category group. The test phase was forced-choice and used only the category words as stimuli for both groups. Hence, the exemplar group is the particularly interesting experimental group in which generalisation from exemplars to categories is tested (while the category group serves as a control group). RTs were shorter with an acquisition-congruent mapping compared with an incongruent mapping and this RT difference was similar for both groups. Thus, it seems that perceiving the exemplar words as effects includes an effect representation beyond the actually seen exemplar word, a generalised representation that includes the corresponding superordinate categories (see Rosch et al., 1976). Later encountering these categories as stimuli similarly activates the respective response. Experiment 2 of the Hommel et al. study extended this observation to transfers to other category members, because such transfer would be another indication for generalisation of the learned links. That is, participants learned exemplar words like “chair” and “dog” during the acquisition phase and were then tested with other exemplars like “table” and “cat.” Finally, Experiment 3 tested whether transfer of R-E relations can also be mediated by mere perceptual features. Here, participants produced the words “orange” and “blackboard” during the acquisition phase and were then tested with the words “circle” and “rectangle.” Both

experiments revealed a congruency effect. Yet, no control group was run in these experiments and hence no comparison of the congruency effects’ size in relation to the congruency effect without generalisation was possible.

A recent study by Eichfelder et al. (2023) aimed at a conceptual replication of Experiment 1 reported by Hommel et al. (2003). In contrast to the forced-choice test phase, this study used a free-choice test phase though. Against the background of ideomotor literature, a free-choice test phase should—if anything—be more sensitive to demonstrate the existence of R-E relations as it has been argued that only an action mode which is induced by free-choice, but not by forced-choice tasks, guarantees expression of this knowledge (e.g., Pfister et al., 2011). Nevertheless, Eichfelder et al. observed the typical response bias only for the category group, but not for the exemplar group. In other words, this study did not replicate the transfer from exemplars to categories reported by Hommel et al. (2003). On top of this, a bimodal distribution of the percentage of congruent choices (see Sun et al., 2020, Exp. 1) was observed for the category group. For the exemplar group, the distribution was unimodal and centred around 50%.

In another study (Esser et al., 2023), participants first learned to relate (four) pictures of objects to four different horizontally aligned responses. On each trial, one of these responses was prompted by a spatially compatible stimulus. In particular, one of four horizontally aligned rectangles was highlighted and if, for example, the left-most rectangle was highlighted, the left-most response key was to be pressed (i.e., a forced-choice task instead of a free-choice task was used in the acquisition phase). In Experiment 1, pictures of a violin, a banana, a pig, and a dress were used during the acquisition phase and each picture resulted predictably from pressing one particular response key. In each trial of the test phase, participants were first presented with a cue picture according to the experimental manipulation described below. Afterward, they were presented with the same stimulus prompting a response as was already used in the acquisition phase. Two manipulations were implemented. First, the picture could be congruent to the response (and thus to the learned picture) required on this trial or not (these conditions are labelled “learned” and “unlearned response location” in the original study). Assume, for example, that the left-most key resulted in the presentation of the violin during the acquisition phase. Then, presenting the violin followed by the stimulus that requires to press the left-most key constitutes a congruent test phase trial, while presenting any other stimulus demanding a different response constitutes an incongruent trial. Second, the nature of the pictures was varied (within-participants). Participants received the learned pictures, pictures of exemplars from the same category (i.e., pictures of a piano, a pineapple, a donkey, and a short), or dissimilar, unrelated pictures belonging to

another category (body parts; i.e., pictures of a foot, a hand, a nose, and an ear). RTs were (significantly) shorter in congruent than in incongruent conditions for the old pictures (~ 23 ms, $d_z = 0.50$ [calculated as $d_z = \frac{t}{\sqrt{n}} = \frac{2.67}{\sqrt{29}}$]), for the similar pictures (~ 10 ms, $d_z = 0.41$), but not for the dissimilar pictures (~ 2 ms, $d_z = 0.08$). It thus seems that experiencing pictures similar to the old (and learned) ones also induces a response activation yielding a congruency effect. While this experiment construed similarity as “belonging to the same category,” Experiment 2 used the same approach, but similarity was construed as objects that require “similar responses in terms of motor commands without the effects belonging to the same object category (e.g., a pushcart and a wheelchair)” (Esser et al., 2023, p. 2254). By and large, similar results were obtained. Thus, these results seem to support the view that effect representations do generalise from one exemplar to another exemplar of the same category, or in other words, R-E relations can become effective when being transferred to a novel-but-similar situation.

As acknowledged by the authors, however, the results are less straightforward than they seem at first glance. First, the forced-choice task might actually have induced stimulus–stimulus relations (e.g., the left-most stimulus [that instructed the left-most response] was always paired with the effect picture of the violin). Under this account, in the test phase, experiencing a stimulus unrelated to the preceding picture might have violated expectations and thereby lengthened RTs. For example, the incongruent picture might have delayed processing of the stimulus rather than preparation and/or execution of the required response that was learned during the acquisition phase. As such, the results can entirely or in parts be explained by learned stimulus–stimulus relations. Of course, the effect representations still generalised to similar ones in this case; yet, something different than R-E learning would be responsible for the observed RT effect in this case. An account in terms of stimulus–stimulus learning gains further plausibility when considering that Experiment 2 of Watson et al. (2015) suggests that R-E learning with four effects is impossible. Second, the test phase used by Esser et al. differs in a crucial aspect from the one in Eichfelder et al. (2023; and also in Hommel et al., 2003): In Esser et al., all participants re-experienced the old pictures (i.e., the learned effects), whereas in Eichfelder et al., the exemplar group encountered only new words (i.e., the respective category words) as stimuli. Recall that only the category group, which re-experienced the old category words as stimuli, showed an overall response bias.

The present experiments

The preceding two sections can be summarised in the following way: (a) Support for R-E learning has been obtained in many studies; yet, whether associations or rather rules

(propositional knowledge) are learned has been questioned recently. (b) The evidence for generalisation in R-E learning is mixed at best.

Unfortunately, the data base for both contentions is small and more data seem necessary to place further considerations and conclusions on a firmer ground. We here present three experiments aiming to provide more empirical data on these two issues. Experiments 1 and 2 built on Eichfelder et al. (2023) and use the same free-choice test procedure, but focus on different forms of generalisation (rather than generalisation from an exemplar to its superordinate category). Experiment 1 investigates whether experiencing visual effects in the top or bottom part of the computer screen as effects would abstract to conceptual representations as would result from linguistic input, that is, the corresponding words “up” and “down.” Experiment 2 investigates the same question for pictures of objects. Experiment 3 features an acquisition phase designed to increase learning strength and focuses on the bimodal distribution of the percentage of congruent choices in the test phase.

In all experiments, we first analysed and compared the response bias between a control and an experimental group, as is traditionally done. In addition, we analysed whether the percentage of congruent choices exhibits a bimodal distribution (as in Eichfelder et al., 2023, and Sun et al., 2020). Several scenarios are possible for the distribution of the percentage of congruent choices (see Figure 1). First, under the assumption of ideomotor theorists that bidirectional associations are learned automatically and thus activation of the effect representation results in co-activation of the respective response (see Footnote 1 for more details), one would expect participants to show overall a tendency of more than 50% congruent choices. This would result in a shift of the distribution towards a mean larger than 50% (see Figure 1a). This would be the situation one would expect with the standard assumptions of ideomotor theory. Second, if no associations are learned or effective in the test phase, while some participants learn and apply rule-like knowledge, we would expect a bimodal distribution with one peak around 50% (representing those participants who respond rather randomly) and another peak at a higher percentage, for example, around 90% (representing those participants who apply the rule-like knowledge and respond almost always in a congruent way; see Figure 1b). Third, both associations and application of rule-like knowledge might be at play as well. In this case, a bimodal distribution would be expected as well, though the first peak would be shifted to a higher percentage (see Figure 1c).

Experiment 1

This experiment was similar to the one by Eichfelder et al. (2023). Two groups of participants experienced different visual effects during a free-choice acquisition phase. A left

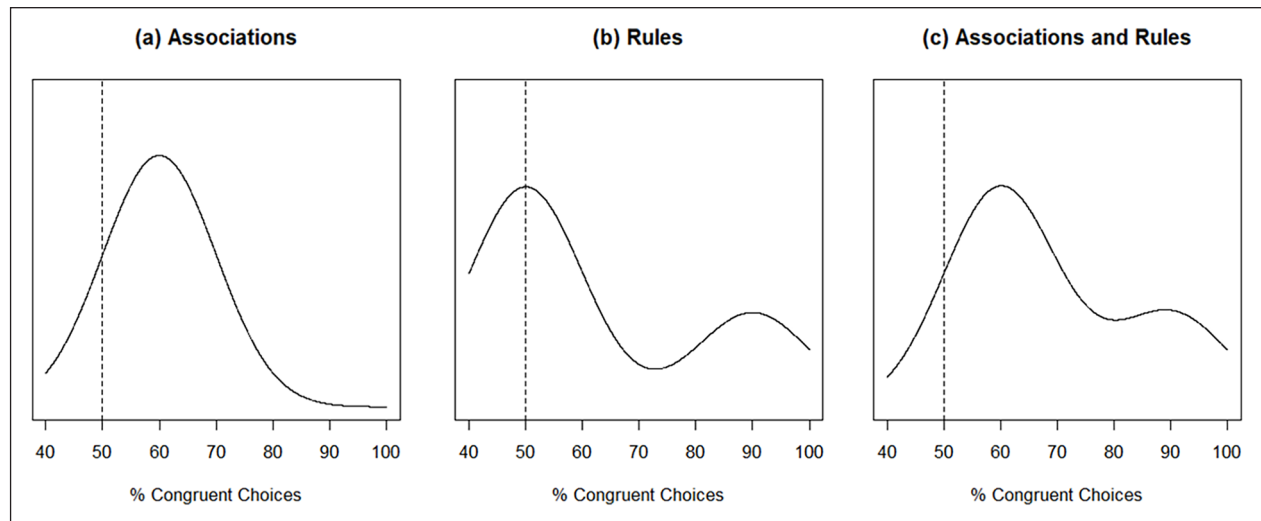


Figure 1. Illustrations of the expected distributions of the percentage of congruent choices separately for different kinds of learned knowledge being responsible for the choices.

or right key press resulted in the German words for “up” and “down” in the control group. Participants in the experimental group produced a filled white circle in the upper or lower half of the computer screen (i.e., above or below the screen centre). The test phase was a free-choice task with the words of the control group as stimuli (in addition to no-go trials that were intermixed to discourage participants from preparing their next response already before seeing the next stimulus; see Elsner & Hommel, 2001, Exp. 3). As a manipulation check, we tested for an overall response bias in the control group. If abstraction from the effect location to a conceptual representation of *up* and *down* occurs, we expect the same response bias in the experimental group. To the extent that such abstraction does not fully occur, the response bias is expected to be smaller in the experimental group, and perhaps absent if no abstraction occurs at all.

Method

Open practices statement. The preregistration of this experiment is available at <https://aspredicted.org/uy265.pdf> and the data are provided at <https://osf.io/jpvq8/>.

Data were analysed using the R software (v4.3.1; R Core Team, 2023). Bayes factors (BFs) were calculated using the R package BayesFactor and the default settings of the function `ttestBF()` (v0.9.12-4.4; Morey & Rouder, 2022). Kernel densities were used to visualise the distributions of the percentage of congruent choices and were calculated with the package `sm` (Bowman & Azzalini, 2021); bimodality coefficients were calculated with the corresponding function of the `mousetrap` package (Wulff et al., 2023). Figures were created with base R functions and the `plotrix` package (Lemon, 2006) and statistical analyses

made use of the `schoRsch` package (Pfister & Janczyk, 2016).

Sampling plan and participants. Sample size was determined by sequential sampling with BFs (Schönbrodt & Wagenmakers, 2018; Schönbrodt et al., 2017) in the same way as was done by Eichfelder et al. (2023). BFs were calculated as BF_{10} meaning that $BF > 1$ is in favour of the alternative hypothesis, and $0 < BF < 1$ is in favour of the null hypothesis. The minimum sample size was set to $n_{\min} = 40$ (i.e., 20 participants per group) and the following stopping rules applied:

1. When a $BF < 1/10$ was obtained for the percentage of congruent choices in the control group tested against 50% in a Bayesian one-sample test (i.e., the Bayesian equivalent of a t test). This result would indicate that control group participants have overall not learned the R-E relation (or do not express this relation in the test phase).
2. When the same one-sample test yielded $BF > 6$ (i.e., suggesting that learning occurred in the control group), and at the same time, a $BF > 6$ or $< 1/6$ was obtained for a Bayesian two-sample test comparing the percentages of congruent choices between the control and the experimental groups. In the first case, the extent of the response bias would differ between groups; in the second case, it would be comparable for both groups.
3. When a maximum number of $n = 50$ participants per group has been reached.

Data were obtained from 111 students of the University of Bremen who participated for course credit. Exclusion

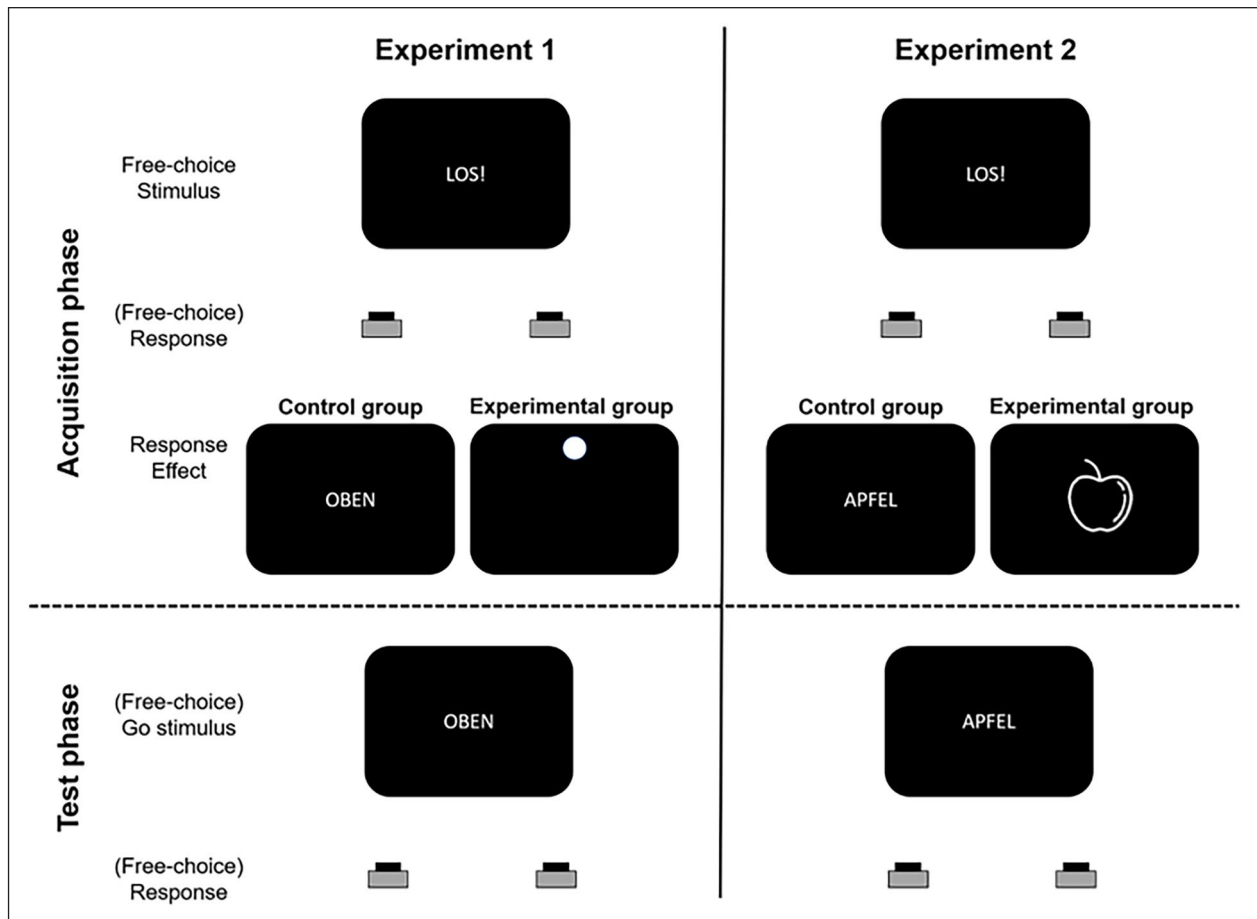


Figure 2. Illustration of the procedure and design of Experiments 1 and 2.

criteria were preregistered and followed the procedure of Eichfelder et al. (2023). Data from 11 participants were discarded according to the preregistered criteria: Ten participants did not press the left and the right key at least 80 times each during the acquisition phase and one participant responded in more than 20.0% of the no-go-trials.

The final sample size was $n=100$ participants (mean age=24.99 years; 72 women, 28 men, 0 non-binary; 88 right-handed, 11 left-handed, 1 ambidextrous). All participants reported normal or corrected-to-normal vision and to be either native German speakers or to have advanced written and spoken knowledge of German. All participants were naïve to the hypotheses of this experiment.

Stimuli and apparatus. Stimulus presentation and response collection were controlled by a standard PC connected to a 17-inch CRT-screen in a sound-attenuated and dimly lit experimental cabin at the Department of Psychology of the University of Bremen. The German words “OBEN” (Engl. “TOP”) and “UNTEN” (Engl. “BOTTOM”) and a white circle (presented in the upper or lower part of the screen) were used as effects and stimuli. In addition, the German word “LOS!” (Engl. “GO!”) was the (free-choice)

stimulus in the acquisition phase and the German word “MITTE” (Engl. “MIDDLE”) and the circle being presented in the screen centre were catch stimuli during the acquisition phase. The letter string “XXXXXX” indicated a no-go trial in the test phase. All stimuli and effects were presented in white colour against a black background. Responses were given with the left and right index finger on the “D” and the “L” key of a standard QWERTZ keyboard and the spacebar was the response key in catch trials.

Tasks and procedure. Trial procedures are illustrated in Figure 2 (left column). A trial of the *acquisition phase* began with the presentation of a white fixation cross in the screen centre for 500 ms, followed by a blank interval with a randomly determined length between 200 and 400 ms. Then, the (free-choice) stimulus “LOS!” was presented in the screen centre for 200 ms and participants were to press the left or right key as fast as possible within 1,000 ms. Each key produced a different visual effect for 500 ms. For the control group, these were the effect words “OBEN” versus “UNTEN,” and for the experimental group, the circle appearing above or below the screen centre. The

particular effects were fully contingent on the given response and R-E mappings were counterbalanced between participants.

Five percent of the acquisition phase trials were catch trials in which either the word “MIDDLE” (control group) or a central circle (experimental group) appeared instead of the regular effects. Catch trials were presented at random positions within the acquisition phase and required pressing the space bar within 2,000 ms, during which the stimulus remained on the screen. If this was not achieved, a corresponding error message (“Please react faster by pressing the SPACE BAR.” in German language) was displayed for 500 ms in the screen centre.

All trials associated with RTs longer than 1,000 ms or shorter than 100 ms were considered omissions and anticipations, respectively. In these cases, an error message appeared for 500 ms in the screen centre (“too fast!” or “too slow!” in German language) and the trial was repeated at a random position of the block so that data from 200 valid acquisition phase trials were obtained. A trial ended with an intertrial interval of 2,000 ms before the next trial started.

In the subsequent *test phase*, each trial started with a central fixation cross for 500 ms followed by a blank interval of 100 ms. Then, in 50% of the trials, the imperative (free-choice) go stimulus (i.e., one of the two words “OBEN” and “UNTEN”; both equally often) appeared for 200 ms on the screen and required a freely selected left or right response within 1,000 ms. In the other 50% of the trials, the no-go stimulus appeared in the screen centre and participants had to withhold any response for 2,000 ms. In case of errors (anticipations, omissions, responses in no-go trials), the corresponding error feedback was presented for 500 ms and the trial was repeated at a random position of the block so that data from 200 valid test phase trials were obtained. All stimuli were intermixed randomly.

Participants received written instructions prior to the acquisition and the test phase. With regard to the free-choice tasks, the written instructions were:

“Please choose yourself freely and as spontaneously as possible between pressing the left or the right response key. Try to press them about equally often while avoiding response patterns such as alternating both keys. Please avoid pressing only one response key throughout the experiment.” (the original German instructions were: “Sie sollen dann selber und so spontan wie möglich wählen, ob Sie die linke oder die rechte Taste drücken. Bitte drücken Sie beide Tasten in etwa gleich häufig und vermeiden Sie dabei Muster, wie beide Tasten abwechselnd drücken. Bitte vermeiden Sie es, immer nur die gleiche Taste zu drücken.”).

Additional verbal instructions at the beginning of the experiment essentially repeated this, after a participant read the respective instructions for the acquisition phase.

Design and analyses. Following the preregistration, we calculated one-sample Bayesian *t* tests for the control and the experimental groups separately to compare the percentage of congruent choices against a chance level of 50%. A two-sample Bayesian *t* test was calculated to compare both groups. To allow for an easier comparison with published research in the field, we also provide the corresponding frequentist *t* tests.

In addition to these preregistered analyses, we analysed the distribution of the percentage of congruent choices separately for both groups via kernel density plots to assess a possible bimodal distribution. Bimodality was quantified by calculating bimodality coefficients (SAS Institute Inc. 1990; see also Freeman & Dale, 2013; Pfister et al., 2013). Values larger than 0.55 are commonly interpreted as pointing towards a bimodal distribution. This analysis was motivated by the corresponding results reported by Sun et al. (2020) and Eichfelder et al. (2023).

Results

Acquisition phase. Anticipations, omissions, and missed catch trials occurred in 3.82%, 3.68%, and in 0.78% of all trials, respectively. Both response keys were used about equally often (on average, 100.08 and 99.92 times per participant for the left and the right key, respectively). Mean RTs were 354 and 358 ms for the control and the experimental groups, respectively, $t(98)=0.26$, $p=.793$, $d=0.05$, $BF=0.22$.

Test phase. Anticipations, omissions, and false alarms in no-go trials occurred in 0.02%, 0.30%, and 2.18% of the trials. Mean RTs in go trials were 419 and 406 ms in the control and the experimental groups, respectively, $t(98)=1.12$, $p=.264$, $d=0.22$, $BF=0.37$.

Mean percentages of congruent choices are visualised in Figure 3 (top panel). The percentage of congruent choices was larger than 50% for the control group, $t(49)=3.12$, $p=.003$, $d=0.44$, $BF=10.57$, but not for the experimental group, $t(49)=0.64$, $p=.526$, $d=0.09$, $BF=0.19$. The comparison of both groups was significant, $t(98)=2.23$, $p=.028$, $d=0.45$, $BF=1.88$. That this test did not reach any of the preregistered thresholds of $BF < 1/6$ or $BF > 6$ was the reason why we stopped data collection at the maximum sample size. Note, however, that the anticipated group effect is significant from a traditional frequentist perspective.

Bimodality of percentages of congruent choices. Figure 3 (lower panel) visualises kernel density plots of the percentages of congruent choices for both groups. The visual impression is similar to what Sun et al. (2020) and Eichfelder et al. (2023) reported. Both groups exhibit a normal-like distribution centred at 50%. In addition, however, a

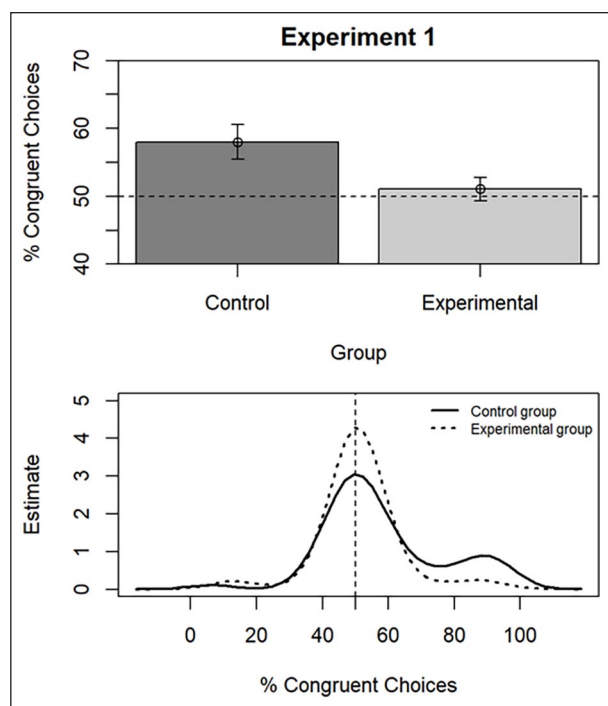


Figure 3. Results of Experiment 1.

Note. Upper panel: Mean percentages of congruent choices for the control and experimental groups. Error bars indicate one standard error of the mean. Lower panel: Kernel density plots for the percentages of congruent choices for the control and experimental groups.

second peak at the upper end of the scale is visible for the control group. Despite this visual impression, the bimodality coefficients were 0.31 and 0.11 for the control and the experimental groups and thus both below 0.55.

Discussion

The results of this experiment are similar to those reported by Eichfelder et al. (2023) who examined generalisation of R-E learning to a superordinate category. First, the results obtained with the control group replicate an overall response bias, that is, a mean percentage of congruent choices larger than 50%. Second, this effect was significantly smaller and not significantly different from 50% for the experimental group. In fact, the BF provided evidence for the absence of an R-E-learning transfer effect in the experimental group. Thus, abstraction from the location of the effect to a conceptual representation did not happen. Third, a bimodal distribution of the percentages of congruent choices was visible for the control group, but not for the experimental group. Although the respective bimodality coefficients did not exceed the critical value of 0.55, the visual impression is clearly in line with bimodality. Thus, we consider the control group data a (somewhat weak) replication of the bimodal pattern first reported by Sun et al. (2020).

Experiment 2

Given the results of Experiment 1 and of Eichfelder et al. (2023), the question arises as to whether the present experimental approach can yield evidence for abstraction or generalisation at all. To address this question, we presented participants in the experimental group with line drawings of the corresponding words used in the acquisition phase of the control group in Experiment 2. Only the words were then used in the test phase, as in the previous experiment (see Figure 2). There are several reasons why abstraction might be facilitated with these materials. First, previous studies have shown marked similarities (though also a contribution of non-overlapping neural machinery) between processing linguistic and non-linguistic input, such as line drawings (e.g., Ganis et al., 1996) and surface material (Dudschig et al., 2021), that seem to activate conceptual representations as well. Second, seeing a picture of an object conceivably suggests phonological recoding of the seen information. Critically, phonological recoding has been shown in experiments on R-E compatibility to be one potential factor that encourages generalisation (Földes et al., 2018; we will come back to this in the General Discussion in more detail). In addition, we assessed more directly whether participants do acquire the correct R-E relations in the acquisition phase by using a post-session questionnaire.

Method

Open practices statement. The preregistration of this experiment is available at <https://aspredicted.org/uy265.pdf> and the data are at <https://osf.io/jpvq8/>.

Sampling plan and participants. The sampling plan followed that of Experiment 1. Data were obtained from 108 students of the University of Bremen who participated for course credit or monetary compensation. Data from six participants were excluded, because they did not press the left and the right key at least 80 times each (of which 1 participant also responded in more than 20.0% of the no-go-trials). Two additional participants were excluded, because they participated in Experiment 1 already.

The final sample size was $n=100$ participants (mean age=25.03 years; 68 women, 32 men, 0 non-binary; 86 right-handed, 14 left-handed, 0 ambidextrous). Other criteria and characteristics apply as for Experiment 1.

Stimuli, apparatus, tasks, procedure, design, and analyses. By and large, the experiment resembled Experiment 1 with different stimuli and effects (see also Figure 2, right column). More precisely, in the control group, the words “APFEL” (Engl. “APPLE”) and “KATZE” (Engl. “CAT”) were used as effects and go stimuli, and the word “HAUS” (Engl. “HOUSE”) was used as the catch stimulus in the

acquisition phase. Schematic pictures of an apple, a cat, and a house were used for the experimental group. During the test phase, the words “APFEL” and “KATZE” were the go stimuli and the letter string “XXXXX” indicated a no-go trial.

After the experiment proper, a questionnaire was filled by the participants. The first part comprised two multiple-choice questions regarding the R-E relation during the acquisition phase, and the second part was an open question regarding response strategies during the test phase. We report results from the first part; the second part was mainly used to screen for participants not taking the experiment seriously.

Results

Acquisition phase. Anticipations, omissions, and missed catch trials occurred in 2.81%, 3.13%, and in 0.17% of all trials, respectively. Both response keys were used about equally often (on average, 99.52 and 100.48 times per participant for the left and the right key, respectively). Mean RTs were 381 and 370 ms for the control and the experimental groups, respectively, $t(98)=0.68$, $p=.500$, $d=0.14$, $BF=0.26$.

Test phase. Anticipations, omissions, and false alarms in no-go trials occurred in 0.01%, 0.44%, and 1.81% of the trials. Mean RTs in go trials were 417 and 419 ms in the control and the experimental groups, respectively, $t(98)=0.10$, $p=.918$, $d=0.02$, $BF=0.21$.

Mean percentages of congruent choices are visualised in Figure 4 (upper panel). The percentage of congruent choices was larger than 50% for the control group, $t(49)=4.33$, $p<.001$, $d=0.61$, $BF=301.71$, as well as for the experimental group, $t(49)=3.48$, $p=.001$, $d=0.49$, $BF=26.65$. In addition, the comparison of both groups yields some evidence for the null hypothesis of no group difference, $t(98)=1.14$, $p=.257$, $d=0.23$, $BF=0.38$.

Bimodality of percentages of congruent choices. Figure 4 (lower panel) visualises kernel density plots of the percentages of congruent choices for both groups. Both groups exhibit a normal-like distribution centred at 50%. In addition, however, a second peak at the upper end of the scale was clearly visible for the control group. While no clear second peak is visible for the experimental group, the bimodality coefficients were 0.71 and 0.67 for the control and the experimental groups, respectively, and thus both exceeded the critical threshold of 0.55.

Questionnaire. Ninety-two participants identified both R-E relations correctly. Within the control group, 3, 1, and 46 participants identified no, one, or two of the R-E relations correctly. The corresponding values for the experimental

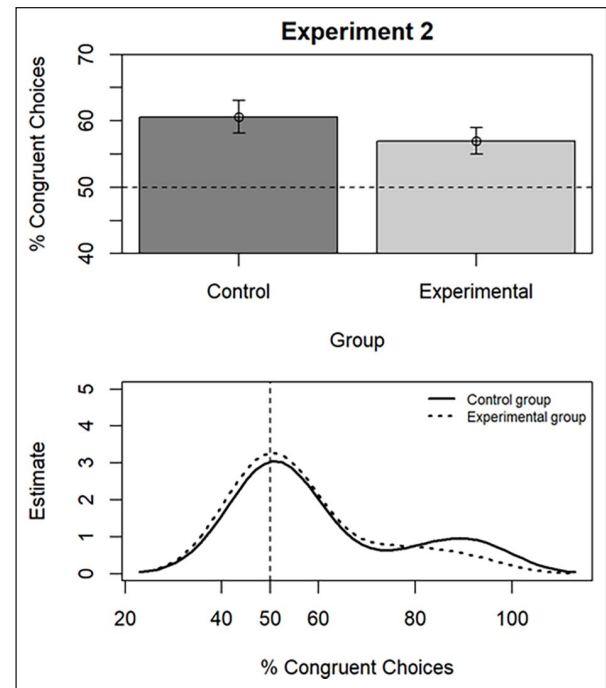


Figure 4. Results of Experiment 2.

Note. Upper panel: Mean percentages of congruent choices for the control and experimental groups. Error bars indicate one standard error of the mean. Lower panel: Kernel density plots for the percentages of congruent choices for the control and experimental groups.

groups were 2, 2, and 46. These frequencies were comparable in both groups, $\chi^2(2)=0.53$, $p=.766$.

Discussion

First, and as in Experiment 1, the results from the control group replicate an overall response bias. Second, and in contrast to Experiment 1, we observed a response bias of almost similar size in the experimental group as well. Thus, abstracting the content of pictures to yield a conceptual representation similar to the one resulting from linguistic input can apparently occur. This shows that the experimental approach can—in principle—yield results consistent with generalisation or abstraction. Yet, it also shows that this happens only under very restricted conditions, for example, situations where phonological recoding seems very likely. Third, bimodal distributions of the percentages of congruent choices were again observed; visually more pronounced for the control group, although the bimodality coefficient exceeded the critical value of 0.55 for both groups in this experiment.

Experiment 3

Although we observed evidence for bimodality of the percentages of congruent choices in Experiments 1 and 2 (and in Eichfelder et al., 2023), it was not particularly

pronounced. In this third experiment, we aimed to increase the learning strength and thereby the overall response bias with a new acquisition phase design to investigate whether the bimodality also becomes more pronounced and more clearly visible. Sun et al. (2020) compared two different instructions, one not mentioning the response effects and the other mentioning them, but did not observe effects of this manipulation, although other research showed that instructions do matter (Eder & Dignath, 2017). Our approach was to make the effects relevant in the experimental group, as this manipulation has been shown to increase the R-E compatibility effect in several studies (e.g., Ansorge, 2002; Janczyk et al., 2015). For participants in one group, the control group, the acquisition phase was standard (as in the previous experiments), although we changed some details to make this phase more similar to earlier research on R-E learning (e.g., the effects were low- and high-pitch tones as in the original study by Elsner & Hommel, 2001). The task of the experimental group, however, was not merely pressing response keys to produce the effect tones, but these participants had to replay sequences of tones during the acquisition phase. Specifically, they heard a sequence of tones and were then asked to press the corresponding sequence of response keys to re-produce the heard sequence. The length of the sequences, and thereby the difficulty, varied as a function of correct replay. This gamification-like aspect was intended to improve participants' motivation (see, for example, Sailer et al., 2013).

Overall, we expected a larger response bias in the experimental group. The new acquisition phase also introduced other features that might further increase learning strength. As the purpose of Experiment 3 was not to further understand mechanisms of the acquisition phase, but to maximise the degree of learning and then to study the bimodality in the testing phase, these additional influences on learning strength are desirable. One particularly relevant design feature is that the tone sequences served as the stimulus instructing the response sequence as well as its effect. This match of stimulus–effect relations was not present in the control group and there is some indication that it might affect learning strength. In particular, the experiments reported in Elsner and Hommel (2001) were run in an A- and a B-variant. In the A-variant, pressing a response key in the test phase also re-played the associated tone, which did not occur in the B-variant. At least descriptively, the congruency effects in RTs (Exp. 1) and the response biases (Exp. 2–4) were larger in the A- than in the B-variant. Hence, this feature may also help to increase the response bias in the experimental group.

In sum, Experiment 3 provides an independent data set where we tested again for bimodality with a different type of acquisition phase. To avoid misunderstandings at this point, Experiment 3 is not concerned with generalisation,

but with increasing the response bias and studying the bimodal distribution of test responses.

Method

Open practices statement. The preregistration of this experiment is available <https://aspredicted.org/uw862.pdf> and the data are provided at <https://osf.io/jpvq8/>.

Sampling plan and participants. Sample size was determined by sequential sampling with BFs (Schönbrodt & Wagenmakers, 2018; Schönbrodt et al., 2017) in a similar way as was described for the preceding experiments. Starting with $n_{\min}=40$ (i.e., 20 participants per group), the following stopping rules applied:

1. When a $BF > 6$ or a $BF < 1/6$ was obtained in a Bayesian two-sample t test, and at the same time, a $BF > 6$ is calculated for a Bayesian one-sample t test either for the control group or the experimental group (or both).
2. When a maximum number of $n=50$ participants per group has been reached.

Data were obtained from 103 participants of the same pool as for the preceding experiments. Exclusion criteria were preregistered: The data from one participant of the control group were excluded for choosing one response key less than 80 times during the acquisition phase. Two additional participants were excluded, because they had participated in previous experiments already.

The final sample consisted of $n=100$ participants (mean age=24.45 years; 64 women, 34 men, 0 non-binary; 90 right-handed, 8 left-handed, 2 ambidextrous). Other criteria and characteristics apply as for Experiment 1.

Stimuli and apparatus. The same equipment was used as for the preceding experiments. In contrast to those experiments, however, two sinusoidal tones (400 Hz [low-pitch] and 800 Hz [high-pitch], duration: 200 ms) were presented via loudspeakers as effects and (go) stimuli (as were also used in Exp. 3 of Janczyk et al., 2023). In addition, a bell chiming sound of approximately 200 ms was the no-go stimulus in the test phase. A centrally presented white square was used as the go stimulus in the acquisition phase. In addition, a centrally presented red square and a white exclamation mark appeared during the acquisition phase of the experimental group, serving to inform participants about the trial progress as detailed below. All visual stimuli were presented against a black background. Responses were given on the left and right “CTRL” key of a standard QWERTZ keyboard.

Tasks and procedure. The major change to the previous experiments concerned the *acquisition phase* for the control and the experimental groups.

For the *control group*, the acquisition phase was similar to a typical one used in earlier research (Elsner & Hommel, 2001; Janczyk et al., 2023, Exp. 3). More precisely, each trial started with the white square appearing for 200 ms as a go signal. Then, participants were to press the left or right response key as fast as possible within 1,200 ms. Each key produced either the low- or the high-pitch tone as its effect (the key-tone assignment was counterbalanced and perfectly predictable for each participant). The next trial started after an intertrial interval of 1,500 ms. Trials associated with RTs shorter than 100 ms were considered anticipations and those with RTs longer than 1,200 ms were considered omissions. Error feedback was as in the preceding experiments and these trials were repeated at a random position of the block. The experiment began with an unanalysed familiarisation block of 16 trials followed by four experimental blocks of 50 trials each. After each block, participants were informed about the number of left and right key presses given in the respective block.

For the *experimental group*, each trial of the acquisition phase started with a red square visible as long as the tone sequence was played. Random sequences of (low- and high-pitch) tones were presented in intervals of 1,200 ms. The sequences started with a length of two tones and could reach a maximum length of five tones. Once a sequence was presented, the exclamation mark occurred for 1,000 ms to indicate that participants are now to repeat the sequence by pressing the corresponding keys. As for the control group, each key press resulted in either the low- or the high-pitch tone. Also similar to the control group, each key press was prompted by the white square and had to be given within 1,200 ms. Pressing a key deleted the square and the next square appeared until the length of the sequence had been reached. Participants were not informed about the key-tone mapping and thus had to explore this relation during the first trials. (An erroneous assignment by the experimenter led to a slight imbalance of the two possible key-tone assignments in the experimental group [24 vs. 26 participants].) When three successive sequences of one length were reproduced correctly, the length was increased by one (until the maximum length of 5 was reached). When a sequence was not reproduced correctly, the length was reduced by one (until the minimum length of 2 was reached). Omission and anticipation errors were fed back as in the control group. Producing an erroneous tone led to the error message “FALSCHER TON!” (Engl. “wrong tone”) displayed for 500 ms. The acquisition phase ended when the participant would exceed 200 valid key presses (i.e., effect productions) with the next sequence. This was done to ensure that both groups performed a roughly equal number of key presses during the acquisition phase. Certainly, however, the overall duration of the acquisition phase is longer for the experimental than for the control group. When the number of valid key presses exceed a multiple of 50, a break was allowed (to remain similar to the acquisition phase of the control group).

The subsequent *test phase* was identical for both groups. Fifty percent of the trials started with the white square for 200 ms together with either the low- or the high-pitch tone as the (free-choice) go stimulus and participants were to press the left or the right key within 1,200 ms. In the other 50% of the trials, the no-go sound was played and participants had to withhold all responses for 1,200 ms. Error messages were as for the preceding experiments and erroneous trials were repeated at a random position within the block. All stimuli were randomly intermixed. The test phase comprised two blocks of 100 trials each. After the experiment proper, a questionnaire similar to that of Experiment 2 was administered.

Design and analyses. The analyses followed those of the preceding experiments. Mean RTs for the acquisition phase of the experimental group were calculated by first averaging RTs within each sequence of key presses and then taking the average of these averages. Anticipations, omissions, and wrong key presses refer to sequences, as a sequence was aborted once one of these errors was committed.

Results

Acquisition phase. In the control group, anticipations and omissions occurred in 3.30% and 1.16% of all trials. Both response keys were used about equally often (on average, 101.24 and 98.76 times per participant for the left and the right key, respectively). In the experimental group, anticipations, omissions, and wrong key presses occurred in 7.14%, 2.84%, and 6.33% of the sequences.

Mean RTs were 279 and 293 ms for the control and the experimental groups, respectively, $t(98)=1.26$, $p=.212$, $d=0.25$, $BF=0.42$. As expected, the acquisition phase of the experimental group took longer (12.18 min) than that of the control group (7.49 min), $t(98)=23.10$, $p<.001$, $d=4.62$, $BF=5.99$. The test phases of the experimental group (8.68 min) and the control group (8.78 min) were roughly of the same length, $t(98)=1.89$, $p=.062$, $d=0.38$, $BF=1.02$.

Test phase. Only one anticipation was registered; omissions and false alarms in no-go trials occurred in 1.11% and 1.88% of the trials. Mean RTs in go trials were 696 and 690 ms in the control and the experimental groups, respectively, $t(98)=0.40$, $p=.689$, $d=0.08$, $BF=0.23$.

Mean percentages of congruent choices are visualised in Figure 5 (upper panel). The percentage of congruent choices was larger than 50% for the control group, $t(49)=5.27$, $p<.001$, $d=0.75$, $BF>10^3$, and for the experimental group, $t(49)=7.76$, $p<.001$, $d=1.10$, $BF>10^6$. In addition, the comparison of both groups yields evidence towards the alternative hypothesis of a group difference, $t(98)=2.22$, $p=.029$, $d=0.44$, $BF=1.82$. As in Experiment 1, the group difference is significant, but

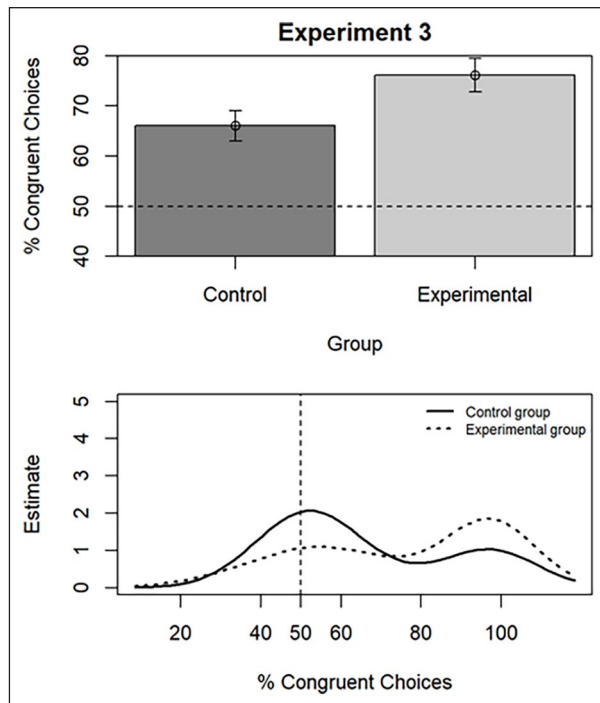


Figure 5. Results of Experiment 3.

Note. Upper panel: Mean percentages of congruent choices for the control and experimental groups. Error bars indicate one standard error of the mean. Lower panel: Kernel density plots for the percentages of congruent choices for the control and experimental groups.

not very strong in terms of the BF, and was again the reason why we collected the maximum sample size.

Bimodality of percentages of congruent choices. Figure 5 (lower panel) visualises kernel density plots of the percentages of congruent choices for both groups. The visual impression suggests bimodal distributions for both groups with one peak around 50% and a second peak at around 90%–100%. This second peak was even more pronounced for the experimental group. The bimodality coefficients were 0.72 and 0.65 for the control and the experimental groups and thus both larger than 0.55.

Questionnaire. Eighty-nine of the participants correctly identified both R-E relations. Within the control group, 7, 2, and 41 participants identified no, one, or two of the R-E relations correctly. The corresponding values for the experimental groups were 2, 0, and 48. Although a χ^2 -test missed significance, $\chi^2(2) = 5.38$, $p = .070$, the descriptive trend suggests that the R-E relations were more correctly picked up in the experimental group.

Discussion

The results of this experiment show that changing the acquisition phase successfully increased the response bias in the experimental group. Admittedly, this effect cannot be attributed with certainty to one particular feature of the acquisition

phase (see Introduction to this Exp. 3). While future research might aim at identifying particularly important features in comparison with the standard acquisition phase, here we were merely interested in increasing the learning strength and investigating how this affects the bimodality in the distribution of congruent choices. This also serves to test whether this bimodality generalises to different types of acquisition phases. Indeed, clear bimodal distributions were observed in both groups' data. Thus, for both groups, the overall response bias seems to be a result of averaging data from participants pursuing two different strategies: some participants responding more or less randomly (as required by the instructions), and other participants responding congruently in a large number of trials. This pattern was even more pronounced in the experimental group (where the overall response bias was also slightly larger).

General discussion

The present study reports three experiments addressing generalisation of effect representations and the representational nature of R-E learning. We begin by summarising the main results, followed by discussing their theoretical implications for the two main questions of the study. Based on this, we outline limitations of the present study and outstanding questions to be addressed in future research.

Summary of the main results

In all experiments, an on-average response bias was observed for the control groups, thus replicating earlier results with a free-choice test phase (e.g., Elsner & Hommel, 2001, Exp. 2–4; Janczyk et al., 2023, Exp. 3; Pfister et al., 2011). The transfer of acquired R-E learning was examined in the experimental groups of Experiments 1 and 2. Experiment 1 did not provide hints of a response bias in the experimental group. This was only the case in Experiment 2, where the response biases of both groups were statistically not different.

In addition, if an on-average response bias was observed, the whole distribution of the individual percentages of congruent choices was bimodal, replicating earlier results (Eichfelder et al., 2023; Sun et al., 2020, Exp. 1): A first peak occurred at around 50%, which means that many participants had no bias towards congruent choices. A second peak was visible at around 90%–100%. This indicates that some participants responded in an acquisition congruent way. This effect was particularly strong in the experimental group of Experiment 3, where we aimed to make the R-E relation particularly strong with a different acquisition phase.

Generalisation of effect representations

The results obtained in Experiment 1 do not suggest any abstraction of the effect locations to their semantic meaning, similar to the results of Eichfelder et al. (2023). Thus,

this result is a further empirical demonstration contradicting earlier claims that effect representations generalise to further similar or related effects—at least when a free-choice test phase is employed. At the same time, the results of Experiment 2 show that some sort of abstraction and generalisation is possible, when a situation known to facilitate such generalisation is created. It shows that the experimental setup can—in principle—yield results in line with generalisation. Taken together, these results suggest limiting factors under which generalisation or abstraction can occur at all (as discussed in more detail at the end of this section).

One might object that comparing the response bias between the control and the experimental groups (in Exp. 1 and 2) is not the fairest comparison. After all, changes to stimuli typically lead to generalisation decrement (e.g., Wheeler et al., 2006). Against this background it might seem favourable to analyse whether the experimental groups show any response bias on their own. Indeed, this was not the case in Experiment 1 (and in Eichfelder et al., 2023), but was the case for Experiment 2. Hence, the conclusions in general remain unchanged. In this regard, it is of interest to note that in Experiment 1 of Hommel et al. (2003), the reported congruency effect was similarly large for both groups, hence showing no evidence for any generalisation decrement (Exp. 2 and 3 had no control group).

The previous successful demonstrations of generalisation in the literature (Esser et al., 2023; Hommel et al., 2003) used forced-choice tasks in the test phase and thus focused on a congruency effect on RTs. Admittedly, it is unclear why we and Eichfelder et al. (2023) could not conceptually replicate these results with free-choice test phases. One possibility is that free-choice tasks are not well-suited to measure what is intended to be measured here (see also Custers, 2023).

Before we discuss this point in more detail, we first discuss results on generalisation in previous R-E compatibility studies, to provide a more comprehensive picture. Such R-E compatibility experiments are different from the current R-E learning experiments, but speak to a conceptually related issue. Their focus, however, is more on the second assumption of the ideomotor principle as laid out in the Introduction, that is, that actions are selected by anticipating their effects. The first systematic demonstration was provided by Kunde (2001). In a proto-typical experiment with spatial R-E compatibility, participants respond with a left or right response key and pressing a key makes an effect appear on the left or right side of the computer screen. Two compatibility conditions are usually implemented in separate blocks: In compatible blocks, the left key produces the left effect and the right key produces the right effect; in incompatible blocks, the left key produces the right effect and the right key produces the left effect. The crucial observation is that RTs are shorter in the compatible than in the incompatible condition, even though the

(predictable and thus anticipate-able) effect occurs only after RT has been measured. This result has been (conceptually) replicated many times (e.g., Janczyk et al., 2017, 2023; Janczyk & Lerche, 2019; Kunde, 2003; Pfister & Kunde, 2013) and similar considerations have successfully been used to explain interference in dual-tasking as well (Janczyk & Kunde, 2020).

With regard to generalisation, results from such R-E compatibility studies are mostly consistent with our findings though. In an early study, Koch and Kunde (2002) found some generalisation in two experiments with verbal colour-word responses (e.g., uttering “red” or “blue”) and written colour words as effects. The R-E compatibility effect was larger when the effects were also written in the respective colour, and it was absent when the effect was a coloured neutral letter string. These results might be taken to suggest that some abstraction of the effect colour indeed takes place. Alternatively, reading a colour word may automatically result in phonological recoding which could be (in)compatible with the verbal responses. Indeed, an R-E compatibility effect was observed in one study with phonological overlap of vocal responses and the effects in the same (German) language where phonological recoding of the effects could interfere with the vocal responses, but not when the effect words appeared in their translated English version (Földes et al., 2018). Also, no R-E compatibility effect was observed with category words as responses and exemplar words as effects or vice versa (Koch et al., 2021). A recent study took an even more simplified approach (Janczyk & Miller, 2024). In one group of Experiment 1, the visual effects appeared at predictable locations on the left or right side of the screen and an R-E compatibility effect was observed. In a second group (and in Exp. 2 and 3 of that study), the effect still occurred on the left or right side of the screen, but unpredictably at one of three locations. With this manipulation, no consistent R-E compatibility effect was observed any longer. Apparently, participants could not abstract from the three possible locations to the broader information that the effect occurs on the left or right side of the screen.

In sum, neither experiments on R-E learning nor on R-E compatibility have provided clear evidence for generalisation or abstraction of the effect representations. It thus seems as if the established representations of effects do not transfer to new situations, although stimuli are known to generalise. One boundary condition, however, might be that transfer occurs if phonological recoding takes place (our Experiment 2 and Földes et al., 2018). This requires, however, the implicit assumption that such phonological recoding did not occur in our Experiment 1. Our main argument is that the location of the circle in the experimental group of Experiment 1 is not part of the object’s immediate semantic/conceptual meaning to the same degree as compared to seeing the picture of a cat (as in Exp. 2). In addition, a reviewer of this article suggested that

participants might be less likely to encode the spatial (up/down) feature categorically or phonologically to avoid confusion with the spatial (left/right) dimension of the responses.

We concur that the present experiments cannot clarify whether any transfer, if a transfer occurs at all, happens during acquisition or at test. Overall, however, we believe that the present state of available empirical results warrants primarily the identification of situations under which generalisation and abstraction can be observed reliably. If these could be identified, it is of further theoretical interest whether an associative and a propositional account make different predictions regarding the nature of generalisation (e.g., perceptual vs. conceptual generalisation). In the next section, we discuss the underlying learned representations from the two perspectives in more detail.

What is learned in R-E learning?

The second question raised in the introduction concerns the representational nature of R-E learning. Traditionally, it is assumed that experiencing the co-occurrence of responses and their ensuing effects leads to “the incidental, implicit acquisition of action-outcome associations” and re-activation of the effect representation “will spread activation to the associated motor pattern” (Watson et al., 2015, p. 46; see Elsner & Hommel, 2004, for the role of contingency and contiguity). Thus, the traditional view entertains associative learning “characterized in terms of the establishment of links between representations” (McLaren et al., 1994, p. 316) which are created “automatically, regardless of the subject’s plans or intentions” (p. 321; see also Moeller & Pfister, 2022). If this applied to R-E learning, this learning should occur in all individuals, albeit possibly to different degrees. Against the background of Figure 1, the results obtained with the control groups of Experiments 1 and 2 and both groups of Experiment 3 match best with the scenario visualised in Panel 1b, that is, a bimodal distribution of the percentage of congruent choices with one peak around 50% and another at a much higher percentage.³ This scenario, however, is not in line with the classic ideomotor assumption of learned bidirectional R-E associations (see also Custers, 2023; Sun et al., 2020) and might also point to a difference between stimulus-response and R-E learning.

In contrast, the observed bimodality and the fact that almost all participants can report the R-E relation of the acquisition phase invites a different explanation (see also Sun et al., 2020): Participants rapidly extract the relevant rules in the acquisition phase and some of them use them in the test phase while others do not. What could be the reason for those participants responding almost always with a congruent choice? Simple demand characteristics seem to be a very likely explanation for this behaviour. This means that participants either responded in the

simplest way or in the way they reasoned the experimenter would be interested in. Notably, however, Experiment 1 and the experiment reported by Eichfelder et al. (2023) also indicate that such behaviour does not generalise to situations with different stimuli involved in the test phase as compared to the acquisition phase. This is noteworthy, as there are studies showing that analogical transfer to novel stimuli is almost perfect when rules have been learned (e.g., Casale et al., 2012). This comes close to our situation and suggests that—in principle—generalisation of rules to new stimuli could be observed.

This reasoning certainly deviates from the classic thinking about R-E learning as outlined in the Introduction (see, in particular, Footnote 1), but appears highly interesting for a full understanding of the ideomotor principle and the evidence existing in favour of it (see also Custers, 2023, and Kunde & Janczyk, 2024). Note, however, that even if some results that have traditionally been interpreted in favour of the ideomotor principle can be better explained by a propositional account, we do not aim at advancing a propositional account of ideomotor theory here. Rather, we offer an account for the present results that acknowledges mechanisms going beyond those assumed in the classical ideomotor principle view.

Limitations and future avenues

One clear limitation of the present Experiments 1 and 2 (and also of Eichfelder et al., 2023) is that the to-be-abstracted stimuli in the experimental group were not varied. This was done to closely replicate the original experimental approach first used by Hommel et al. (2003). However, variation of the exemplars seems to be an important factor for categorical abstraction and increased variability yields more stable category representations (e.g., Cohen et al., 2001; Hahn et al., 2005; see also Reichmann et al., 2023, for related work on attitudes and a recent summary of the impact of variability on abstraction). Similarly, increasing the variability of semantic content in artificial languages improves the detection of invariant structures (Gómez, 2002). Thus, a straightforward extension would be testing an experimental group that received more than one effect (e.g., three exemplars of the category “animal” and three exemplars of the category “furniture”) in future research.

Furthermore, given the susceptibility of free-choice test phases to response strategies, it might be worthwhile to focus more on forced-choice test phases and RTs as the dependent variables. Although Sun et al. (2020) did not observe compatibility effects on RTs in their Experiments 2 and 3 with traditional test phase designs, many others did (e.g., Eder & Dignath, 2017; Elsner & Hommel, 2001, Exp. 1; Hommel et al., 2003; Watson et al., 2015, Exp. 1). In addition, such compatibility effects were observed in a modified design in which learning and testing occur

simultaneously on each trial (Sun et al., 2022). Thus, using this latter experimental approach in combination with multiple effects in the experimental group might be worthwhile. It should be noted though that the RT effects reported by Sun et al. were rather small (e.g., $d=0.24$ in their Exp. 1). The approach used by Esser et al. (2023) appears promising in this regard as it yielded larger effect sizes. However, future research should first address the mentioned limitations, that is, isolating R-E from stimulus–stimulus learning and assessing the impact of seeing the old items again, prior to drawing strong conclusions.

Can we explain congruency effects in RTs (and error rates) with a propositional rule-like account as well? We believe that two variants of such tasks must be distinguished. First, if participants are split into two groups and one group is then required to respond with a reversed (i.e., incongruent) mapping (as in Exp. 1 of Elsner & Hommel, 2001), it is certainly possible that other aspects than only putatively established associations contribute to the observed RT effect (see also Custers, 2023, who brought up such thoughts). Second, consider a situation in which participants learn two response-colour links during the acquisition phase. At test, however, they are instructed to respond to the identity of coloured letters. The task-irrelevant colour can be either the one produced by the required response (congruent trial) or not (incongruent trial). RTs are shorter in congruent compared with incongruent trials (see Paelecke & Kunde, 2007, Exp. 4 and 5, or Wolfensteller & Ruge, 2011, for such experiments). Arguably, effects below 100 ms as observed in most RT tasks are unlikely to be produced strategically (i.e., such RT effects are unlikely being “faked”). Therefore, we do not see a convincing explanation without bidirectional associations for such results so far (see also Kunde & Janczyk, 2024).

At present, one might also embrace the following perspective when considering all available evidence: In general, the ideomotor principle’s assumption of bidirectional associations holds, that is, activation of an effect representation leads to some activation of the linked response. Certainly, this does not suffice to actually produce this response (if this were true, humans could be operated like a remote control car). The important point is that forced- and free-choice test phases are not interchangeable and likely tap into two different learning mechanisms: While associations might be at work in free-choice tasks as well, their impact is rather limited and easily overshadowed by potential strategies participants opt to choose. Some participants seem to pick up demand characteristics and then respond with a congruent choice in a high percentage of trials. By contrast, others might interpret the instructions to mean that they should respond randomly (which was indeed intended, but not explicitly mentioned). Some studies, including the one by Elsner and Hommel (2001), explicitly instruct random responding and seem to argue that, despite this explicit instruction, an on-average response bias indicated an influence of learned R-E

associations that could not be avoided by participants. The critical point here is that the average effect, which has been taken as clear evidence towards an automatic response priming by means of associations, is actually a misleading representation of the underlying bimodal distribution, which results from (at least) two different response patterns. Note also that a possible resemblance of free-choice tasks and random generation tasks has been discussed in the literature (Frith, 2013; Naefgen & Janczyk, 2018). These points add to the uncertainty of whether free-choice tasks are suited at all to assess the existence of R-E associations.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This project was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation)—Project 422180965 (jointly awarded to M.J. and V.H.F.) within the Research Group 2718: Modal and Amodal Cognition (Project 381713393).

ORCID iD

Markus Janczyk  <https://orcid.org/0000-0002-9958-3220>

Data accessibility statement



The data from the present experiment are publicly available at the Open Science Framework website: <https://osf.io/jpvq8/>

Notes

1. Similar statements can be found elsewhere, too. For example, Dignath et al. (2014) noted that “observations of response-contingent effects create bidirectional links between cognitive representations of responses and effects” (p. 1701); Kunde et al. (2007) wrote that these “action-effect links can be used in both ways, to predict the sensory consequences when the movement is known and, more importantly, to recruit the required movement when a certain effect is intended” (p. 72); and the bidirectionality is emphasised as well by Watson et al. (2015) who noted that “Ideomotor theories of action control propose that the anticipation of action goals emerges from the acquisition of bidirectional action–outcome associations” (p. 45). Custers (2023) summarised this as “thinking of or imagining a particular outcome, would lead to activation of the motor programs that produced the outcome in the past, without explicit knowledge of how motor programs are related to outcomes” (p. 262). While there is little systematic investigation on the limits of building these automatic and implicit associations, Watson et al. (2015) pointed to the idea that this might only occur when the possible R-E associations are limited in

number. In sum, the acquisition of bidirectional associations seems a fundamental and widespread assumption in ideomotor principle-inspired research and provides the theoretical basis for the rationale of the experiments presented in the current article.

2. Unfortunately, the term “free-choice task” and the related term “forced-choice task” (which will be introduced soon in the current article) have different meanings in different research contexts. Our use goes back to the work of Berlyne (1957) who used two stimuli in his experiments. If only one stimulus was presented, the participants were to correctly provide an instructed response. If both stimuli were presented, participants had to choose which of the two possible responses they give. In Berlyne’s words, these “experiments enabled the traditional choice reaction time (which we may call ‘forced-choice RT’) to be compared with what we may call ‘free-choice RT’” (p. 107, emphases in original). However, in other fields such as decision-making, the term “forced-choice” is sometimes used when participants need to select from several alternatives, while “free-choice” means that they can also select to not choose any alternative (e.g., Brazell et al., 2006). Being aware of this, we prefer to use the terms in the sense of Berlyne here, as this is the standard meaning in ideomotor research, but also in the broader context of action research in general.
3. In exploratory analyses, we fitted a mixture of two Gaussian distributions to the proportion of congruent choices data from the control groups of Experiments 1 and 2 and both groups of Experiment 3 using the `normalmixEM()` function of the `mixtools` package in R (Benaglia et al., 2010). The best-fitting components had means(sds) at 0.50(0.05)/0.69(0.22) (Exp. 1), 0.51(0.06)/0.88(0.07) (Exp. 2), 0.53(0.08)/0.97(0.04) (Exp. 3, control group), and 0.58(0.18)/0.98(0.02) (Exp. 3, experimental group). The main point here is that indeed one of the underlying Gaussian distributions was (more or less) centred at 0.50 (i.e., at 50% congruent choices), while the other had its centre at a much larger value. The relatively small mean of 0.69 of the second Gaussian distribution for Experiment 1 (together with the wide standard deviation) is the result of few participants who almost always responded with in-congruent choices (a similar result can be seen in Fig. 2 in Sun et al., 2020).

References

- Ansorge, U. (2002). Spatial intention–response compatibility. *Acta Psychologica*, 109, 285–299.
- Badets, A., Koch, I., & Philipp, A. M. (2016). A review of ideomotor approaches to perception, cognition, action, and language: Advancing a cultural recycling hypothesis. *Psychological Research*, 80, 1–15.
- Benaglia, T., Chauveau, D., Hunter, D. R., & Young, D. S. (2010). `mixtools`: An R Package for analyzing mixture models. *Journal of Statistical Software*, 32(6), 1–29.
- Berlyne, D. E. (1957). Conflict and choice time. *British Journal of Psychology*, 48, 106–118.
- Bowman, A. W., & Azzalini, A. (2021). *sm: Nonparametric smoothing methods* (R package version 2.2-5.7). <http://www.stats.gla.ac.uk/~adrian/sm>
- Brazell, J. D., Diener, C. G., Karniouchina, E., Moore, W. L., Séverin, V., & Uldry, P.-F. (2006). The no-choice option and dual response choice designs. *Marketing Letters*, 17, 255–268.
- Carpenter, W. B. (1852). On the influence of suggestion in modifying and directing muscular movement, independently of volition. *Royal Institution of Great Britain (Weekly Evening Meeting)*, (pp. 147–154).
- Casale, M. B., Roeder, J. L., & Ashby, F. G. (2012). Analogical transfer in perceptual categorization. *Memory & Cognition*, 40, 434–449.
- Cohen, A. L., Nosofsky, R. M., & Zaki, S. R. (2001). Category variability, exemplar similarity, and perceptual classification. *Memory & Cognition*, 29(8), 1165–1175.
- Custers, R. (2023). Thoughts about actions and outcomes (and what they lead to). *Motivation Science*, 9(4), 261–271.
- Dignath, D., Pfister, R., Eder, A. B., Kiesel, A., & Kunde, W. (2014). Representing the hyphen in bi-directional action-effect associations: Automatic integration of time intervals into cognitive action structures. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 40(6), 1701–1712.
- Dudschig, C., Kaup, B., Leuthold, H., & Mackenzie, I. G. (2021). Conceptual representation of real-world surface material: Early integration with linguistic-labels indicated in the N400-component. *Psychophysiology*, 58, Article e13916.
- Eder, A., & Dignath, D. (2017). Influence of verbal instructions on effect-based action control. *Psychological Research*, 81, 355–365.
- Eichfelder, L. A., Franz, V. H., & Janczyk, M. (2023). Is there hierarchical generalization in response-effect learning? *Experimental Brain Research*, 241, 135–144.
- Elsner, B., & Hommel, B. (2001). Effect anticipation and action control. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 229–240.
- Elsner, B., & Hommel, B. (2004). Contiguity and contingency in action-effect learning. *Psychological Research*, 68, 138–154.
- Esser, S., Haider, H., Lustig, C., Tanaka, T., & Tanaka, K. (2023). Action–effect knowledge transfers to similar effect stimuli. *Psychological Research*, 87, 2249–2258.
- Földes, N., Philipp, A. M., Badets, A., & Koch, I. (2018). Exploring the representational basis of response-effect compatibility: Evidence from bilingual verbal response-effect mappings. *Acta Psychologica*, 186, 1–7.
- Freeman, J. B., & Dale, R. (2013). Assessing bimodality to detect the presence of a dual cognitive process. *Behavior Research Methods*, 45(1), 83–97.
- Frith, C. (2013). The psychology of volition. *Experimental Brain Research*, 229, 289–299.
- Ganis, G., Kutas, M., & Sereno, M. I. (1996). The search for “common sense”: An electrophysiological study of the comprehension of words and pictures in reading. *Journal of Cognitive Neuroscience*, 8(2), 89–106.
- Gómez, R. L. (2002). Variability and detection of invariant structure. *Psychological Science*, 13(5), 431–436.
- Greenwald, A. G. (1970a). Sensory feedback mechanisms in performance control: With special reference to the ideomotor mechanism. *Psychological Review*, 77, 73–99.
- Greenwald, A. G. (1970b). A choice reaction time test of ideomotor theory. *Journal of Experimental Psychology*, 86(1), 20–25.
- Guttman, N., & Kalish, H. I. (1956). Discriminability and stimulus generalization. *Journal of Experimental Psychology*, 51(1), 79–88.

- Hahn, U., Bailey, T. M., & Elvin, L. B. C. (2005). Effects of category diversity on learning, memory, and generalization. *Memory & Cognition*, 33(2), 289–302.
- Harleß, E. (1861). Der Apparat des Willens [The apparatus of will]. *Zeitschrift für Philosophie und Philosophische Kritik*, 38, 50–73.
- Herbart, J. F. (1825). *Psychologie als Wissenschaft neu gegründet auf Erfahrung, Metaphysik und Mathematik [Psychology as a science newly founded on experience, metaphysics, and mathematics]*. August Wilhelm Unzer.
- Hommel, B., Alonso, D., & Fuentes, L. (2003). Acquisition and generalization of action effects. *Visual Cognition*, 10, 965–986.
- Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The theory of event coding: A framework for perception and action. *Behavioral and Brain Sciences*, 24, 849–878.
- Janczyk, M., Durst, M., & Ulrich, R. (2017). Action selection by temporally distal goal-states. *Psychonomic Bulletin and Review*, 24, 467–473.
- Janczyk, M., Giesen, C. G., Moeller, B., Dignath, D., & Pfister, R. (2023). Perception and action as viewed from the Theory of Event Coding: A multi-lab replication and effect size estimation of common experimental designs. *Psychological Research*, 87, 1012–1042.
- Janczyk, M., & Kunde, W. (2020). Dual tasking from a goal perspective. *Psychological Review*, 127, 1079–1096.
- Janczyk, M., & Lerche, V. (2019). A diffusion model analysis of the response-effect compatibility effect. *Journal of Experimental Psychology: General*, 148, 237–251.
- Janczyk, M., & Miller, J. (2024). Generalization of unpredictable action effect features: Large individual differences with little on-average effect. *Quarterly Journal of Experimental Psychology*, 77, 898–908.
- Janczyk, M., Naefgen, C., & Kunde, W. (2020). Are freely chosen actions generated by stimulus codes or effect codes? *Attention, Perception, & Psychophysics*, 82, 3767–3773.
- Janczyk, M., Yamaguchi, M., Proctor, R. W., & Pfister, R. (2015). Response-effect compatibility with complex actions: The case of wheel rotations. *Attention, Perception, & Psychophysics*, 77, 930–940.
- Kaup, B., Ulrich, R., Bausenhardt, K. M., Bryce, D., Butz, M. V., Dignath, D., Dudschig, C., Franz, V. H., Friedrich, C., Gawrilow, C., Heller, J., Huff, M., Hütter, M., Janczyk, M., Leuthold, M., Mallot, H., Nürk, H.-C., Ramscar, M., Said, N., & Wong, H. Y. (2024). Modal and amodal cognition: An overarching principle in various domains of psychology. *Psychological Research*, 88, 307–337.
- Koch, I., Földes, N., Kunde, W., & Philipp, A. M. (2021). Exploring the role of verbal-semantic overlap in response-effect compatibility. *Acta Psychologica*, 215, Article 103275.
- Koch, I., & Kunde, W. (2002). Verbal response-effect compatibility. *Memory & Cognition*, 30(8), 1297–1303.
- Kunde, W. (2001). Response-effect compatibility in manual choice reaction tasks. *Journal of Experimental Psychology: Human Perception and Performance*, 27(2), 387–394.
- Kunde, W. (2003). Temporal response-effect compatibility. *Psychological Research*, 67(3), 153–159.
- Kunde, W., Elsner, K., & Kiesel, A. (2007). No anticipation—No action: The role of anticipation in action and performance. *Cognitive Processing*, 8, 71–78.
- Kunde, W., & Janczyk, M. (2024). Thoughts about thoughts about actions and outcomes: Comment on Custers (2023). *Motivation Science*.
- Lemon, J. (2006). Plotrix: A package in the red light district of R. *R-News*, 6(4), 8–12.
- McLaren, I. P. L., Green, R. E. A., & Mackintosh, N. J. (1994). Animal learning and the implicit/explicit distinction. In N. C. Ellis (Ed.), *Implicit and explicit learning of languages* (pp. 313–332). Academic Press.
- Moeller, B., & Pfister, R. (2022). Ideomotor learning: Time to generalize a longstanding principle. *Neuroscience and Biobehavioral Reviews*, 140, Article 104782.
- Morey, R., & Rouder, J. (2022). *BayesFactor: Computation of Bayes Factors for common designs* (R package version 0.9.12-4.4). <https://CRAN.R-project.org/package=BayesFactor>
- Naefgen, C., Dambacher, M., & Janczyk, M. (2018). Why free choices take longer than forced choices: Evidence from response threshold manipulations. *Psychological Research*, 82, 1039–1052.
- Naefgen, C., & Janczyk, M. (2018). Free choice tasks as random generation tasks: An investigation through working memory manipulations. *Experimental Brain Research*, 236, 2263–2275.
- Paelecke, M., & Kunde, W. (2007). Action-effect codes in and before the central bottleneck: Evidence from the PRP paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, 33, 627–644.
- Pfister, R., & Kunde, W. (2013). Dissecting the response in response-effect compatibility. *Experimental Brain Research*, 224(4), 647–655.
- Pfister, R., & Janczyk, M. (2016). schoRsch: An R package for analyzing and reporting factorial experiments. *The Quantitative Methods for Psychology*, 12, 147–151.
- Pfister, R., & Janczyk, M. (2012). Harleß' apparatus of will: 150 years later. *Psychological Research*, 76, 561–565.
- Pfister, R., Kiesel, A., & Hoffmann, J. (2011). Learning at any rate: Action-effect learning for stimulus-based actions. *Psychological Research*, 75(1), 61–65.
- Pfister, R., Schwarz, K. A., Janczyk, M., Dale, R., & Freeman, J. B. (2013). Good things peak in pairs: A note on the bimodality coefficient. *Frontiers in Psychology*, 4, Article 700.
- R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.
- Reichmann, K., Hütter, M., Kaup, B., & Ramscar, M. (2023). Variability and abstraction in evaluative conditioning: Consequences for the generalization of likes and dislikes. *Journal of Experimental Social Psychology*, 108, Article 104478.
- Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, 8, 382–439.
- Sailer, M., Hence, J., Mandl, H., & Klevers, M. (2013). Psychological perspectives on motivation through gamification. *Interaction Design and Architecture(s) Journal*, 19, 28–37.
- SAS Institute Inc. (1990). *SAS/STAT user's guide, version 6* (4th ed.).
- Schönbrodt, F., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic Bulletin Review*, 25, 128–142.

- Schönbrodt, F., Wagenmakers, E. J., Zehetleitner, M., & Perugini, M. (2017). Sequential hypothesis testing with Bayes factors: Efficiently testing mean differences. *Psychological Methods*, 22(2), 322–339.
- Shin, Y. K., Proctor, R. W., & Capaldi, E. J. (2010). A review of contemporary ideomotor theory. *Psychological Bulletin*, 136, 943–974.
- Stock, A., & Stock, C. (2004). A short history of ideo-motor action. *Psychological Research*, 68, 176–188.
- Sun, D., Custers, R., Marien, H., & Aarts, H. (2020). Ideomotor action: Evidence for automaticity in learning, but not execution. *Frontiers in Psychology*, 11, Article 185.
- Sun, D., Custers, R., Marien, H., Liefhooge, B., & Aarts, H. (2022). Examining mechanistic explanations for ideomotor effects. *Journal of Experimental Psychology: Human Perception and Performance*, 48(5), 458–466.
- Thorndike, E. L. (1913). Ideo-motor action. *Psychological Review*, 20, 91–106.
- Watson, P., van Steenbergen, H., de Wit, S., Wiers, R. W., & Hommel, B. (2015). Limits of ideomotor action-outcome acquisition. *Brain Research*, 1626, 45–53.
- Wheeler, D. S., Amundson, J. C., & Miller, R. R. (2006). Generalization decrement in human contingency learning. *Quarterly Journal of Experimental Psychology*, 59, 1212–1223.
- Wolfensteller, U., & Ruge, H. (2011). On the timescale of stimulus-based action–effect learning. *Quarterly Journal of Experimental Psychology*, 64(7), 1273–1289.
- Wulff, D. U., Kieslich, P. J., Henninger, F., Haslbeck, J. M. B., & Schulte-Mecklenbeck, M. (2023). *Movement tracking of psychological processes: A tutorial using mousetrap*. PsyArXiv. <https://doi.org/10.31234/osf.io/v685r>