Acta Psychologica 217 (2021) 103330

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/locate/actpsy

Not so social after all: Video-based acquisition of observational stimulus-response bindings

Carina G. Giesen [a,*], Christian Frings [b]

[a] Friedrich Schiller University Jena, Jena, Germany
[b] Trier University, Trier, Germany

ARTICLE INFO

Keywords:
Stimulus-response binding
Episodic retrieval
Observational learning
Dyadic interaction
Video-based learning

ABSTRACT

Merely observing how another person responds to a stimulus results in incidental stimulus-response (SR) bindings in memory. These observationally acquired SR bindings can be retrieved on a later occasion. Retrieval will bias current behavioral response tendencies towards re-execution of the observed response. Previous demonstrations of this effect endorsed a dyadic interaction paradigm in which two co-actors respond in alternating fashion. The present paper investigates a video-based version of the observational SR binding task in which videotaped responses are observed on screen. Whereas findings from the dyadic paradigm indicate that retrieval of observationally acquired SR bindings is modulated by social relevance, the video-based paradigm is not influenced by social moderators. Data of four experiments show that manipulations of visual perspective, natural and artificial group membership had no modulatory effect on retrieval of observationally acquired SR bindings in the video-based paradigm. The absence of any socially modulated effect in the video-based paradigm is supported by Bayesian statistics in favor of the null hypothesis. Data from a fifth experiment suggests that observational SR binding and retrieval effects in the video-based paradigm reflect the influence of spatial attention allocated towards response keys of observed responses. Implications for the suitability of both paradigms to study observational learning and joint action phenomena are discussed.

1. Introduction

The idea of episodic stimulus-response (SR) bindings or "event files" (Hommel, 1998) is central for recent accounts of automatic, stimulus-based action regulation (Frings et al., 2020; B. Hommel et al., 2001; Logan, 1988; Rothermund et al., 2005). Accordingly, whenever an action is executed in response to a stimulus on a given *prime* trial, their mental codes become integrated, resulting in transient SR bindings that are stored in memory. Stimulus repetition on a later occasion (e.g., in the subsequent *probe* trial) triggers retrieval of the associated response. This will facilitate or impede probe performance, depending on whether the retrieved response is appropriate or not in a new situation. To date, a burgeoning amount of findings attests that storage and retrieval of SR bindings are pervasive principles of action regulation and apply to a broad scope of stimuli and responses (for an overview, see Henson et al., 2014).

1.1. The observational SR binding task

Although the vast majority of studies focus on bindings between stimuli and self-executed responses, this is no necessary condition for the formation of SR bindings. Giesen et al. (2014) (see also Giesen et al., 2016) recently showed that stimuli also become bound to responses that are merely observed (but not executed by oneself). They developed the *observational SR binding task*, which represents a color categorization task shared between a dyad of two co-actors. Participants classified the color of word stimuli as red or green via button press in alternating fashion in a sequential prime-probe design. One participant (the prime actor) categorized the color of a word stimulus presented in the prime trial. At the same time, the other participant (the prime observer) saw only the word stimulus (but no color) and had to observe the action of the prime actor. To test whether mere observation of the prime response results in an observationally acquired SR binding, the former prime observer became actor in the subsequent probe trial and had to categorize the color of the probe word stimulus. Stimulus relation from prime to probe (word repetition vs. change) and compatibility between

* Corresponding author at: Friedrich Schiller University Jena, Department of Psychology, General Psychology II, Am Steiger 3, Haus 1, 07743 Jena, Germany.
E-mail address: carina.giesen@uni-jena.de (C.G. Giesen).

https://doi.org/10.1016/j.actpsy.2021.103330
Received 18 February 2020; Received in revised form 2 February 2021; Accepted 3 May 2021
Available online 12 May 2021

observed prime and to-be-performed probe responses (compatible vs. incompatible) were manipulated orthogonally. Similar to "standard" SR binding and retrieval effects, stimulus repetition in the probe trial should retrieve observed responses from memory, producing facilitation (interference) when retrieved responses are compatible (incompatible) with to-be-performed probe responses.

In sum, findings by Giesen and colleagues indicate that people rely on observationally acquired SR bindings to regulate their own actions. However, retrieval effects in this dyadic variant of the observational SR binding task only emerged if the observed responses were performed by *socially relevant* others. Specifically, in the study by Giesen et al. (2014), retrieval effects were only obtained when participants had to interact in a cooperative or competitive way, but were absent when participants worked independently of each other. In addition to this "situationally induced" interdependence between co-actors, Giesen et al. (2018) showed that retrieval of observationally acquired SR bindings also occurs for more chronic forms of interdependence, as is constituent for romantic relationships. In their study, retrieval effects for observationally acquired SR bindings emerged only for romantically involved interaction partners, but not for pairs of strangers. This highlights that chronic interdependence will produce automatic retrieval of observational SR bindings independently of the task itself.

Intriguingly, findings from the dyadic observational SR binding task bear a close structural resemblance to research on observational learning. Since the seminal work of Bandura (1986), it is widely accepted that our action knowledge does not solely rest on our own experiences. Instead of relying on our individual "trial-and-error" learning history stemming from novel or problematic situations, many behaviors can be acquired by observing the behavior of others in similar situations. Observing how other persons act and behave in particular situations provides us with a wide variety of vicarious experiences. However, one does not simply copy every observed behavior. Whether an action will be expressed in one's own behavior (or not) is strongly influenced by motivational factors. That is, individuals are motivated to interact and, thus, learn from people that are socially relevant for them (Bandura, 1986).

Whereas social learning theory provides a sound analysis of the macro-processes that are involved in observational learning phenomena (e.g., attention, memory retention, action reproduction, and motivation; cf. Bandura, 1986), far less is known about the underlying cognitive micro-processes and mechanisms by which observed actions become automatized. Although prominent accounts of automatic action regulation such as Logan's Instance Theory (1988) the Theory of Event Coding (TEC; B. Hommel et al., 2001), or Binding and Retrieval in Action Control (BRAC, Frings et al., 2020) provide some views on the micro-processes involved in automatization and action planning, these accounts have neglected the influence of social learning by mere observation so far. Given that social relevance affects observational SR binding and retrieval phenomena in a way that mimics effects known from observational learning research, the observational SR binding task bridges the gap between social learning theory and recent accounts of behavior automatization.

1.1.1. Cognitive mechanisms underlying the observational SR binding task

Finding a modulation of observationally acquired SR bindings by social relevance provides us with deeper insight into the underlying (socio-)cognitive mechanisms involved in this task. It is reasonable to assume that effects of social relevance on transient SR binding and retrieval are mediated by attentional processes: One will automatically attend more to actions that were performed by a socially-relevant (compared with a socially-irrelevant) interaction partner.

The beneficial influence of attention on both, the emergence and retrieval of SR bindings is an established and well documented finding for bindings between stimuli and self-performed actions (Logan, 1988; Moeller & Frings, 2014; Singh et al., 2018; but see B. Hommel, 2005; for a discussion, see Henson et al., 2014;). Apparently, the same argument can be applied to SR bindings that were acquired by mere observation (Giesen et al., 2014; 2016; see also Giesen et al., 2018, for a more elaborate discussion).

1.2. Socially modulated effects in joint action tasks

Of course, the dyadic observational SR binding task is not the only paradigm which reflects close parallels between merely observing versus self-performing an action. Additional evidence stems from a variety of other tasks like observational acquisition of response-effect bindings (Paulus et al., 2011), imitation tasks (Brass et al., 2001), behavioral mimicry (Van Baaren et al., 2009) or joint action tasks (Sebanz et al., 2003). Importantly, many of the effects investigated in these tasks are also moderated by social relevance. For instance, interference effects in the Joint Simon task (Sebanz et al., 2003) become stronger as a function of interdependence among co-actors (B. Hommel et al., 2009; Iani et al., 2011; Ruys & Aarts, 2010). The same holds true for behavioral mimicry and automatic imitation (for a distinction see Genschow et al., 2017). For instance, individuals imitate others less strongly when they observe actions from a third-person perspective rather than a first-person perspective (Bortoletto et al., 2013; O. Genschow et al., 2013; Lamm et al., 2007; Vogt et al., 2003), when they face out-group members (O. Genschow & Schindler, 2016), non-human agents (e.g., Liepelt et al., 2010; Longo & Bertenthal, 2009), or when they are in a competitive mode as compared to a cooperative mode (e.g., Weyers et al., 2009). Together, evidence from the mentioned tasks documents that social factors exert a strong top-down modulatory influence on the cognitive micro-processes in contexts in which people act together.

1.3. The present study: video-based observational SR binding effects

In the joint action tasks (Sebanz et al., 2003) it has been controversially discussed whether participants performing a task with an unseen co-actor need to believe that the unseen co-actor is actually a human and not a computer (Sellaro et al., 2013; Welsh et al., 2007) – or whether the belief of an intentional agent together with a spatial event is needed to produce joint action effects. Against this background, one may ask whether observational SR binding and retrieval effects emerge only in the dyadic setting of the task, that is, when two participants sit opposite to each other and respond in alternating fashion, or whether these bindings can be acquired via other ways of behavior observation that do not require the presence of a co-actor. If so, it should be possible to acquire observational SR bindings also via observing responses only in a *video*, rather than executed by another person.

Based on this reasoning, Experiment 1 was designed to test whether observational SR binding effects also emerge in a video-based variant of the task. To anticipate results, this was indeed the case. However, something was odd with the findings from Experiment 1. In particular, these findings were obtained although participants received *no information at all* about the person performing the videotaped prime responses. This is at odds with findings from the dyadic observational SR binding task because in the dyadic paradigm, observational SR bindings were only retrieved if responses are observed in socially relevant interaction partners (e.g., persons with whom one interacts in cooperative or competitive manner or romantic partners; Giesen et al., 2014, 2018) On the other hand, this discrepancy resembles the inconsistent data pattern concerning interpersonal relationships of co-actors as observed in the joint Simon task (see Dolk et al., 2014, for an overview).

Thus, Experiment 2–5 were motivated to further investigate the discrepant findings between the dyadic and video-based version of the observational SR binding task. Whereas the former yielded retrieval effects only when interaction partners were socially relevant for observers, the latter yielded retrieval effects even in the absence of any social relevance of hand models carrying out observable actions. This discrepancy of findings suggests that social relevance is not the only factor that affects attention allocation in the observational SR binding task. Thus, one has to look out for other potential factors that will affect attentional processes which then mediate acquisition and retrieval of observational SR bindings.

However, before one can engage in such an endeavour, a number of potential differences between the dyadic and video-based variant of the observational SR binding task need to be resolved first. This is done to rule out that these differences account for why SR retrieval effects occurred in the absence of social relevance manipulation in the video-based version of the task. In the dyadic observational SR binding task, (a) responses are observed from a third-person perspective and (b) it is obvious that observed responses were executed by another person (namely, participants' interaction partner who is sitting in front of them). The present study addresses differences in visual perspective (Exp. 2), and self-other differentiation (Exp.'s 3, 4) between both task variants to investigate whether and to which extent retrieval of observational SR bindings acquired via video observation is affected by social moderators in top-down fashion. In turn, Experiment 5 was designed to test for a simple cognitive alternative explanation, which argues that observational SR binding and retrieval in the video-based paradigm is a function of spatial attention that promotes binding and retrieval even in the absence of social relevance of hand models.

1.4. *Required sample size, ethics vote and access to research data*

A-priori power analyses with the G*Power software (Faul et al., 2007) showed that a sample size of at least $n = 27$ was required to guarantee sufficient statistical power of 1-ß = 0.80 with $\alpha = 0.05$ (one-tailed) to detect a medium-sized two-way (Exp. 1) or three-way (Exp.'s 2 to 5) interaction effect ($d_z = 0.50$). Thus, all experiments aimed to sample at least the required minimum sample size, but collected more data if possible (i.e., in the given time period labs were available). For all experiments, we endorsed a significance level of $\alpha = 0.05$, two-tailed. Follow-up tests on stimulus repetition effects for compatible and incompatible sequences as well as stimulus relation × response compatibility interaction effects in the "first-person perspective" (Exp. 2), "in-group" condition (Exp. 3, 4) and "reference to social agent" (Exp. 5) were tested according to a one-tailed test logic due to the directional nature of their prediction.

All experiments were in accordance with the Ethical standards of the Institute of Psychology of the FSU Jena and in accordance with the Declaration of Helsinki. For all experiments, all participants provided informed consent to take part in the study.

All experiments, stimulus material, data and analyses scripts will be made publically available at the Open Science Framework (OSF; https://osf.io/bh4wa/) after initial acceptance of the paper.

2. Experiment 1 (video-based observational SR binding)

A video-based variant of the observational SR binding task was developed that is performed individually by participants. First, a word stimulus is presented during a prime display, together with a videotaped color categorization response. Participants are instructed to closely observe which response is shown in the video (an occasional memory test ensures this). In the subsequent probe display, it is then the participants' turn to execute a response to categorize the color of a word stimulus. The to-be-executed probe response can be compatible or incompatible with the observed videotaped prime response. Furthermore, the word stimulus from the prime display either repeats in the probe display or not. Following the same rationale underlying all "standard" SR binding and retrieval effects, stimulus repetition from prime to probe display should retrieve the observationally acquired SR binding and affect probe performance: Accordingly, performance *benefits* should result if observed prime responses are compatible with to-be-executed probe responses, since the retrieved SR binding facilitates response selection. In turn, performance *costs* should be obtained whenever observed prime responses are incompatible with to-be-executed probe responses, since the retrieved SR binding interferes with response selection. This pattern is reflected in an interaction effect of stimulus relation (stimulus repetition vs. change) from prime to probe and compatibility between observed prime responses and required probe responses (compatible vs. incompatible).

2.1. *Method*

2.1.1. *Participants*

Forty-two German native speaking participants took part in the experiment. One participant had to be excluded due to excessive error rates (more than 15% probe errors and/or more than 25% errors in the memory test), another participant due to too many slow responses (more than 25% responses slower than 1000 ms). Thus, data of 40 participants were analyzed (29 female, $M_{\text{age}} = 23$ years, range: 19–32 years). Participants were tested individually and received partial course credit or sweets for participating. The experiment lasted 25–30 min.

2.1.2. *Material*

The experiment was programmed with *E*-Prime 2.0. Twenty-five neutral, mono- or disyllabic German adjectives (see Appendix 1) served as word stimuli and were presented centrally in Times New Roman font (16 pts) on a black 19" LED monitor. In prime displays, word stimuli were presented in white font; in probe displays, word stimuli were presented in either red or green font. Furthermore, during prime displays, a brief video sequence (duration: 1000 ms) was presented, which showed video-taped categorization responses on a response pad of either a red or green response. Videos displayed the following movement sequence: First, both left and right index fingers of a hand model were presented in resting position placed on the red and green response keys. Subsequently, one index finger was lifted upwards and moved downwards again to press the respective (red or green) response key and resumed to the resting position. Response keys in the videos were tagged with red or green adhesive labels which were clearly visible to participants.Although video-taped red/green responses were executed by the left/right hand of a hand model, we refrained from any reference to spatial positions of responses keys throughout the experiment. Instead, instructions framed observed (prime) and to-be-executed (probe) responses as "red responses" or "green responses". Participants had the task to observe both word stimulus and the videotaped responses during the prime display. In the subsequent probe display, participants task was to categorize the color of the word stimulus themselves by pressing the corresponding key on a response pad (identical to the response pad depicted in the prime videos, see Fig. 1). Response pads were connected to computers via the parallel port.

2.1.3. *Design*

The experiment comprised a 2 × 2 within subjects design with the factors stimulus relation (word stimulus repetition vs. word stimulus change [baseline]) and response compatibility (response compatible vs. response incompatible) of observed prime and to-be-executed probe responses. Stimulus relation was manipulated by randomly choosing a word stimulus for prime displays that was repeated (50%) or changed (50%) for probe displays with the restriction that a given word stimulus could not be sampled for more than two prime-probe sequences in a row. Response compatibility was manipulated by requiring a probe response that was either compatible (50%) or incompatible (50%) to observed videotaped prime responses. Type of observed prime response (red or green key press) was counterbalanced within participants. Probe reaction times (RT, in ms) were dependent measures of interest; however, probe error rates (in %) were analyzed as well.

2.1.4. *Procedure*

Upon their arrival in the lab, participants were seated at individual computers. Instructions were given on screen. Participants placed their left and right index fingers on the red and green keys on the response

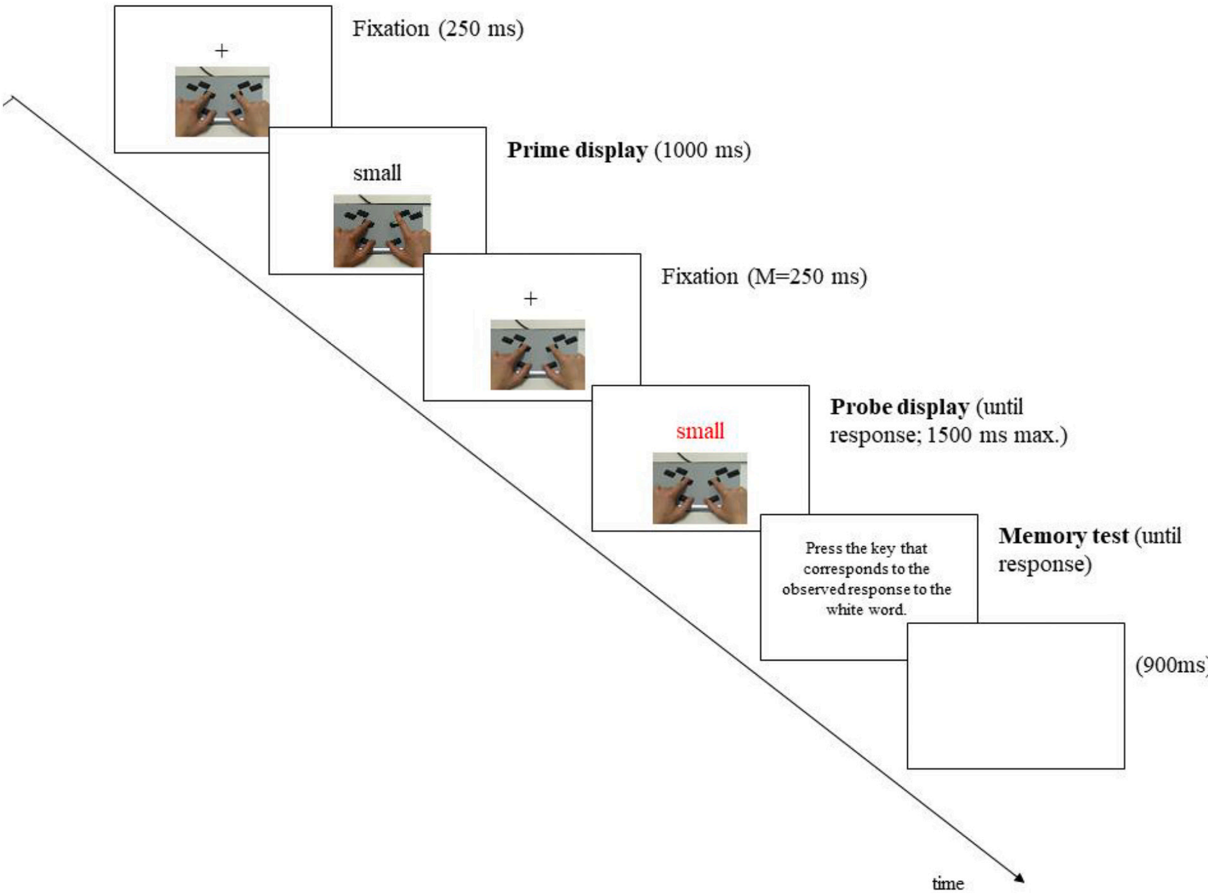

Fig. 1. Example for trial sequence in Experiment 1. Stimuli are not drawn to scale.

pads and kept them there during the entire experiment ("resting position"). Two additional keys, labelled with "go" were operated via pressing either the left or right thumb and served to navigate through instructions and to initiate the practice and experimental block. Participants were informed that trials consisted of a prime-probe sequence. During the prime display, they would first see a word stimulus and a brief video sequence in which either a red or green response was shown. Participants were instructed to observe these videotaped responses. Apart from that, no additional information about videotaped responses (e.g., which task was executed by hand models) was provided. Participants were further informed that in the subsequent probe display, another word stimulus would appear in either red or green font and that they had to categorize its color by pressing the corresponding key on the response pad. The meaning of the word stimuli was irrelevant for the probe task and served as a distractor (Rothermund et al., 2005). Participants were asked to respond as quickly and accurately as possible. To ensure that participants paid sufficient attention while observing the videotaped prime responses, a memory test was randomly presented after 25% of all probe displays that asked participants to press the key that corresponded to the videotaped response they just observed in the preceding prime display.

After reading instructions, participants worked through 32 prime-probe sequences for practice in which they received feedback for erroneous or too slow ($>$ 1000 ms) probe responses. After completion of the practice block, participants worked through 320 experimental prime-probe sequences that were constructed according to the experimental design.

Trial structure of experimental prime-probe sequences was as follows (Fig. 1): A sequence started with a fixation display, consisting of a centrally presented fixation cross and a still picture of a response pad with two hands in resting position placed below (duration: 250 ms). Then, the prime display followed in which a word stimulus was presented centrally in white font; below the word stimulus, the video of a green or red response press was shown (duration: 1000 ms). Then, another fixation display was presented for a variable duration (150–350 ms; $M = 250$ ms) to prevent anticipation of probe responses. Then, the probe display followed: Word stimuli were centrally presented in either red or green font (until response; maximum duration: 1500 ms) together with the still picture. After 25% of all probe displays, a memory test appeared, asking participants to press the response key that corresponded to the observed videotaped prime response (until response) that was executed when the word was white. Finally, a blank black screen was shown (duration: 900 ms). Then, the next trial started.

2.2. Results

2.2.1. Data preparation

Prior to analyses, 11.5% of all probe responses were discarded either because of errors (2.2%), or because of erroneous performance in the memory test (8.3%; overall: 1.9%), or because they qualified as RT outlier values[1] (1.3%). Then, means were computed for every condition of the factorial design for both RTs and error rates (see Table 1) and were entered to a 2×2 repeated measures analysis of variance (ANOVA) with the factors stimulus relation and response compatibility.

2.2.2. Probe performance

The ANOVA revealed only a significant Stimulus Relation $\times$

[1] Probe responses faster than 200 ms or slower than 3 interquartile ranges above the third quartile of the sample's RT distribution were regarded as outliers (Tukey, 1977).

Table 1
Mean Probe RT (in ms) and error rates (in %; printed in parentheses) for the factorial design of Experiment 1.

Probe RT (error rates)		
Stimulus Relation	Response Compatibility	
	C	IC
SR	426 (1,8)	439 (2.5)
SC	431 (2.5)	430 (1.7)
ΔSC-SR	5 (0.7)	−9 (−0.8)
$S \times R = (SC\text{-}SR)_C$ - $(SC\text{-}SR)_{IC}$	14 (1.5)	

SR = Stimulus repetition. SC = Stimulus change. C = compatible. IC = incompatible. ΔSC-SR = stimulus repetition effects. S × R = interaction effects of stimulus relation and response compatibility.

Response Compatibility interaction for probe RT, $F(1,39) = 12.18$, $p = .001$, $\eta_p^2 = 0.24$, and error rates, $F(1,39) = 5.55$, $p = .024$, $\eta_p^2 = 0.13$, reflecting retrieval of bindings between word stimuli and observed prime responses. Follow-up tests showed that compared to stimulus changes in the probe, word stimulus repetition produced a performance benefit (significant for RT: $\Delta = 5$ ms; $t[39] = 1.87$, $p = .035$ (one-tailed), $d_z = 0.30$; nonsignificant for errors: $\Delta = 0.7\%$; $t[39] = 1.57$, $p = .063$ (one-tailed), $d_z = 0.25$) for probe responses that were compatible with observed prime responses. In turn, compared to word stimulus changes in the probe, stimulus repetition produced a significant performance cost (RT: $\Delta = -9$ ms; $t[39] = 3.37$, $p = .001$ (one-tailed), $d_z = 0.53$; errors: $\Delta = -0.8\%$ ms; $t[39] = 2.17$, $p = .018$ (one-tailed), $d_z = 0.34$) for probe responses that were incompatible with observed prime responses.

2.3. Discussion

Experiment 1 showed robust retrieval effects of stimulus-response bindings between word stimuli and observed videotaped responses, indicated by significant interactions of stimulus relation and response compatibility. These effects emerged for both RT and error rates. Importantly, these findings were obtained although participants received *no information at all* about the person performing the videotaped prime responses. This indicates that in the video-based variant of the observational SR binding task, retrieval of observationally acquired SR bindings occurred in the *absence* of any social relevance. In and of itself, this insight is particularly striking, because it is at odds with findings from the dyadic variant of the observational SR binding task. In the dyadic paradigm, SR bindings that were acquired by observation are only retrieved if responses are observed in socially relevant interaction partners (e.g., persons with whom one interacts in cooperative or competitive manner or romantic partners, Giesen et al., 2014, 2016, 2018).

However, there is a striking difference between the video-based and the dyadic variant of the observational SR binding task concerning the visual perspective of observations: That is, the perspective of videotaped responses in the former variant reflects a "first-person" perspective that resembles one's view on one's own hands (e.g., while observing one's own probe response execution). In turn, the visual perspective in the latter reflects a "third-person" perspective, since participants are seated opposite each other in the dyadic paradigm. Effects of visual perspective on imitation are documented in the literature, showing reduced imitation when participants observe actions from a third-person perspective rather than a first-person perspective (Bortoletto et al., 2013; O. Genschow et al., 2013; Lamm et al., 2007; Vogt et al., 2003).

It is thus possible that observed responses in the video-based variant were more similar to one's own responses, which made them more likely to be retrieved automatically from memory. This would imply that only the retrieval of observational SR bindings that were obtained from a third-person perspective are prone to the influence of social relevance and consequently are retrieved only if responses were observed in relevant others. To address this alternative explanation, Experiment 2 used videotaped responses that differed in visual perspective (first- vs. third-person perspective). If the suggested explanation is correct, the video-based observational SR binding task should yield retrieval effects only when SR bindings were observed from a first-person perspective, but not when SR bindings were acquired from a third-person perspective (reflected in a three-way interaction between visual perspective, stimulus relation, and response compatibility).

3. Experiment 2

3.1. Method

3.1.1. Participants

Fifty German native speaking participants took part in the experiment. Applying the same criteria as in Experiment 1, six participants had to be excluded due to excessive error rates (more than 15% errors). Thus, data of 44 participants were analyzed (27 female, $M_{age} = 22.5$ years, range: 18–30 years). Participants were tested individually and received €2 and sweets for participating. The experiment lasted 30 min.

3.1.2. Material, design, and procedure

Material, design, and procedure paralleled Experiment 1 except for the following changes: Visual perspective of videotaped prime responses was added as an additional factor that was varied within subjects (first vs. third person perspective). Thus, the videos from Experiment 1 were replaced by four new videos (two for each perspective × two responses) that depicted red or green key presses either from the first- or third-person perspective (see Fig. 2). Visual perspective of videotaped prime responses was then varied blockwise within subjects (with 160 prime-probe sequences for each visual perspective). Block order (first-third vs. third-first person perspective) was counterbalanced between subjects. Preliminary analyses showed that block order did not modulate the effects of interest; hence, all analyses were performed with this factor collapsed. As a manipulation check, the identification with videotaped prime responses was assessed with three items at the end of each block: (1) I had the feeling that the filmed hands were my own; (2) Watching the videos felt like watching my own responses; (3) It was easy for me to differentiate between the responses shown in the videos and my own responses (reverse coded). Participants rated their agreement on a 7-point Likert scale (1 = do not at all agree; 7 = do fully agree).

3.2. Results

3.2.1. Manipulation checks

3.2.1.1. Identification with videotaped prime responses. Means were computed from all three items for each visual perspective (first-person perspective: Cronbach's $\alpha = 0.87$; third-person perspective: Cronbach's $\alpha = 0.71$). In general, identification scores were rather low and also did not differ between first- and third-person perspective ($M_{\text{first-person}} = 2.72$; $M_{\text{third-person}} = 2.44$, $t[43] = 1.25$, $p = .217$, $d_z = 0.19$).

3.2.1.2. Memory test for observed responses. To ensure that all probe actors paid sufficient attention to observed prime responses, average error scores for performance in the memory test were computed for each block. Importantly, error rates in the memory test did not differ between both visual perspectives ($M_{\text{first-person}} = 5.97\%$; $M_{\text{third-peron}} = 5.85\%$; $|t| < 1$, $p = .899$. This indicates that all participants adequately attended which response was executed in the prime videos, irrespective of the videos' visual perspective.

3.2.2. Probe performance

For RT only correct probe responses were analyzed. According to the same criteria as in Experiment, 6.0% of all probe responses were excluded: 2.6% because of errors, 5.9% because of erroneous responses

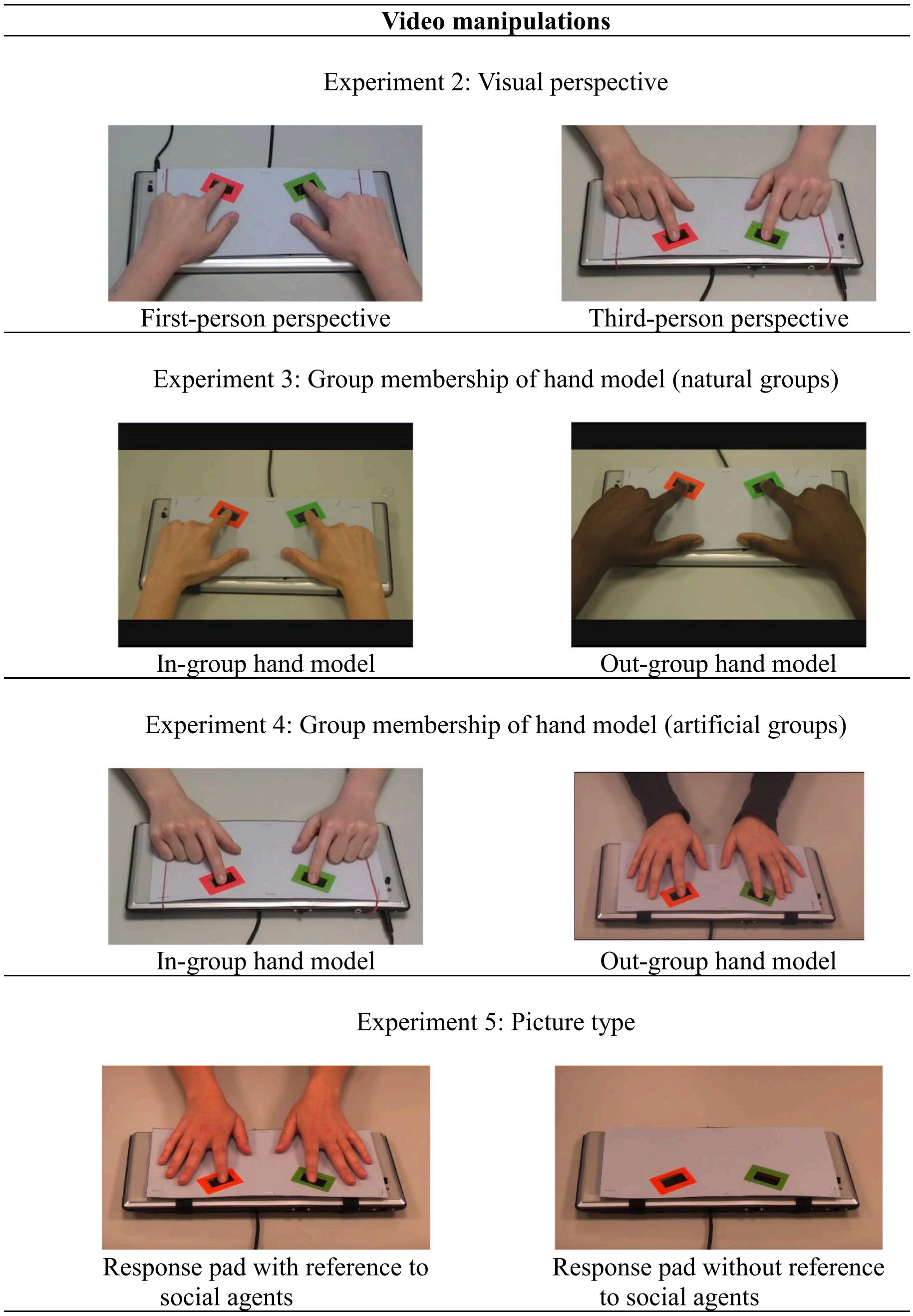

Fig. 2. Illustrations of video manipulations in Experiments 2 to 5.

in the memory test (1.5% of overall trials), and 1.9% because of RT outlier values. Means were then computed for the factorial design for RT and error rates (see Table 2) and were entered into a 2 (stimulus relation: word repetition vs. word change) × 2 (response compatibility: compatible vs. incompatible) × 2 (visual perspective: first- vs. third-person) ANOVA with repeated measures.

The ANOVA revealed a main effect of response compatibility for RT, $F(1,43) = 4.55$, $p = .039$, $\eta_p^2 = 0.10$, indicating that probe responses that were compatible with observed prime responses (M = 379 ms) were executed faster than incompatible probe responses (M = 387 ms). This effect was further qualified by a significant Stimulus Relation × Response Compatibility interaction for probe RT, $F(1,43) = 5.95$, $p = .019$, $\eta_p^2 = 0.12$, reflecting retrieval of bindings between word stimuli and observed prime responses. Follow-up tests showed that compared to word stimulus changes in the probe, word stimulus repetition produced a significant performance benefit ($\Delta = 6$ ms; $t[43] = 2.31$, $p = .013$ (one-tailed), $d_z = 0.35$) for probe responses that were compatible with observed prime responses. In turn, compared to stimulus changes in the probe, stimulus repetition produced a nonsignificant performance cost ($\Delta = -2$ ms; $t[43] = 1.28$, $p = .103$ (one-tailed), $d_z = 0.19$) for probe responses that were incompatible with observed prime responses. Importantly, the three-way interaction was not significant, $F < 1$, $BF_{01} = 7.413$.[2] Follow-up inspection showed that net S × R interaction effects were significant for first-person (S × $R_{\text{first-person}} = 7$ ms, $t[43] = 1.86$, $p =$

[2] All Bayes Factor analyses were conducted with the JASP software package (version 0.13.1.0). In detail, S × R interaction effects were used as dependent variable and were analyzed as a function of the third factor in a Bayesian *t*-test for independent groups (this test is mathematically equivalent to the test of three-way interaction in the ANOVA, with $F = t^2$).

Table 2
Mean Probe RT (in ms) and error rates (in %; printed in parentheses) for the factorial design of Experiments 2 to 5.

Experiment	Factorial Design				
2		Visual perspective			
		first-person		third-person	
	Stimulus Relation	Response Compatibility		Response Compatibility	
		C	IC	C	IC
	SR	377 (3.0)	389 (2.1)	375 (3.6)	386 (2.4)
	SC	381 (2.4)	386 (1.0)	381 (4.0)	385 (1.9)
	ΔSC-SR	4 (−0.6)	−3 (−1.1)	6 (0.4)	−1 (−0.5)
	S × R	7 (0.5)		7 (0.9)	
3		Group Membership (natural groups)			
		in-group		out-group	
		Response Compatibility		Response Compatibility	
		C	IC	C	IC
	SR	402 (1.3)	422 (2.0)	401 (1.5)	423 (2.2)
	SC	412 (2.0)	420 (2.3)	413 (2.2)	417 (1.1)
	ΔSC-SR	10 (0.7)	−2 (0.3)	12 (0.7)	−6 (−1.1)
	S × R	12 (1.0)		18 (1.8)	
4		Group Membership (artificial groups)			
		in-group		out-group	
		Response Compatibility		Response Compatibility	
		C	IC	C	IC
	SR	403 (2.1)	419 (2.0)	422 (0.9)	439 (1.7)
	SC	409 (2.0)	415 (1.4)	430 (1.6)	433 (1.2)
	ΔSC-SR	6 (−0.1)	−4 (−0.6)	8 (0.7)	−6 (−0.5)
	S × R	10 (0.5)		14 (1.2)	
5		Picture Type			
		With Hand		Without hand	
		Response Compatibility		Response Compatibility	
		C	IC	C	IC
	SR	415 (3.2)	432 (3.5)	414 (2.7)	433 (2.6)
	SC	426 (4.2)	426 (2.2)	426 (3.4)	429 (2.0)
	ΔSC-SR	9 (1.0)	−6 (−1.3)	12 (0.7)	−4 (−0.6)
	S × R	15 (2.3)		16 (1.3)	

SR = Stimulus repetition. SC = Stimulus change. C = compatible. IC = incompatible. ΔSC-SR = stimulus repetition effects. S × R = interaction effects of stimulus relation and response compatibility (see Table 1 for effect computation).

.035 (one-tailed, since this effect was predicted), $d_z = 0.28$, see Table 2 for computation of S × R interaction effects); but missed conventional levels of significance for third-person perspective ($S \times R_{\text{third-person}} = 8$ ms, $t[43] = 1.85$, $p = .072$, $d_z = 0.28$). All other effects were not significant (all $Fs < 1$).

For error rates, ANOVA revealed only significant main effects of visual perspective, $F(1,43) = 7.89$, $p = .007$, $\eta_p^2 = 0.16$, and response compatibility, $F(1,43) = 9.99$, $p = .003$, $\eta_p^2 = 0.19$, indicating that fewer probe errors occurred after observing responses in the first-person (M = 2.0%) compared with the third-person perspective (M = 3.0%), whereas more errors were made for probe responses that were compatible (M = 3.0%) with observed prime responses compared with incompatible probe responses (M = 2.0%). All other effects were not significant (Stimulus Relation × Response Compatibility interaction: $F[1,43] = 1.65$, $p = .21$, $\eta_p^2 = 0.04$; three-way interaction: $F < 1$; all other: $Fs < 3.8$; $ps > 0.058$).

3.3. Discussion

The results of Experiment 2 are clear: Visual perspective of videotaped prime responses did not modulate retrieval of observationally acquired SR bindings in die video-based paradigm. Further on, analysis of memory test performance yield no differences between visual perspectives, suggesting that participants paid sufficient attention to videotaped prime responses from both perspectives. Thus, the findings from Experiment 2 render it unlikely that participants from Experiment 1 were more prone to retrieve prime responses viewed from the first-person perspective as compared to prime responses viewed from the third-person perspective. This implies that the difference in visual perspective between the video-based and dyadic variants of the observational SR binding task cannot explain why retrieval effects in the video-based variant emerged in the absence of social relevance.

Finding no modulation by visual perspective is in line with findings from related tasks like the imitation-interference task (e.g., Brass et al., 2001) that yield robust imitation effects with pictures/picture sequences that adopt a third-person perspective. In this regard, one might argue that videos from the third-person perspective are still very similar to one's own response and just give the impression of watching one's *mirrored response.* Perceived similarity appears to be a central mediating process underlying imitative behavior (e.g., Kokal et al., 2009). In line with this reasoning, Giesen and Genschow (2020, in prep) gathered initial evidence that similarity between model and observer is also a mediating process underlying the retrieval of observational SR bindings. This would explain why manipulating the visual perspective in Experiment 2 yielded no modulatory effects: If observed responses felt like watching a mirrored response, this renders observed responses very similar to one's own responses. It is thus not surprising that this did not induce the feeling of observing *another person* to a sufficiently strong degree. Thus, Experiment 3 used a stronger manipulation to make it more apparent to participants that videos depict responses from another person. To make the differentiation between oneself and the other person as strong as possible, social group membership was used as an experimental factor. To make social group membership visually salient in the videos without explicitly directing participants' attention towards a social group context, prime videos either stemmed from hand models with white versus dark skin. It was reasoned that videos from white hands (representing the in-group) should be perceived as more similar to one's own responses compared with videos from dark hands (representing out-group members). As is known from research using other imitation tasks (e.g, O. Genschow & Schindler, 2016; Lakin et al., 2008), participants imitate less strongly if they face out-group rather than in-group members. If this reasoning this correct, then social group membership should also modulate retrieval of observational SR bindings. Retrieval of SR bindings should be more likely if responses belonged to in-group members (i.e., hand models with white skin), but should be reduced or even absent for responses that were observed in out-group members (i.e., hand models with dark skin; reflected in a three-way interaction of stimulus relation, response compatibility, and group membership).

4. Experiment 3

4.1. Method

4.1.1. Participants

Fifty-six German native speaking participants with white skin color took part in the experiment. Two participants reported a migration background; however, excluding these participants did not change the data pattern. Applying the same criteria as in Experiment 1, one participant had to be excluded due to excessive errors in the memory test. Thus, data of 55 participants were analyzed (36 female, M_{age} =

21.2 years, range: 18–29 years). Participants were tested individually and received partial course credit for participating. The experiment lasted 30 min.

4.1.2. Material, design, and procedure

Material, design, and procedure paralleled Experiment 2 except for the following changes: Visual perspective was replaced by group membership of videotaped prime responses by showing responses that were performed by hands of white- (representing the in-group) versus dark-skinned (representing the out-group) hand models in the videos. Thus, the videos from Experiment 2 were replaced by four new videos (2 group membership × 2 color responses) that depicted red or green key presses either from a hand model with white or dark skin (see Fig. 2). All videos adopted a first person perspective. Group membership of videotaped prime responses was then varied blockwise within subjects (with 160 prime-probe sequences for each hand model). Block order (in- vs. out-group hand model first) was counterbalanced between subjects. Preliminary analyses showed that block order did not modulate the effects of interest; hence, all analyses were performed with this factor collapsed. Again, the identification with videotaped prime responses was assessed as a manipulation check by using the same three items as in Experiment 2 after each block (i.e., for videos of each hand model).

4.2. Results

4.2.1. Manipulation checks

4.2.1.1. Identification with videotaped prime responses. Means were computed from all three items for each hand model (in-group [IG]: Cronbach's α =0.84; out-group [OG]: Cronbach's α =0.64). Again, identification scores with videotaped responses were rather low; however, scores were significantly higher for videos of responses carried out by in-group compared with out-group members ($M_{IG} = 2.47$; $M_{OG} = 2.17$, $t[54] = 2.15$, $p = .036$, $d_z = 0.29$), indicating that the manipulation was successful.

4.2.1.2. Memory test for observed responses. To ensure that all probe actors paid sufficient attention to observed prime responses, average error scores for performance in the memory test were computed for each block. Importantly, error rates in the memory test did not differ between hand models of different skin color ($M_{IG} = 3.95\%$; $M_{OG} = 4.64\%$; $t(54) = 1.06$, $p = .292$, $d_z = 0.14$). This indicates that all participants adequately attended which response was executed in the prime videos, irrespective of the hand models' social group membership/skin color.

4.2.2. Probe performance. For RT

Only correct probe responses were analyzed. According to the same criteria as in Experiment, 5.6% of probe responses were excluded: 1.9% because of errors, 4.3% because of erroneous responses in the memory test (1.1% of overall trials), and 1.5% because of RT outlier values. Means were then computed for the factorial design for RT and error rates (see Table 2) and were entered into a 2 (Stimulus relation: word repetition vs. word change) × 2 (Response compatibility: compatible vs. incompatible) × 2 (hand models' group membership: in- vs. out-group) ANOVA with repeated measures.

The ANOVA revealed a main effect of response compatibility for RT, $F(1,54) = 23.75$, $p < .001$, $\eta_p^2 = 0.31$, indicating that probe responses that were compatible with observed prime responses ($M = 410$ ms) were executed faster than incompatible probe responses ($M = 423$ ms). This effect was further qualified by a significant Stimulus Relation × Response Compatibility interaction for probe RT, $F(1,54) = 24.19$, $p < .001$, $\eta_p^2 = 0.31$, reflecting retrieval of bindings between word stimuli and observed prime responses. Follow-up tests showed that compared to word stimulus changes in the probe, word stimulus repetition produced a significant performance benefit ($\Delta = 11$ ms; $t[54]$ =4.23, $p < .001$ (one-tailed), $d_z = 0.57$) for probe responses that were compatible with observed prime responses. In turn, compared to word stimulus changes in the probe, word stimulus repetition produced a significant performance cost ($\Delta = -4$ ms; $t[43]$ =1.75, $p = .044$ (one-tailed), $d_z = 0.23$) for probe responses that were incompatible with observed prime responses. Importantly, the three-way interaction was not significant, $F < 1$, $BF_{01} = 14.518$. The net S × R interaction effects were significant for both hand models ($S \times R_{IG} = 11$ ms, $t[55] = 2.96$, $p = .002$ (one-tailed, since effect was predicted), $d_z = 0.40$; $S \times R_{OG} = 18$ ms, $t[55]$ =4.59, $p < .001$, $d_z = 0.61$). All other effects were not significant (all Fs < 3.83, all ps > 0.055).

For error rates, ANOVA revealed only a significant Stimulus Relation × Response Compatibility interaction, $F(1,54) = 6.20$, $p = .016$, $\eta_p^2 = 0.10$, reflecting retrieval of bindings between word stimuli and observed prime responses. Follow-up tests showed that compared to word stimulus changes in the probe, word stimulus repetition produced a significant performance benefit ($\Delta = 0.7\%$; $t[54]$ =2.17, $p = .017$ (one-tailed), $d_z = 0.29$) for probe responses that were compatible with observed prime responses. In turn, compared to word stimulus changes in the probe, word stimulus repetition produced a nonsignificant performance cost (errors: $\Delta = -0.4\%$; $t[54]$ =1.55, $p = .064$ (one-tailed), $d_z = 0.21$) for probe responses that were incompatible with observed prime responses. The three-way interaction failed the significance criterion, $F(1,54) = 2.51$, $p = .119$, $\eta_p^2 = 0.04$, as did all other effects (all Fs < 3.9; all ps > 0.084).

4.3. Discussion

Against expectations, manipulating the social group membership of hand models and observers did not modulate retrieval of observational SR bindings in Experiment 3, indicated by the absent three-way interaction. In fact, retrieval effects were of comparable magnitude for both, videotaped prime responses that were executed by white- as well as dark-skinned hands. This implies that observationally acquired SR bindings were retrieved irrespective of the similarity in terms of social group membership between one's own and videotaped responses. Indeed, on a descriptive level, retrieval effects even tended to be stronger for videos of out-group members (see Table 2). This insight is important, since results from the manipulation check showed that identification was higher for white- than for dark-skinned hands, indicating that the manipulation in and of itself was successful. Thus, although participants identified stronger with prime videos of white-skinned hands (representing in-group members) than with videos of dark-skinned hands (representing out-group members), this had no influence on retrieval of observationally acquired SR bindings in the video-based paradigm.

The absence of any modulatory influence (despite a successful manipulation) suggests that retrieval the video-based variant of the observational SR binding task is not as prone to the influence of social factors like group membership as the dyadic paradigm. However, this reasoning is based on a null-finding that can be due to other reasons. One possible alternative explanation for the absence of any social modulation is reactance in participants: Social group membership was operationalized by representatives of natural social groups rather than working with artificially created social groups (e.g. by giving hand models and observers gloves of different colors). Use of natural social groups adds an (unwanted) racial note to the experiment. Although speculative, it might be that after observing hand models with dark skin, participants tried to avoid behaving in a way that could be misinterpreted as discriminatory or prejudiced in nature. This could have motivated reactant behavior in participants, who then increased their attention to videos from dark-skinned hand models. This would explain the presence of SR retrieval for the videos of out-group members (see also General Discussion for the influence of attention on SR binding and retrieval effects).

Experiment 4 was thus designed to follow up on the idea whether differences in social group membership affect retrieval of observationally acquired SR bindings in the video-based paradigm, but without using natural social groups (like different skin colors). Instead, Experiment 4 used a minimal group paradigm manipulation (e.g., Tajfel & Turner, 2004) with artificial social groups. At the beginning of the experiment, participants saw pictures with dot clouds for a few seconds and were asked to estimate the number of dots. Participants then received (bogus) feedback of ostensibly belonging to the social group of "underestimators" or "overestimators", respectively. However, assignment of participants to both groups was arbitrary. Participants then read a description of typical members of their own group (in-group) and of typical members of the other group (out-group). To strengthen the manipulation of group membership, descriptions of the in-group were more favorable than descriptions of the out-group. Then, participants performed the video-based observational SR binding task. Prime videos either stemmed from an in-group versus an out-group member. Following the same reasoning as Experiment 3, videos from the in-group member should be perceived as more similar to one's own responses compared with videos from the out-group member. If this reasoning this correct, then group membership of hand models and observers should modulate retrieval of observational SR bindings. Retrieval of SR bindings should be more likely if responses belonged to in-group members, but should be reduced or even absent for responses that were observed in out-group members (reflected in a three-way interaction of stimulus relation, response compatibility, and group membership).

5. Experiment 4

5.1. Method

5.1.1. Participants

Fifty-two German native speaking participants took part in the experiment. Applying the same criteria as in Experiment 1, two participants had to be excluded due to excessive error rates. Ten additional participants did not correctly identify their own and/or the video players' group memberships in a manipulation check. Thus, data of 40 participants were analyzed (31 female, $M_{\text{age}} = 22.4$ years, range: 18–32 years). Participants were tested individually and received partial course credit for participating. The experiment lasted 40 min.

5.1.2. Material, design, and procedure

Material, design, and procedure paralleled Experiment 2 except for the following changes: At the beginning of the experiment, participants were asked to work through a computerized "personality test". To this end, they saw seven different pictures with dot clouds. Every picture was presented for 5 s. Participants were asked to estimate the number of dots. After the last picture, participants then received (bogus) feedback on their personality and were told to belong to the social group of "underestimators" or "overestimators", respectively. However, assignment of participants to both groups was arbitrary and based on subject number. Participants then read a description of typical members of their own group (in-group) and of typical members of the other group (out-group). To strengthen the manipulation of group membership, descriptions of the in-group were more favorable than descriptions of the out-group. Participants were told that in the oncoming task, they would see videotaped responses from other persons (referred to as "interaction partners") who also answered the "personality test" and thus either belonged to their own versus the other group. To allow for easy identification of group membership, two arrow symbols (four arrows arranged in the shape of a plus sign) were introduced that indicated the group membership of the person carrying out the videotaped responses (arrows pointing inwards represented "underestimators"; arrows pointing outwards represented "overestimators"). Thus, the videos from Experiment 2 were replaced by four new videos (2 group membership × 2 color responses) that depicted red or green key presses either from an in-group or out-group member (see Fig. 2; assignment of video model to in- or out-group was counterbalanced). To be as similar to the dyadic variant of the observational SR binding task as possible, all videos adopted a third-person perspective and referred to "interaction partners" whenever videotaped responses were mentioned in the instructions. Instructions were made more comparable to the dyadic variant of the observational SR binding task: Participants were instructed to imagine that they would perform each experimental block with another person and that the color-categorization task would be shared between the two of them. Participants were further instructed that whenever they saw a colored word stimulus, it was their turn to give a color categorization response. However, when the word stimulus was presented in white font, it was their "interaction partner's" turn to categorize the word color. In that case, each participant had the task to observe the response carried out by their "interaction partner". Group membership of videotaped prime responses varied blockwise within subjects (with 160 prime-probe sequences for each group member). Block order (in- vs. out-group member first) was counterbalanced between subjects. Preliminary analyses showed that block order did not modulate the effects of interest; hence, all analyses were performed with this factor collapsed.

As a manipulation check, participants answered the following questions in a forced-choice questionnaire: (1) Which personality type are you? (2) Which personality type had your interaction partner in the first block? (3) Which personality type had your interaction partner in the second block? Participants gave their answers via mouse click on one of the following response buttons presented on screen: "underestimator", "overestimator", "can't remember". Participants who did not correctly identify their own as well as their interaction partners' personality type were excluded from all analyses. Like in Experiment 3, the identification with videotaped prime responses was assessed as another manipulation check after each block (i.e., for videos of each hand model). Unfortunately, answers were not saved due to a programming error.

After all data were collected, participants were debriefed about the aims of the experiment and were informed about the bogus personality test.

5.2. Results

5.2.1. Manipulation checks

5.2.1.1. Memory test for observed responses. To ensure that all probe actors paid sufficient attention to observed prime responses, average error scores for performance in the memory test were computed for each block. Importantly, error rates in the memory test did not differ between hand models of in- and outgroup ($M_{\text{IG}} = 3.50\%$; $M_{\text{OG}} = 3.38\%$; $|t| < 1$). This indicates that all participants adequately attended which response was executed in the prime videos, irrespective of the hand models' group membership.

5.2.2. Probe performance. For RT

Only correct probe responses were analyzed. According to the same criteria as in Experiment, 4.2% of probe responses were excluded: 1.7% because of errors, 3.4% because of erroneous responses in the memory test (0.9% of overall trials), and 1.7% because of RT outlier values. Means were then computed for the factorial design for RT and error rates (see Table 2) and were entered into a 2 (Stimulus relation: word repetition vs. word change) × 2 (Response compatibility: compatible vs. incompatible) × 2 (hand model's group membership: in-group vs. out-group) ANOVA with repeated measures.

For RT, the ANOVA revealed main effects of response compatibility, $F(1,39) = 7.68$, $p = .009$, $\eta_p^2 = 0.16$, and group membership, $F(1,39) = 6.94$, $p = .012$, $\eta_p^2 = 0.15$. Probe responses that were compatible with observed prime responses were executed faster ($M = 416$ ms) than incompatible probe responses ($M = 427$ ms); also, probe responses were faster when observing responses of the in-group member ($M = 412$ ms) compared with the out-group member ($M = 431$ ms). Furthermore, the Stimulus Relation × Response Compatibility interaction was significant, $F(1,39) = 10.13$, $p = .003$, $\eta_p^2 = 0.21$, reflecting retrieval of bindings between word stimuli and observed prime responses. Follow-up tests showed that compared to word stimulus changes in the probe, word stimulus repetition produced a significant performance benefit ($\Delta = 7$ ms; $t[39] = 2.61$, $p = .007$ (one-tailed), $d_z = 0.41$) for probe responses that were compatible with observed prime responses. In turn, compared to word stimulus changes in the probe, word stimulus repetition produced a significant performance cost ($\Delta = -5$ ms; $t[39] = 2.33$, $p = .013$ (one-tailed), $d_z = 0.37$) for probe responses that were incompatible with observed prime responses. Importantly, the three-way interaction was not significant, $F < 1$, $BF_{01} = 9.202$. Net S × R interaction effects were significant for both, the in-group ($S \times R_{IG} = 10$ ms, $t[39] = 1.91$, $p = .032$ (one-tailed, since this effect was predicted), $d_z = 0.30$) as well as the out-group ($S \times R_{IG} = 14$ ms, $t[39] = 2.56$, $p = .014$, $d_z = 0.40$). No other effects were significant (all $Fs < 1$).

For error rates, ANOVA revealed only a main effect of group membership, $F(1,39) = 5.03$, $p = .031$, $\eta_p^2 = 0.11$, reflecting fewer probe errors after observing an outgroup ($M = 1.35\%$) compared with an ingroup member ($M = 1.88\%$). No other effect was significant (all $Fs < 2.4$; all $ps > 0.13$).

5.3. Discussion

The results of Experiment 4 are clear-cut: Social group membership of hand did not modulate retrieval of observational SR bindings, given that the three-way interaction was absent. Once again, retrieval effects were of comparable magnitude for responses observed in in-group as well as in out-group members. This implies that observationally acquired SR bindings were retrieved irrespective of the social group membership of the observed hand model. Together with findings from Experiment 3, this argues against a strong modulatory influence of social factors like group memembership on retrieval of observational SR bindings in the video-based paradigm.

So far, retrieval of observational SR bindings in the video-based paradigm variant was immune to manipulations of visual perspective (Exp. 2) and to "standard" manipulations of social group membership (Exp.'s 3 and 4). This is add odds with findings from the dyadic variant of the observational SR binding task, where retrieval effects are socially moderated and vary as a function of social relevance (Giesen et al., 2014, 2016, 2018). If social relevance of hand models was also an important factor within the video-based observational SR binding task, then retrieval effects should have been absent for responses observed in out-group members (Exp.'s 3 and 4), for responses observed from a third-person perspective (Exp. 2), and Experiment 1 as a whole (since hand models were not socially relevant).

Tentatively, the present findings suggest that other factors will impact on attentional processes that underlie retrieval effects in both paradigms. To follow-up on this reasoning, Experiment 5 was designed to test for a simple cognitive alternative explanation, which argues that observational SR binding and retrieval in the video-based paradigm is a function of attention that promotes binding and retrieval even in the absence of social relevance of hand models. Going even one step further, it could be that observational SR binding and retrieval in the video-based paradigm might even work without any reference to "social" agents (e.g., hand models) at all. According to this view, it is sufficient to draw spatial attention to a response key for creating a binding between the highlighted response key's color (i.e., the target feature) and the word stimulus presented in the prime display. Stimulus repetition in the subsequent probe display will then retrieve this stimulus feature binding, which facilitates or hampers probe performance, depending on whether the retrieved feature matches or mismatches in the probe (Giesen & Rothermund, 2014). Indeed, there is evidence that irrelevant but attended features are more likely to become integrated into SR episodes or more likely to retrieve SR episodes upon repetition (Singh et al., 2018; but see B. Hommel, 2005; see Henson et al., 2014, for a discussion).

To test this non-social alternative account, prime videos were replaced by still pictures of a response pad (viewed from the third-person perspective, see Fig. 2). Pictures either showed two hands placed on the response pad (social reference) or simply depicted an empty response pad (no social reference). Instead of videotaped responses, either the red or green key was cued with a yellow, blinking square during the prime display. Accordingly, if observational SR binding and retrieval is a function of (spatial) attention due to stimulus-based retrieval of feature bindings, Experiment 5 should yield robust retrieval effects even when the keys of a response pad without hands are highlighted (i.e., in the absence of any reference to social agents). If, however, retrieval of observationally acquired bindings requires at least a minimal reference to a social agent performing a response, retrieval effects should emerge only for pictures that include human hands, but should be reduced or absent for pictures without social reference (indicated by a significant three-way interaction of stimulus relation, response compatibility, and picture type).

6. Experiment 5

6.1. Method

6.1.1. Participants

Fifty-four German native speaking participants took part in the experiment. Applying the same criteria as in Experiment 1, no participants had to be excluded. Thus, data of all participants were analyzed (39 female, $M_{age} = 21.7$ years, range: 19–29 years). Participants were tested individually and received partial course credit for participating. The experiment lasted 30 min.

6.1.2. Material, design, and procedure

Material, design, and procedure paralleled Experiment 2 except for the following changes: The videos from Experiment 2 were replaced by four still pictures (2 picture type × 2 color responses) that depicted a response pad either with or without human hands placed on the red and green response key (see Fig. 2). All pictures adopted a third-person perspective. Trial procedure changed as follows: A given trial started with a brief fixation (250 ms), followed by the prime display, consisting of a word stimulus in white font that was presented simultaneously with a still picture of a response pad. After 200 ms, a yellow frame (blinking three times) appeared, highlighting either the red or green response key on the picture (prime display's overall duration: 800 ms). After another fixation (250 ms), the probe display followed (similar to Experiment 1, see above). Picture type varied blockwise within subjects (with 160 prime-probe sequences for each picture type). Block order (picture with or without hands first) was counterbalanced between subjects. Preliminary analyses showed that block order did not modulate the effects of interest; hence, all analyses were performed with this factor collapsed.

Similar to Experiment 1, a memory test appeared after 25% of all probe displays, asking participants which response key was highlighted in the preceding prime display. Participants gave their response by pressing the corresponding key on their response pad.

6.2. Results

6.2.1. Manipulation checks

6.2.1.1. Memory test for highlighted prime response keys.

To ensure that all probe actors paid sufficient attention to prime displays, average error scores for performance in the memory test were computed for each block. Importantly, error rates in the memory test did not differ between pictures with or without hands ($M_{hand} = 6.9\%$; $M_{w/o\ hand} = 6.5\%$; $|t| <$

1). This indicates that all participants adequately attended which response key was highlighted in the prime display, irrespective of the picture type.

6.2.2. Probe performance. For RT

Only correct probe responses were analyzed. According to the same criteria as in Experiment 1, 5.5% of probe responses were excluded: 3.0% because of errors, 6.7% because of erroneous responses in the memory test (1.7% of overall trials), and 1.5% because of RT outlier values. Means were then computed for the factorial design for RT and error rates (see Table 2) and were entered into a 2 (Stimulus relation: word repetition vs. word change) × 2 (Response compatibility: compatible vs. incompatible) × 2 (picture type: social reference vs. no social reference) ANOVA with repeated measures.

For RT, the ANOVA revealed only a main effect of response compatibility, $F(1,53) = 5.44$, $p = .024$, $\eta_p^2 = 0.09$. Probe responses that were compatible with highlighted response keys were executed faster (M = 420 ms) than incompatible probe responses (M = 430 ms). Furthermore, the Stimulus Relation × Response Compatibility interaction was significant, $F(1,53) = 10.89$, $p = .002$, $\eta_p^2 = 0.17$, reflecting retrieval of bindings between word stimuli and highlighted color of prime response keys. Follow-up tests showed that compared to word stimulus changes in the probe, word stimulus repetition produced a significant performance benefit (RT: $\Delta = 11$ ms; $t[53] = 3.82$, $p < .001$ (one-tailed), $d_z = 0.52$) for probe responses that were compatible with highlighted prime response keys. In turn, compared to word stimulus changes in the probe, word stimulus repetition produced a nonsignificant performance cost (RT: $\Delta = -5$ ms; $t[53] = 1.43$, $p = .078$ (one-tailed), $d_z = 0.19$) for probe responses that were incompatible with highlighted prime response keys. Importantly, the three-way interaction was not significant ($F < 1$), $BF_{01} = 6.338$. Net S × R interaction effects were significant for both, prime picture with reference to social agents ($S \times R_{hand} = 16$ ms, $t[53] = 2.59$, $p = .006$ (one-tailed, since this effect was predicted), $d_z = 0.35$) as well as for the prime picture with no social reference ($S \times R_{w/o\ hand} = 16$ ms, $t[39] = 2.16$, $p = .035$, $d_z = 0.29$). No other effects were significant (all $Fs < 2.8$, all $ps > 0.10$).

The same ANOVA on error rates revealed only the Stimulus Relation × Response Compatibility interaction as significant, $F(1,53) = 4.47$, $p = .039$, $\eta_p^2 = 0.08$. Follow-up tests showed that compared to word stimulus changes in the probe, word stimulus repetition produced a nonsignificant performance benefit (i.e., less errors, $\Delta = 0.8\%$; $t[53] = 1.52$, $p = .068$ (one-tailed), $d_z = 0.21$) for probe responses that were compatible with highlighted prime response keys. In turn, compared to word stimulus changes in the probe, word stimulus repetition produced a significant performance cost (i.e., more errors, $\Delta = -0.9\%$; $t[53] = 1.75$, $p = .043$ (one-tailed), $d_z = 0.24$) for probe responses that were incompatible with highlighted prime response keys. No other effects were significant (all $Fs < 2.9$, all $ps > 0.09$).

6.3. Discussion

Experiment 5 yield clear-cut results: Stimulus-based retrieval of observationally acquired feature bindings was apparent after participants observed highlighting of response keys on a still picture during the prime display. Importantly, these effects emerged regardless of whether pictures included a reference to social agents (i.e., showed a response pad with human hands) or not (i.e., showed an empty response pad). Finding robust retrieval effects in the latter condition is particularly striking and supports the non-social explanation for retrieval effects in the video-based paradigm. In this respect, one may understand the observation of a videotaped prime response as another means of drawing attention to a particular response key, thereby increasing the likelihood of integrating the response key's main defining feature – it's color – with the word stimulus (Moeller & Frings, 2014; Singh et al., 2018). Importantly, these feature bindings can be accessed from memory via stimulus repetition. Stimulus-based retrieval will then either promote or impede performance, depending on whether the retrieved feature matches or mismatches with the current task requirements (cf. Giesen & Rothermund, 2014).

7. General discussion

Evidence from Experiment 1 showed that observational SR binding and retrieval effects can also be acquired by observing video-taped responses instead of observing responses executed by a "real" interaction partner. Notably, retrieval of these bindings occurs even in the absence of any information about the person executing the videotaped responses. Against this background, the present paper sought to unravel potential differences between the dyadic and the video-based variant of the observational SR binding task that might explain why retrieval effects vary as a function of social relevance in the dyadic paradigm (Giesen et al., 2014, 2016, 2018), but are absent in the video-based paradigm. In the dyadic observational SR binding task, interaction partners are sitting opposite each other, which implies that (a) responses are observed from a third-person perspective and (b) it is obvious that observed responses were executed by another person. Thus, the dyadic paradigm differs in meaningful ways from the video-based paradigm variant tested in Experiment 1. Because of this, several manipulations aimed to make the video-based paradigm more similar to the dyadic paradigm to eradicate subtle differences between both paradigm variants. One experiment endorsed prime videos with both first- and third-person visual perspective (Exp. 2). Two additional experiments highlighted that videotaped responses were executed by another (socially relevant) person by manipulating social group membership of hand models (Exp.'s 3 and 4). If effects in the video-based observational SR binding task would follow the same principles as in the dyadic paradigm, then retrieval effects should have been absent for (a) responses observed from a third-person perspective (Exp. 2) and for (b) responses observed in out-group members who are distinctly different from oneself (Exp.'s 3 and 4). However, results of Experiment 2 showed that retrieval of observational SR bindings emerged to similar degree for prime videos that endorsed a first- as well as third-person perspective. Furthermore, in Experiments 3 and 4, retrieval effects of observational SR bindings emerged to similar degree for prime responses that were executed by an in-group as well as an out-group member. Together, Experiments 2 to 4 provided no evidence for a social modulation of retrieval effects in the video-based observational SR binding task.

The conclusion that socially modulated effects are absent in the video-based variant of the observational SR binding task rests on the interpretation of the nonsignificant three-way interactions in Experiments 2 to 4. To bolster this conclusion – which is based on a series of null findings – *post-hoc* power analyses as well as Bayesian statistics were performed. *Post-hoc* power computations with G*Power showed that achieved power ($1\text{-}\beta$) to detect an effect of $d = 0.5$ in Experiments 2 to 4 ranged between 0.93 and 0.97. It is thus save to say that the present experiments were sufficiently strong to detect differences in the strength of observational SR binding and retrieval effects between the manipulated visual perspective and group membership conditions that were of at least medium size. Note also that in terms of the actual data pattern, SR binding and retrieval effects were either *equal* to or *larger* (but never smaller or absent) in the third-person perspective and for outgroup members than in the respective comparison condition (see Table 2), which argues against the alternative hypothesis. The absence of any social modulatory effects in the video-based paradigm is further supported by Bayesian statistics, showing that the null-hypothesis (i.e., the absence of a three-way interaction) was seven to fourteen times more likely than the alternative hypothesis in Experiments 2 to 4, which represents strong to decisive evidence in favor of the null hypothesis (Kass & Raftery, 1995).

As such, the null-findings of Experiments 2 to 4 are at odds with findings from the dyadic variant of the observational SR binding task, where retrieval effects are socially moderated and vary as a function of social relevance (Giesen et al., 2014, 2016, 2018). Findings from Experiment 5 show that it is sufficient to draw attention towards a response key press for creating a binding between the highlighted response key's color feature and the word stimulus presented in the prime display. Stimulus repetition in the subsequent probe display will then retrieve this stimulus feature binding, which is reflected in performance advantages (disadvantages) if the retrieved feature binding matches (mismatches) in the probe. According to Giesen and Rothermund (2014), stimulus feature bindings and stimulus-response bindings (which reflect qualitatively different sorts of bindings; see also Hommel, 2007) emerge independently of each other and are also retrieved independently from each other. Thus, although the absence of any observable prime response in Experiment 5 prevented the emergence of stimulus-response bindings, the emergence of stimulus feature bindings was unaffected (see Experiment 2 in Giesen & Rothermund, 2014, for comparable findings).

According to this non-social explanation of retrieval effects in the observational SR binding task, one may understand the observation of a videotaped prime response as another means of "highlighting" a particular response key, meaning that participants' attention is allocated to the response key that was pressed by the video-agent. This non-social attention explanation fits well to all SR binding effects observed in Experiments 1–5: That is because for drawing attention to a response key, it is irrelevant whether the cues drawing the attention are yellow squares, white or black hands observed from first or third-person perspective, and so on.

7.1. Theoretical implications

The present findings are noteworthy in several respects. First, it appears that stimulus-based retrieval effects acquired from video observation most likely reflect attention induced bindings between word stimuli and color features (assigned to response keys). Retrieval of these stimulus feature bindings, in turn, reflects a process that is immune to social modulations as observed in the dyadic variant of the observational SR binding task. Retrieval of stimulus feature bindings can thus explain why we obtained robust retrieval effects even under conditions in which retrieval of observationally acquired SR bindings should be absent.

Second, it is unresolved to which extend stimulus-feature binding effects in the dyadic variant of the observational SR binding task and the video-based variant of the observational SR binding task reflect the same underlying processes. Given that both effects were modulated by social variables in a different fashion one could argue that the underlying processes are different. From a theoretical perspective, this is plausible, as numerous findings document that different binding types are stored and retrieved simultaneously and independently of each other (Giesen & Rothermund, 2014; Hommel, 1998, 2007; B. Hommel et al., 2001).

Yet, third, it is also theoretically possible and more parsimonious that the very same cognitive process (i.e., attention) might be underlying both task variants. However, both tasks might differ in their responsiveness to particular moderating conditions, which will then affect attention allocation. For instance, whereas social relevance appears to be a key moderator for the occurrence of observational SR bindings in the dyadic task variant, spatial cueing of attention is the dominant moderator for the occurrence of stimulus feature bindings in the video-based task. The fact that stimulus feature bindings were virtually absent when real human interaction partners were there but irrelevant for participants (cf. Giesen et al., 2014, 2018) might hint again at an attention modulation. The mere presence of a human will typically allocate attention towards the other human (as compared to a video at least; see also Longo & Bertenthal, 2009). In a situation in which one competes or cooperates with the other human, this automatic allocation of attention to the other human is actually helpful (and binding effects emerge as usual).Yet, if the other human is actually irrevelant for the task, it is a strong attention-grabbing distractor that has to be ignored, meaning that attention is actively drawn away from the other human. This might explain why stimulus-feature bindings were absent in a dyadic set-up where participants independently responded in alternating fashion. Future research is needed to investigate whether this overarching explanation in terms of attention can reconcile the findings from the dyadic variant of the observational SR binding task and the video-based variant of the observational SR binding task.

7.2. Limitations

Our set of experiments comes with some caveats, though. First, we have to admit that a direct comparison of both task variants is compromised, since no direct replication of findings from the dyadic task was tested in the video-based task, and vice-versa. In this vein, the present series of studies shows that even subtle differences in the experimental set-up of both paradigm variants might have a strong impact on the results. Thus, an exact replication that for instance endorses the "situationally induced" interdependence manipulation from the dyadic task (Giesen et al., 2014) would be an important future contribution. Nevertheless, we want to point out that two experiments (Exp. 3 & 4) endorsed manipulations of social relevance by using hand models of the same vs. different social group. The importance of social identity and our need-to-belong to different social groups is an established and undisputable finding (cf. Tajfel & Turner, 2004). In this respect, the various manipulations that were tested in the video-based task variant of the observational SR binding task are very similar on a conceptual level to the manipulations that were tested in the dyadic task variant. Thus, comparing results from both tasks is still legitimate.

Second, we cannot rule out that presenting both the response and the distractor on screen in the video-based task variant resulted in a different experience compared with the dyadic task variant. One could argue that "response observation" in the latter task is a multisensory experience: Responses are not presented focally on screen but are visible in the periphery (to the left and right from the screen). Additionally, because participants have to permanently press down both release keys with their left and right hands, they will feel (and also hear) whenever their interaction partner hits one of the response buttons. This renders "response observation" as a holistic experience that combines visual, auditory and tactile impressions. Future research is needed to address whether differences in the "visual-only" vs. "holistic" response observation experience is the crucial feature in accounting for the obtained differences between both task variants.

7.3. Practical implications

The present findings illustrate that two apparently similar variants of the same task may differ in their responsiveness to moderating factors. Tentatively, it has to be explored whether they also rely on the same or different underlying processes. This insight is important also for other socio-cognitive tasks that aim to investigate imitative or joint action phenomena. For instance, one may ask whether imitation effects that stem from copying a real person (e.g., Chartrand & Bargh, 1999) derive from the same underlying process as effects captured with the imitation-interference task (Brass et al., 2001) as a result of picture observation. The same question can be asked for joint Simon effects that result from engaging in a shared task with a real person (e.g., Sebanz et al., 2003) or from "interacting" with a video of a human or wooden hand (e.g., Tsai & Brass, 2007). Future research is needed to address these questions in more detail.

On a more general level, the present findings are relevant also for more applied settings. Tentatively, one may understand the present findings as evidence that video techniques (e.g. video conferences, video-based counselling or even psychotherapy) might involve different effects than live interactions and are thus only a poor substitute for true face-to-face interactions.

8. Conclusion

The present set of studies documents that the video-based variant of the observational SR binding task is not (yet) suited to measure genuine "social" effects that are responsive to appropriate manipulations of social factors. This renders the video-based observational SR binding task inappropriate for the study of observational learning and joint action phenomena, despite its efficient and appealing set-up. At present, it appears that the dyadic variant of the observational SR binding task is a more suitable candidate to study the socio-cognitive micro-processes underlying observational learning or related imitative tasks.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author statement

Carina G. Giesen, Friedrich-Schiller-Universität Jena, Jena, Germany and Christian Frings, Trier University, Trier, Germany.

Appendix 1. List of German word stimuli used in Experiments 1 to 5 (English translation in parentheses)

glatt (smooth)	kariert (checked)	warm (warm)	offen (open)	klar (clear)
eben (even)	kurz (short)	niedrig (low)	neu (new)	trocken (dry)
eckig (angular)	scharf (sharp)	unruhig (fidgety)	breit (broad)	steil (steep)
bunt (colourful)	dunkel (dark)	leise (quiet)	spitz (pointed)	dick (thick)
klein (small)	leicht (light)	langsam (slow)	weich (soft)	hoch (high)

References

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory. Prentice-Hall series in social learning theory*. Englewood Cliffs, NJ US: Prentice-Hall, Inc.

Bortoletto, M., Baker, K. S., Mattingley, J. B., & Cunnington, R. (2013). Visual–motor interactions during action observation are shaped by cognitive context. *Journal of Cognitive Neuroscience, 25*(11), 1794–1806. https://doi.org/10.1162/jocn_a_00431

Brass, M., Bekkering, H., & Prinz, W. (2001). Movement observation affects movement execution in a simple response task. *Acta Psychologica, 106*(1–2), 3–22. https://doi.org/10.1016/S0001-6918(00)00024-X

Chartrand, T. L., & Bargh, J. A. (1999). The chameleon effect: The perception–behavior link and social interaction. *Journal of Personality and Social Psychology, 76*(6), 893–910. https://doi.org/10.1037/0022-3514.76.6.893

Dolk, T., Hommel, B., Colzato, L. S., Schütz-Bosbach, S., Prinz, W., & Liepelt, R. (2014). The joint Simon effect: A review and theoretical integration. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00974

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/BF03193146

Frings, C., Hommel, B., Koch, I., Rothermund, K., Dignath, D., Giesen, C., … Philipp, A. (2020). Binding and retrieval in action control (BRAC). *Trends in Cognitive Sciences, 24*(5), 375–387. https://doi.org/10.1016/j.tics.2020.02.004

Genschow, O., Florack, A., & Wänke, M. (2013). The power of movement: Evidence for context-independent movement imitation. *Journal of Experimental Psychology: General, 142*(3), 763–773. https://doi.org/10.1037/a0029795

Genschow, O., & Schindler, S. (2016). The influence of group membership on cross-contextual imitation. *Psychonomic Bulletin & Review, 23*(4), 1257–1265. https://doi.org/10.3758/s13423-015-0983-4

Genschow, O., van Den Bossche, S., Cracco, E., Bardi, L., Rigoni, D., & Brass, M. (2017). Mimicry and automatic imitation are not correlated. *PLoS One, 12*(9). https://doi.org/10.1037/t51181-000

Giesen, C. G., & Genschow, O. (2020). *What if I say I'm (not) like the other? Perceived (dis-) similarity between co-actors mediates retrieval of observationally acquired stimulus-response bindings*. Manuscript in preparation.

Giesen, C., Herrmann, J., & Rothermund, K. (2014). Copying competitors? Interdependency modulates stimulus-based retrieval of observed responses. *Journal of Experimental Psychology: Human Perception and Performance, 40*(5), 1978–1991. https://doi.org/10.1037/a0037614

Giesen, C., Löhl, V., Rothermund, K., & Koranyi, N. (2018). Intimacy effects on action regulation: Retrieval of observationally acquired stimulus-response bindings in romantically involved interaction partners versus strangers. *Frontiers in Psychology, 9*, 1369. https://doi.org/10.3389/fpsyg.2018.01369

Giesen, C., & Rothermund, K. (2014). Distractor repetitions retrieve previous responses and previous targets: Experimental dissociations of distractor–response and distractor–target bindings. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 40*(3), 645–659. https://doi.org/10.1037/a0035278

Giesen, C., Scherdin, K., & Rothermund, K. (2016). Flexible goal imitation: Vicarious feedback influences stimulus-response binding by observation. *Learning & Behavior, 45*(2), 147–156. https://doi.org/10.3758/s13420-016-0250-1

Henson, R. N., Eckstein, D., Waszak, F., Frings, C., & Horner, A. J. (2014). Stimulus–response bindings in priming. *Trends in Cognitive Sciences, 18*(7), 376–384. https://doi.org/10.1016/j.tics.2014.03.004

Hommel, B. (1998). Event files: Evidence for automatic integration of stimulus-response episodes. *Visual Cognition, 5*(1), 183–216. https://doi.org/10.1080/713756773

Hommel, B. (2005). How much attention does an event file need? *Journal of Experimental Psychology: Human Perception and Performance, 31*(5), 1067–1082. https://doi.org/10.1037/0096-1523.31.5.1067

Hommel, B. (2007). Feature integration across perception and action: Event files affect response choice. *Psychological Research/Psychologische Forschung, 71*(1), 42–63. https://doi.org/10.1007/s00426-005-0035-1

Hommel, B., Colzato, L. S., & van den Wildenberg, W. P. M. (2009). How social are task representations? *Psychological Science, 20*(7), 794–798. https://doi.org/10.1111/j.1467-9280.2009.02367.x

Hommel, B., Müsseler, J., Aschersleben, G., & Prinz, W. (2001). The Theory of Event Coding (TEC): A framework for perception and action planning. *Behavioral and Brain Sciences, 24*(5), 849–937. https://doi.org/10.1017/S0140525X01000103

Iani, C., Anelli, F., Nicoletti, R., Arcuri, L., & Rubichi, S. (2011). The role of group membership on the modulation of joint action. *Experimental Brain Research, 211* (3–4), 439–445. https://doi.org/10.1007/s00221-011-2651-x

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795.

Kokal, I., Gazzola, V., & Keysers, C. (2009). Acting together in and beyond the mirror neuron system. *NeuroImage, 47*(4), 2046–2056. https://doi.org/10.1016/j.neuroimage.2009.06.010

Lakin, J. L., Chartrand, T. L., & Arkin, R. M. (2008). I am too just like you: Nonconscious mimicry as an automatic behavioral response to social exclusion. *Psychological Science, 19*(8), 816–822. https://doi.org/10.1111/j.1467-9280.2008.02162.x

Lamm, C., Batson, C. D., & Decety, J. (2007). The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *Journal of Cognitive Neuroscience, 19*(1), 42–58. https://doi.org/10.1162/jocn.2007.19.1.42

Liepelt, R., Prinz, W., & Brass, M. (2010). When do we simulate non-human agents? Dissociating communicative and non-communicative actions. *Cognition, 115*(3), 426–434. https://doi.org/10.1016/j.cognition.2010.03.003

Logan, G. D. (1988). Toward an instance theory of automatization. *Psychological Review, 95*(4), 492–527. https://doi.org/10.1037/0033-295X.95.4.492

Longo, M. R., & Bertenthal, B. I. (2009). Attention modulates the specificity of automatic imitation to human actors. *Experimental Brain Research, 192*(4), 739–744. https://doi.org/10.1007/s00221-008-1649-5

Moeller, B., & Frings, C. (2014). Attention meets binding: Only attended distractors are used for the retrieval of event files. *Attention, Perception, & Psychophysics, 76*(4), 959–978. https://doi.org/10.3758/s13414-014-0648-9

Paulus, M., van Dam, W., Hunnius, S., Lindemann, O., & Bekkering, H. (2011). Action-effect binding by observational learning. *Psychonomic Bulletin & Review, 18*, 1022–1028. https://doi.org/10.3758/s13423-011-0136-3

Rothermund, K., Wentura, D., & De Houwer, J. (2005). Retrieval of incidental stimulus-response associations as a source of negative priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*(3), 482–495. https://doi.org/10.1037/0278-7393.31.3.482

Ruys, K. I., & Aarts, H. (2010). When competition merges people's behavior: Interdependency activates shared action representations. *Journal of Experimental Social Psychology, 46*(6), 1130–1133. https://doi.org/10.1016/j.jesp.2010.05.016

Sebanz, N., Knoblich, G., & Prinz, W. (2003). Representing others' actions: Just like one's own? *Cognition, 88*(3), 11–21. https://doi.org/10.1016/S0010-0277(03)00043-X

Sellaro, R., Treccani, B., Rubichi, S., & Cubelli, R. (2013). When co-action eliminates the Simon effect: Disentangling the impact of co-actor's presence and task sharing on joint-task performance. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00844

Singh, T., Moeller, B., Koch, I., & Frings, C. (2018). May I have your attention please: Binding of attended but response-irrelevant features. *Attention, Perception, & Psychophysics, 80*(5), 1143–1156. https://doi.org/10.3758/s13414-018-1498-7

Tajfel, H., & Turner, J. C. (2004). The social identity theory of intergroup behavior. In J. T. Jost, & J. Sidanius (Eds.), *Key readings in social psychology. Political psychology: Key readings* (pp. 276–293). New York, NY: Psychology Press. https://doi.org/10.4324/9780203505984-16.

Tsai, C.-C., & Brass, M. (2007). Does the human motor system simulate Pinocchio's actions? Coacting with a human hand versus a wooden hand in a dyadic interaction. *Psychological Science, 18*(12), 1058–1062. https://doi.org/10.1111/j.1467-9280.2007.02025.x

Tukey, J. W. (1977). *Exploratory data analysis*. Reading, MA: Addison-Wesley.

Van Baaren, R. B., Jansson, L., Chartrand, T. L., & Dijksterhuis, A. (2009). Where is the love? The social aspects of mimicry. *Philosophical Transitions of the Royal Society of London: B Biological Sciences, 364*, 2381–2389.

Vogt, S., Taylor, P., & Hopkins, B. (2003). Visuomotor priming by pictures of hand postures: Perspective matters. *Neuropsychologia, 41*(8), 941–951. https://doi.org/10.1016/S0028-3932(02)00319-6

Welsh, T. N., Higgins, L., Ray, M., & Weeks, D. J. (2007). Seeing vs believing: Is believing sufficient to activate the processes of response co-representation? *Human Movement Science, 26*(6), 853–866. https://doi.org/10.1016/j.humov.2007.06.003

Weyers, P., Mühlberger, A., Kund, A., Hess, U., & Pauli, P. (2009). Modulation of facial reactions to avatar emotional faces by nonconscious competition priming. *Psychophysiology, 46*(2), 328–335. https://doi.org/10.1111/j.1469-8986.2008.00771.x