Learn Behav (2018) 46:171–181
https://doi.org/10.3758/s13420-017-0299-5

Associative structure of second-order conditioning in humans

Paul Craddock[1,2] · **Jessica S. Wasserman**[3] · **Cody W. Polack**[3] · **Thierry Kosinski**[1] · **Charlotte Renaux**[1] · **Ralph R. Miller**[3]

Published online: 3 November 2017

Abstract Second-order conditioning (SOC; i.e., conditioned responding to S2 as a result of S1–US pairings followed by S2–S1 pairings) is generally explained by either a direct S2→US association or by an associative chain (i.e., S2→S1→US). Previous research found that differences in responses to S2 after S1 was extinguished often depended on the nature of the S2–S1 pairings (i.e., sequential or simultaneous). In two experiments with human participants, we examined the possibility that such differences result from S1 evoking S2 during extinction of S1 following simultaneous but not sequential S2–S1 pairings. This evocation of S2 by S1 following simultaneous pairings may have paired the evoked representation of S2 with absence of the outcome, thereby facilitating mediated extinction of S2. Using sequential S2-S1 pairings, both Experiments 1 and 2 failed to support this account of how extinction of S1 reduced responding to S2. Experiment 1 found that extinguishing S1 reduced responding to S2, while extinguishing S2 had little effect on responses to S1, although forward evocation of S1 during extinction of S2 paired the evoked representation of S1 with absence of the outcome. In Experiment 2, evocation of S2 during S1 nonreinforced trials was prevented because S2–S1 pairings *followed* (rather than proceeded) S1-alone exposures. Nevertheless, responding to S2 at test mimicked S1 responding. Responding to S2 was high in the context in which S1 had been reinforced and low in the context in which S1 had been nonreinforced. Collectively, these experiments provide additional support for the associative-chain account of SOC.

Keywords Second-order conditioning · Associative structure · Human conditioning · Conditioned discrimination task · Extinction

Following operational second-order conditioning (SOC; i.e., S1–US pairings followed by S2–S1 pairings, where S1 is a first-order conditioned stimulus, S2 a second-order conditioned stimulus, and US is an unconditioned stimulus), the second-order (S2) stimulus often acquires the potential to produce a conditioned response (CR) although it has never directly been paired with the US (Pavlov, 1927). Two alternative explanations of SOC in terms of associative linkage can be found in the literature (e.g., Barnet, Arnold, & Miller, 1991; Rizley & Rescorla, 1972). The first of these is that at test S2 elicits the CR through an *associative chain* that includes S1 (i.e., S2→S1→US→CR, or S2→S1→CR).[1] Both of these associative structures are variants of the associative-chain account, which assumes a role for the representation of S1 at test of S2 (e.g., Hall, 1996). The second type of account of SOC assumes that activation of the US representation by S1 during the Phase 2 S2→S1 pairings effectively creates direct pairings between S2 and the reactivated representation of the US,

✉ Paul Craddock
paul.craddock@univ-lille3.fr

[1] Université de Lille, Lille, France

[2] Université de Lille, Nord de France, Domaine universitaire du "Pont de Bois", rue du Barreau, BP 60149, 59653 Villeneuve d'Ascq Cedex, France

[3] State University of New York at Binghamton, Binghamton, NY, USA

[1] In both the associative-chain and the direct association accounts S2's capacity to elicit a CR might be all that is necessary at test, but it might be immediately preceded by activation of the representation of the US. We fully acknowledge this possibility, but discussion of this possibility here is distracting from the present concern regarding the role of S1. Whether activation of the US representation is involved is a parallel feature of both accounts. Our glossing over this issue for clarity should not be taken as an attempt to dismiss a potential role for the US representation.

resulting a *direct association* between S2 and the US (i.e., S2→US; Konorski, 1967). Thus, the direct association account of SOC differs from the associative-chain account in its omission of any role for the representation of S1 at the time at which S2 is tested.

The conventional test to differentiate between the associative-chain account and the direct association account of SOC involves assessment of the consequences of extinction of S1 (following the S2–S1 pairings) on responding to S2. A reduction in responding to S2 as a result of extinction of S1 is conventionally viewed as consistent with the associative-chain account of SOC in which presentation of S2 at test activates the representation of extinguished S1 (see Fig. 1). However, most published studies in which S2 and S1 were *sequentially presented* during S2–S1 pairing failed to report a decrease in responding to S2 following subsequent extinction of S1 (e.g., Rizley & Rescorla, 1972; but see Molet, Miguez, Cham, & Miller, 2012, which found evidence for the associative-chain view). Results similar to those of Rizley and Rescorla were observed in a human causal learning study (Jara, Vila, & Maldonado, 2006). Consequently, the prevailing view is that, given sequential S2–S1 pairings in Phase 2, responding to S2 depends on a direct association that was established during Phase 2 of SOC treatment between S2 and either the response evoked by the US or the representation of the US (i.e., an S2→CR or S2→US→CR associative sequence, or both as was suggested by Polack, Molet, Miguez, & Miller, 2013).

In contrast to the conventional absence in reduction of responding to S2 following post-SOC extinction of S1, Rescorla (1982) reported that when the S2–S1 pairings had been *simultaneous* during Phase 2, subsequent extinction of S1 did reduce responding to S2. He concluded that with simultaneous S2–S1 pairings, the representation of S1 played a role at test of S2. Thus, Rescorla concluded that simultaneous and sequential S2–S1 pairings resulted in different associative structures for the resultant SOC. Thus, Rescorla suggests that in Fig. 1a represents the structure underlying simultaneous SOC, whereas 2a represents the associative structure of sequential SOC.

There is, however, an alternative account of why, following simultaneous S2–S1 pairings, extinction of S1 decreases responding to S2. That is, with simultaneous S2–S1 pairings, S1 during Phase 2 may mediate formation of a direct S2–US association based on S2 being present and S1 activating a representation of the US with which it had been paired during Phase 1. Then, during Phase 3 (i.e., extinction of S1), presentation of S1 may evoke the representation of S2 in the absence of US, which could result in extinction of the evoked representation of S2 and reduce responding to S2 at test. Note that within this account of why extinguishing S1 reduces responding to S2 is the assumption that extinction of an invoked representation also reduces responding to the actual cue at test. We call this explanation of why extinction of S1 decreases responding to S2 the *mediated-extinction hypothesis*.

The mediated-extinction hypothesis is consistent with the direct association account of SOC in that the direct association view assumes that responding to S2 depends on a direct association between S2 and the evoked representation of the US. Furthermore, it contrasts with the associative-chain account of SOC, in which S2 at test activates a representation of S1 and consequently S1's current response potential. Extinction of the evoked representation of S2 may have occurred either in addition to or in the absence of S1 playing a contributing role in an associative chain at test of S2. That is, direct associations may contribute to SOC in parallel with an associative chain or be the sole basis of SOC.

Presumably, responding to S2 is not subject to appreciable extinction as a result of nonreinforced evocation of S2 in the case of prior sequential S2–S1 pairings because S1 does not strongly evoke the representation of S2 during S1's extinction. This is because, in the sequential case, S1 would have to activate the representation of S2 through a backward association. The possibility of an absent cue entering into the same type of association as its companion cue when the companion cue is presented alone was first proposed by Holland (1981, 1983; Holland & Forbes, 1982) and elaborated on by Hall (1996).[2]

The central point of what has been discussed above is that the presence or absence of decreased responding to S2 when S1 is extinguished does not necessarily reveal the associative structure underlying SOC (i.e., direct S2–US association or S2–S1–US associative chain). Some sequential SOC studies did observe decreased responding to S2 as a result of S1 being extinguished (e.g., Molet et al., 2012) while others did not (e.g., Rizley & Rescorla, 1972). In contrast, simultaneous S2–S1 pairings seem to consistently yield decreased responding to S2 (e.g., Rescorla, 1982). The presence or absence of decreased responding to S2 could be related to S1's potential during its extinction to activate a representation of S2 and affect direct extinction of the evoked representation of S2. The present experiments sought to investigate the potential role of mediated extinction in reducing responding to S2. Experiment 1 used sequential S2–S1 pairings with the expectation that extinction of S1 would not reduce responding to S2. If sequential SOC is less sensitive to post-SOC extinction of S1 because of the greater difficulty for S1 to backward evoke the representation of S2 during extinction, then extinction of

[2] It should be noted that Dickinson and Burke's (1996) modification of Wagner's (1981) SOP model also made predictions concerning the modification of the associations to an absent cue when its companion cue is extinguished. But their explanation cannot account for the decrease in responding to S2 when S1 is extinguished in the simultaneous condition (e.g., Rescorla, 1982) because their model assumes that retrieved representations of stimuli are activated into the A2 state (of the Dickinson & Burke model) and that representations evoked into the same state of activation establish an excitatory association. Therefore, because S2 and the US are evoked in the same state of activation (A2) when S1 is presented alone following simultaneous presentations of S2 and S1, extinction of S1 is predicted to increase rather than decrease responding to S2.

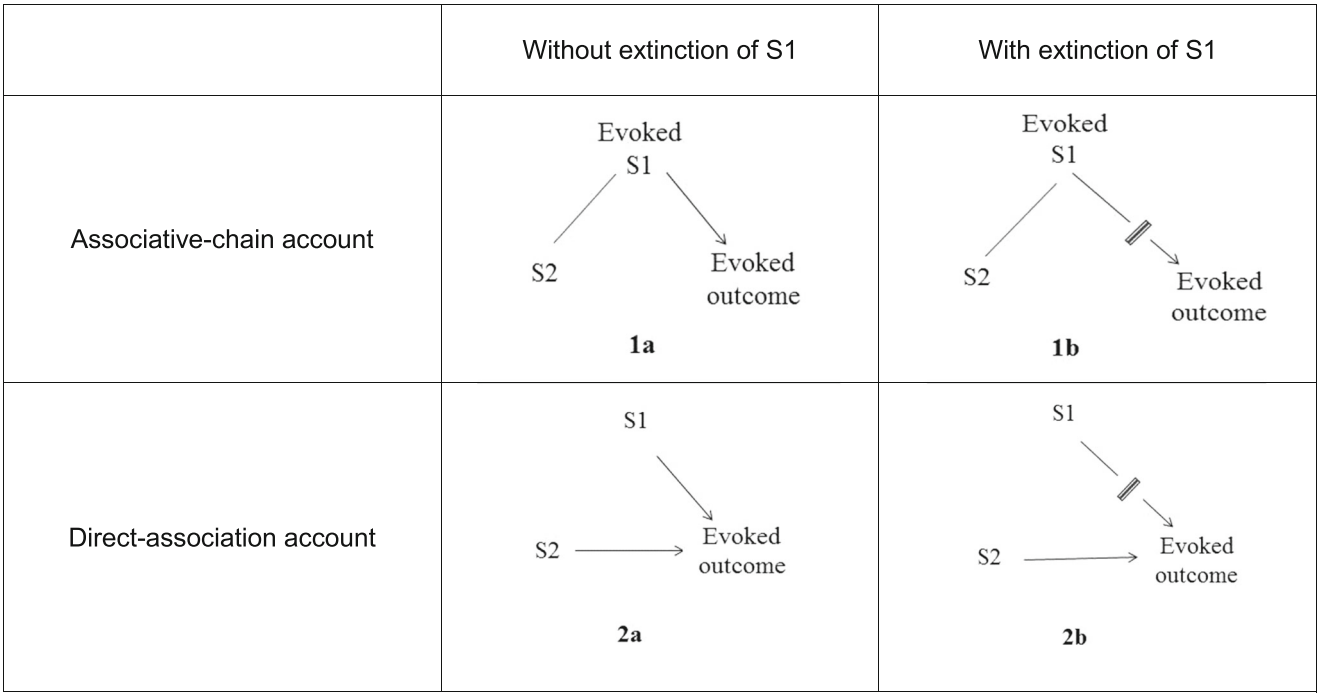

Fig. 1 Hypothetical associative structures underlying responding to S2 at test, following SOC (1a and 2a) and following post-SOC extinction of S1 (1b and 2b), according to the associative-chain account (1a and 1b) and the direct association account (2a and 2c), respectively

S2, which has a forward association with S1, should produce pronounced mediated extinction of S1. Hence, in the sequential case the mediated-extinction account anticipates extinction of the evoked representation of S1 (i.e., decreased responding of S1) as a result of extinguishing S2. However, little or no decrease of responding to S2 as a result of S1's extinction is expected if S1's backward association prevents effective mediated extinction of S2. Thus, the two theoretical accounts make different predictions with regard to responding to S1 when S2 is extinguished.

Experiment 1

Experiment 1 compared the influence of extinguishing S1 on responding to S2 (relative to its appropriate control condition) with the influence of extinguishing S2 on responding to S1 (relative to its appropriate control condition) in a SOC preparation in which sequential S2–S1 pairings occurred during Phase 2. Of central interest was whether the evoked representation of S1 (forward activated by S2 presentations during extinction of S2) along with the absence of the US would decrease responding to S1 (i.e., the first-order cue) as well as decrease responding to S2 (i.e., the extinguished second-order cue). If this did not occur, it would be highly improbable that extinction of S1 would result in pairings of the evoked representation of S2 (backward activated by S1 presentations during extinction of S1) with the absence of the US leading to extinction of S2.

Method

Participants

Twenty-four University of Lille (France) male and female students ($N = 24$), 18–29 years of age, volunteered.

Apparatus and materials

Individual cubicles were used to provide a quiet space for participants. The program was written using Inquisit. Seven white digits served as the S1 cues, and seven faces (512 × 383 pixels) portraying neutral affective states (Minear & Park, 2004) served as the S2 cues. Which digits and which faces served in each condition of the experiment were randomly determined for each participant. All cues were presented on a light gray background of a computer screen. The digits (Helvetica 150) were presented in the middle of the screen and the faces on the left side of the screen during the entire experiment. The US (hereafter referred to as an outcome, given its dubious biological significance) was the word *WIN* (Helvetica 80) in red letters that always appeared on the right side of the screen. All stimuli were presented equidistant from the top and bottom of the screen.

Procedure

Throughout the experiment, instructions were provided to the participants on the computer screen. Participants were first

asked to learn which S1s (i.e., digits) were followed by the WIN outcome and which were not. They were also instructed to predict the outcome on each trial by pressing on a "WIN" or an "empty screen" key (left and right cursor keys, counterbalanced). The S1s remained on the screen until participants responded. Each condition started with Phase 1 in which the seven digits were randomly assigned to the seven S1 roles ($S1_{SOC}$, $S1_{S1ext}$, $S1_{S2ext}$, $S1_{Filler1}$, $S1_{Filler2}$, $S1_{Filler3}$, $S1_{Novel}$; see Table 1). Participants were then presented with three successive blocks of the first six of these seven S1s ($S1_{Novel}$ was only presented at test). Each block contained one presentation of each of the six presented S1s with the order randomly determined anew for each block. Upon termination of each S1 presentation, that is, when the participant had made a prediction, there was a 1.5-s presentation of the WIN outcome on the right side of the screen following three of the S1s ($S1_{SOC}$, $S1_{S1ext}$, and $S1_{S2ext}$), while the three others (i.e., fillers: $S1_{Filler1}$, $S1_{Filler2}$, and $S1_{Filler3}$) were followed by the absence of the outcome (i.e., only the dark gray screen background was visible) for the same 1.5 s. The fillers were added to equate the number of reinforced cues and nonreinforced cues as well as to encourage discrimination learning as opposed to simply responding on all trials. The intertrial intervals (ITIs) were 2 s (so that the absent nonreinforcement outcome and the immediately subsequent ITI was simply a blank, 3.5-s, dark-gray screen).

Prior to Phase 2, participants were instructed on the screen that they would now see a face (i.e., S2) paired with each digit (i.e., S1), and that they should try to learn which face goes with which digit without making any prediction concerning the face that followed each digit. All participants were exposed once to each S2–S1 sequential pairing in random order. Each S2 appeared for 2 s on the left-hand side of the screen, and following termination of S2 its S1 associate was presented for 2 s in the center of the screen. After a 2-s ITI, the next S2 was presented.

Before Phase 3 began, all participants were instructed that they would have to learn new associations between a digit (i.e., S1) or a face (i.e., S2) and the outcome. Similar to Phase 1, they were instructed to predict, on each trial, what the outcome would be. Six within-subject conditions were constituted. In the SOC-S1 ext condition, operational extinction of $S1_{S1ext}$ occurred during Phase 3. In the SOC-S2 ext condition, operational extinction of $S2_{S2ext}$ occurred during Phase 3. In the SOC condition, there were no relevant trials during Phase 3. Of the three filler conditions in which S1 had been presented without an outcome in Phase 1, one of them (Filler 1) now had S1 paired with the outcome ($S1_{Filler1}$–outcome). A second condition (Filler 3) had S2 paired with the outcome ($S2_{Filler3}$–outcome). The remaining filler condition (Filler 2) had no new presentations of S1 or S2 during Phase 3. Fifteen blocks of each event occurred in Phase 3. Note that the fillers served to encourage learning about specific cues in Phases 1 and 3 and discourage the learning of nonspecific rules.

Finally, participants were instructed that, for each cue (digits and faces), they would have to respond by pressing the "WIN" key if they thought that the cue predicted the outcome, and the "empty screen" key if they thought that it predicted the absence of the outcome. They were instructed throughout the test to keep the index finger of their preferred hand positioned on the down arrow key except for when they were responding. Participants indicated their choice by pressing the left or right arrow keys (counterbalanced as "WIN" or "empty screen" across participants within conditions in each of the two groups). Because responding to a cue may influence subsequent responding to its associate, test order was recorded for all cues. Participants were not informed that their reaction times (RTs) were being recorded, but they were encouraged in the instructions on the screen to respond as soon as they knew their answer. Also, participants were not informed whether their response was correct or not (i.e., no feedback; specifically, they were told on the screen "This time there will be no feedback (i.e., no outcomes will be displayed even if they occurred.)". Each of the six S1s and the six S2s seen by participants during training was tested plus $S1_{Novel}$ and $S2_{Novel}$. Each test stimulus remained on the screen until the participant had responded; then "next trial" was displayed for 2 s. The 14 test cues were presented in random order.

The type of response (i.e., "WIN" or "empty screen") on each test was recorded as well as the RT to respond. These two dependent variables were transformed into a single variable following Craddock, Molet, and Miller's (2012) transformation rule. To normalize the RTs, each of the 14 test RTs was divided by the participant's largest RT of the 14 (RTmax). Any "WIN" response RT was then transformed into [−ln (RT / RTmax)], and any "empty screen" response RT into [+ln (RT / RTmax)]. Hence, transformed data near zero (i.e., RT / RTmax near 1) denote hesitating responses, while positive or negative transformed data indicate that the participant answered "WIN" or "empty screen" without hesitation, respectively. Because a slight increase or decrease in a short RT is presumably more significant than an equivalent increase or decrease in a long RT, differences between short RTs were amplified relative to differences between long RTs through the use of a logarithmic transformation. This transformation also minimized the positive skew ordinarily seen in RTs and yielded data sets within each condition that were approximately normally distributed.

Our first test of the mediated-extinction hypothesis compared responding to $S1_{S2ext}$ and $S1_{SOC}$. Mediated extinction would take the form of less responding to $S1_{S2ext}$ presumably because activation of the representation of $S1_{S2ext}$ without the outcome during extinction of $S2_{S2ext}$ during Phase 3 provides an opportunity for mediated extinction of $S1_{S2ext}$. Importantly, activation of $S1_{S2ext}$ during extinction of $S2_{S2ext}$ in Phase 3 was expected to be strong due to the forward relationship of $S2_{S2ext}$ to $S1_{S2ext}$ during Phase 2. In contrast, extinction of $S1_{S1ext}$ should have had a smaller decremental effect on

Table 1 Design of Experiment 1

S1	S2	Condition	Phase 1	Phase 2	Phase 3	Test	Mediated Extinction Predictions	Associative-Chain Predictions
$S1_{SOC}$	$S2_{SOC}$	SOC	3 $S1_{SOC}$–Outcome	1 $S2_{SOC}$–$S1_{SOC}$	–	1 $S1_{SOC}$? and 1 $S2_{SOC}$?	$S1_{SOC}$ and $S2_{SOC}$ positive	$S1_{SOC}$ and $S2_{SOC}$ positive
$S1_{S1ext}$	$S2_{S1ext}$	SOC–S1 ext	3 $S1_{S1ext}$–outcome	1 $S2_{S1ext}$–$S1_{S1ext}$	15 $S1_{S1ext}$–no Outcome	1 $S1_{S1ext}$? and 1 $S2_{S1ext}$?	$S1_{S1ext}$ negative $S2_{S1ext}$ positive	$S1_{S1ext}$ negative $S2_{S1ext}$ negative
$S1_{S2ext}$	$S2_{S2ext}$	SOC–S2 ext	3 $S1_{S2ext}$–outcome	1 $S2_{S2ext}$–$S1_{S2ext}$	15 $S2_{S2ext}$–no Outcome	1 $S1_{S2ext}$? and 1 $S2_{S2ext}$?	$S1_{S2ext}$ negative $S2_{S2ext}$ negative	$S1_{S2ext}$ positive $S2_{S2ext}$ negative
$S1_{Filler1}$	$S2_{Filler1}$	Filler1 S1–S1–outcome	3 $S1_{Filler1}$–no Outcome	1 $S2_{Filler1}$–$S1_{Filler1}$	15 $S1_{Filler}$–Outcome	1 $S1_{Filler1}$? and 1 $S2_{Filler1}$?	$S1_{Filler1}$ positive $S2_{Filler1}$ negative	$S1_{Filler1}$ positive $S2_{Filler1}$ positive
$S1_{Filler2}$	$S2_{Filler2}$	Filler 2 S1	3 $S1_{Filler2}$–no Outcome	1 $S2_{Filler2}$–$S1_{Filler2}$	–	1 $S1_{Filler2}$? and 1 $S2_{Filler2}$?	$S1_{Filler2}$ negative $S2_{Filler2}$ negative	$S1_{Filler2}$ negative $S2_{Filler2}$ negative
$S1_{Filler3}$	$S2_{Filler3}$	Filler 3 S1–S2–outcome	3 $S1_{Filler3}$–no Outcome	1 $S2_{Filler3}$–$S1_{Filler3}$	15 $S2_{Filler3}$–outcome	1 $S1_{Filler3}$? and 1 $S2_{Filler3}$?	$S1_{Filler3}$ positive $S2_{Filler3}$ positive	$S1_{Filler3}$ negative $S2_{Filler3}$ positive
$S1_{Novel}$	$S2_{Novel}$	Novel				1 $S1_{Novel}$? and 1 $S2_{Novel?}$	$S1_{Novel}$ negative $S2_{Novel}$ negative	$S1_{Novel}$ negative $S2_{Novel}$ negative

Note. Numbers to the left of each stimulus designation indicate the number of each type of trial. S1s were always digits. S2s were always faces. Positive and negative refer to the predicted sign of transformed RTs. Positive means that the outcome is predicted; negative means that the absence of outcome is predicted

responding to $S2_{S1ext}$, relative to $S2_{SOC}$, because $S1_{S1ext}$'s association with $S2_{S1ext}$ was backwards (i.e., in Phase 2, $S1_{S1ext}$ followed $S2_{S1ext}$). In contrast to the mediated-extinction account, the associative-chain account anticipates decreased responding to S2 when S1 is extinguished because, at test, S2 evokes S1 and its extinguished response.

Results and discussion

Mean transformed RTs for each cue are depicted in Fig. 2. SOC was demonstrated by comparing scores for $S2_{SOC}$ to zero, $t(23) = 2.18$, $p < .04$, $d = 0.40$. Also, $S2_{SOC}$ yielded higher scores than novel $S2_{Novel}$, $t(23) = 3.75$, $p < .002$, $d = 0.76$, and higher scores than $S2_{Filler2}$, an S2 paired with an explicitly unpaired S1, $t(23) = 2.69$, $p < .02$, $d = 0.55$.

Contrary to what is often found in the literature (e.g., Rizley & Rescorla, 1972), the comparison of $S2_{S1ext}$ to $S2_{SOC}$ found that extinction of $S1_{S1ext}$ during Phase 3 decreased responding to $S2_{S1ext}$, $t(23) = 5.36$, $p < .001$, $d = 1.10$, while extinction of $S2_{S2ext}$ had relatively little effect on responding to $S1_{S2ext}$. $S1_{S2ext}$ yielded nonsignificantly lower scores than $S1_{SOC}$, $t(23) = 2.03$, $p > .05$, $d = 0.41$. No ANOVA was performed to quantitatively compare these two differences because S2s were faces and second-order cues, whereas S1s were digits and first-order cues, which would confound any direct comparison between these two differences.

During Phases 2 and 3 for Condition Filler 1, $S1_{Filler1}$ was paired with first $S2_{Filler1}$ and then the outcome in a sequence similar to sensory preconditioning (SPC) of $S2_{Filler1}$, which appears to have increased responding to $S2_{Filler1}$ compared to $S2_{Filler2}$, $t(23) = 2.89$, $p < .01$, $d = 0.59$. However, no evidence was found that reinforcing $S2_{Filler3}$ in Phase 3 influenced responding to $S1_{Filler3}$ in Condition Filler 3 (compared to $S1_{Filler2}$), $t(23) = 0.23$, $d = 0.05$. Note that SPC in the former case was likely attenuated by latent inhibition arising from pre-exposing participants during Phase 1 to $S1_{Filler1}$ alone prior to the pairings of $S2_{Filler1}$ with $S1_{Filler1}$.

Experiment 1 yielded results that differed from what was expected based on the majority of the existing literature concerning sequential SOC in that we observed reduced response strength of S2 as a result of post-SOC extinction of S1. However, we do note that similar results were obtained by Molet et al. (2012) with sequential S2–S1 pairings with rats. Interpretation of the effect of extinguishing $S1_{S1ext}$ on responding to $S2_{S1ext}$ in Condition SOC-S1 ext in terms of the formation of an association between the representation of S2 and the absence of outcome during Phase 3 (i.e., mediated extinction) is unlikely for two reasons. First, it would have to assume that S1 backward activated the representation of S2. Second, if backward activation of $S1_{SOC}$ is assumed, why was it stronger than forward activation of $S1_{S2ext}$ by $S2_{S2ext}$ in Condition SOC-S2 ext, given that forward pairings are widely seen to result in stronger behavioral control than backward pairings (e.g., Pavlov, 1927)? Modulations of responding to $S2_{S2ext}$ did not appreciably influence responding to $S1_{S2ext}$. These observations cast doubt on the mediated-extinction explanation of how extinction of $S1_{S1ext}$ on responding to $S2_{S1ext}$ in Condition SOC-S1 ext.

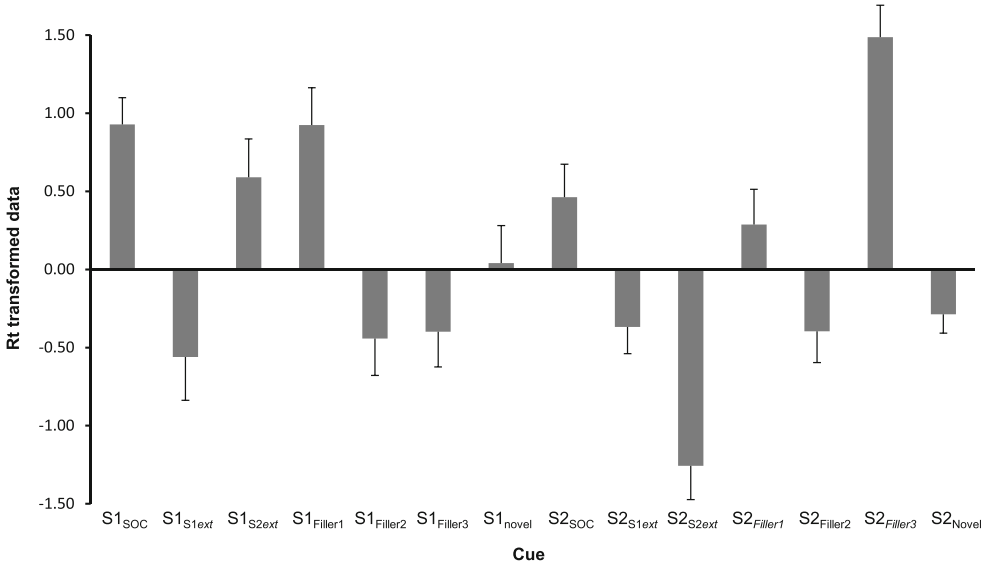

Fig. 2 Mean transformed RT and standard error as a function of cue for Experiment 1. S1s were always digits. S2s were always faces. See Table 1 for the roles of the various S1s and S2s

The associative-chain account more readily explains the present results by assuming that responding to $S1_{S2ext}$ in Condition SOC-S2 ext at test reflected the level of $S1_{S2ext}$–outcome association acquired during Phase 1 because backward activation of $S2_{S2ext}$ by $S1_{S2ext}$ *at test* was weak and hence did not engage $S2_{S2ext}$'s association to "no outcome." However, extinction of $S1_{S1ext}$ greatly influenced responding to $S2_{S1ext}$ at test, presumably because the test of $S2_{S1ext}$ forward activated $S1_{S1ext}$'s representation.

Finally, the data provided no support for the direct association account of SOC according to which neither S2 nor S1 should be influenced by their respective companion's extinction because what counts in that framework is only the association between the tested cue and the outcome. Importantly, in Experiment 1, assessment of the mediated-extinction account of S2's dependency on extinction of S1 was indirect as it hinged on assumptions about directionality of associations being of differential effectiveness and on digits (i.e., S1s) and faces (i.e., S2s) having similar associabilities. Hence, a further test of the mediated-extinction hypothesis of how extinction of S1 might decrease responding to S2 seemed warranted.

Experiment 2

Experiment 2 further assessed whether SOC depends more on an S2–S1–US associative chain than any potential direct S2–US association with mediated extinction explaining the reduction in responding to S2 that is sometimes seen when S1 is extinguished. This was accomplished by presenting the S1-no outcome trials *prior* to the S2–S1 pairings, thus preventing any direct S2–outcome association from being modified post-SOC. In Phase 1, $S1_{SOC1}$ was reinforced in Context A and nonreinforced in Context B, a distinctly different context (see Table 2). Training with $S1_{SOC2}$ was identical to that of $S1_{SOC1}$, but with Contexts A and B playing reversed roles. Critically, although excitation of these two S1s in Context A (or B) initially generalized to Context B (or A), over repeated training trials they were extinguished in Context B (or A) while their excitatory values were maintained in Context A (or B). Thus, all training of $S1_{SOC1}$ and $S1_{SOC2}$ occurred before $S2_{SOC1}$ and $S2_{SOC2}$ were ever paired with them. This precluded the evoked representations of $S2_{SOC1}$ ($S2_{SOC2}$) from being paired with the absence of the outcome during the nonreinforced $S1_{SOC1}$ ($S1_{SOC2}$) trials in Context B (A). Then, $S2_{SOC1}$ and $S2_{SOC2}$ were sequentially paired with $S1_{SOC1}$ and $S1_{SOC2}$, respectively, in a novel context (C) in which generalized conditioned expectation of the outcome was expected. Presenting $S1_{SOC1}$ and $S1_{SOC2}$ in Contexts A or B never evoked $S2_{SOC1}$ or $S2_{SOC2}$ because these S1s and S2s had not yet been paired with each other. Therefore, based on the mediated-extinction hypothesis, responses to $S2_{SOC1}$ and $S2_{SOC2}$ were expected to be equivalent in whichever context they were tested. However, if responses to $S2_{SOC1}$ ($S2_{SOC2}$) mimicked the expected context dependency of responses to $S1_{SOC1}$ ($S2_{SOC2}$), that is, renewal in Context A (C), the results would support an associative-chain account of SOC.

In Experiment 2, following conditioned discrimination training of $S1_{SOC1}$ and $S1_{SOC2}$ between Contexts A and B in Phase 1 and subsequent $S2_{SOC1}$–$S1_{SOC1}$ and $S2_{SOC2}$–$S1_{SOC2}$ pairings in Phase 2 in Context C, all four cues were tested in the contexts in which $S1_{SOC1}$ and $S1_{SOC2}$ had been differentially reinforced (A and B) and the context in which the S2–S1 pairings had subsequently occurred (C). Importantly, the S2s were paired with the S1s in a neutral context (C) just before testing; that is, there was no post-SOC training of $S1_{SOC1}$, $S1_{SOC2}$, $S2_{SOC1}$, or $S2_{SOC2}$. The mediated-extinction

Table 2 Design of Experiment 2

Condition	Phase 1: Contexts A and B intermixed	Phase 2: Context C only	Test in Contexts A, B, and C
SOC1	$(S1_{SOC1}–outcome)_A$ / $(S1_{SOC1}–noOutcome)_B$	3 $(S2_{SOC1}–S1_{SOC1})_C$	$1(S1_{SOC1}$ & $S2_{SOC1})_{A,\ B,\ and\ C}$
SOC2	$(S1_{SOC2}–noOutcome)_A$ / $(S1_{SOC2}–outcome)_B$	3 $(S2_{SOC2}–S1_{SOC2})_C$	1 $(S1_{SOC2}$ & $S2_{SOC2})_{A,\ B,\ and\ C}$
$S2_{alone}$	$(S1_{alone}–outcome)_A$ / $(S1_{alone}–outcome)_B$	3 $(S2_{S2alone})_C$	1 $(S1_{S2alone}$ & $S2_{alone})_{A,\ B,\ and\ C}$
NoS2	$(S1_{noS2}–noOutcome)_A$ / $(S1_{noS2}–outcome)_B$		$1(S1_{noS2})_{A,\ B,\ and\ C}$
Novel			1 $(S1_{Novel}$ & $S2_{Novel})_{A,\ B,\ and\ C}$

Note. Numbers to the left of each stimulus designation indicate the number of each type of trial. S1s were always digits. S2s were always faces. Slashes separate interspersed trials. During Phase 1, the retention criterion had to be reached in 12 or fewer trial cycles

hypothesis, concerning the effects of post-SOC extinction of S1 on S2, as well as the direct association account of SOC upon which mediated extinction hypothesis is based, assumes that decrementing responding to S2 requires activation of the representation of S2 along with the absence of the outcome's representation. As S2 could not have been evoked during the Phase 1 S1-no outcome presentations, in the direct S2–US framework the test context was expected to control responding to S1 ($S1_{SOC1}$ and $S1_{SOC2}$) but not to S2 ($S2_{SOC1}$ and $S2_{SOC2}$). However, if SOC was based on an S2–S1–US associative chain which depended on activation S1 at test, responses to S2 should mimic the context dependency of responding to that specific S1.

Method

Participants

Forty-eight university students (males and females; 18–29 years of age), from SUNY-Binghamton, volunteered. Their participation was approved by the local Institutional Review Board.

Apparatus

As in Experiment 1, S1s and S2s were represented by digits and faces, respectively, and the outcome was the same as in Experiment 1. The background screen was one of three color–texture dyads, different for each experimental context (i.e., A, B, and C.). The roles of the three color–texture dyads were randomly assigned across conditions independently for each participant.

Procedure

Conditioned discrimination training during Phase 1 consisting of eight trials/block of four S1 cues each, which were reinforced in one context and nonreinforced in a second context. Specifically, in Context A, $S1_{SOC1}$ and $S1_{S2alone}$ were reinforced while $S1_{noS2}$ and $S1_{SOC2}$ were nonreinforced. In Context B, $S1_{noS2}$ and $S1_{SOC2}$ were reinforced, whereas $S1_{SOC1}$ and $S1_{S2alone}$ were nonreinforced (see Table 2). Phase 1 was conducted as repeated cycles of the eight trial types depicted in Table 2, each type of trial occurred once in a random order for each subject, and the order was determined anew for each cycle. Phase 1 was continued for each participant until the participant's response was correct on all eight trials for two consecutive cycles, at which time the participant was advanced to Phase 2. Any participant who failed to reach this retention criterion within 12 cycles of Phase 1 training was informed that the experiment was over and was dismissed.

Each trial of Phase 1 was preceded with 0.5 s of gray screen (ITI) that was devoid of any overt contextual features, 1.0 s of the current trial context (A or B), presentation of the cue (S1 or S2 with duration dependent on the participant's response) in the current trial context, 1.5 s of presentation of the outcome or absence of the outcome in the current trial context, and 1.0 s further exposure to the current trial context. Text at the top of the screen reminded participants which key was to be used to predict either the presence or absence of the "WIN" outcome. A purple sticker on one of the keys represented the outcome-present key, while an orange sticker represented the outcome-absent key. When the "WIN" predicted response key was the left key (i.e., when the purple sticker was on the left key), S1 appeared in the middle and the "WIN" outcome appeared on the left. Conversely, when the "WIN" predicted response key is the right key, S1 appeared in the middle and "WIN" appeared on the right. The two keys were the horizontal cursor keys and their assignment was counterbalanced across subjects.

A Phase 2 cycle consisted of one each of three trial types: $S2_{SOC1}–S1_{SOC1}$ and $S2_{SOC2}–S1_{SOC2}$ (sequential pairings corresponding to, respectively, the SOC1 and SOC2 conditions), and $S2_{S2alone}$ presented by itself (i.e., the S2-alone condition), all of which occurred in a novel context (C). In order to minimize potential extinction for $S1_{SOC1}$ and $S1_{SOC2}$ during Phase 2, we limited Phase 2 to three cycles. If the "WIN" outcome appeared on the left side of the screen in Phase 1, S2 appeared on the opposite side (right side) for Phase 2. Conversely, I, if "WIN" appeared on the right side during Phase 1, S2 appeared on the left side in Phase 2. A Phase 2 trial in Context C included a 2.5-s ITI (an increase of the ITI for the S2–S1 pairings from Experiment 1, which was intended to minimize potential learning across adjacent trials), and 1.5-s

presentation of each stimulus (sequentially when there was a pairing). A 2.5-s ITI preceded the first Phase 2 trial and followed the last Phase 2 trial. The order of the three trial types within a cycle was random and drawn without replacement.

Testing occurred once in each of the three contexts (A, B, and C) in one of the following six orders, ABC, BCA, CAB, ACB, BAC, and CBA, with the order of test context counterbalanced across subjects. All cues (including novel $S1_{Novel}$ and $S2_{Novel}$ cues) were tested once in each context with the test trials blocked by context and, within test block, the order of cues randomly assigned to each participant. After having completed the entire experiment, participants were presented with a debriefing screen and were excused.

Results and discussion

In order to have a consistent estimator of effect sizes, ($d = \frac{T}{\sqrt{n}}$) was used to estimate both the effect size of comparisons between two means and the effect size of specific planned 2 × 2 interactions. In the latter case, $d = \sqrt{\frac{F\ of\ interaction}{n}}$.

SOC was tested in Context C by comparing presumably neutral $S2_{S2alone}$ to $S2_{SOC1}$ and $S2_{SOC2}$ (see Fig. 3). $S2_{SOC1}$ yielded higher scores than $S2_{S2alone}$ in Context C, $t(47) = 4.57$, $p < .001$, $d = 0.66$, and $S2_{SOC2}$ yielded higher scores than $S2_{S2alone}$ in C, $t(47) = 5.74$, $p < .001$, $d = 0.83$. No significant difference between responses to $S2_{SOC1}$ and $S2_{SOC2}$ was observed in Context C, $t(47) = 1.51$, $p < .15$, $d = 0.22$.

As previously mentioned, the prediction of the mediated-extinction hypothesis and the underlying S2–outcome direct association account of SOC is that there should be no difference in responding between $S2_{SOC1}$ and $S2_{SOC2}$ in either Context A or Context B. This is because, according to this hypothesis, second-order conditioned discrimination requires S2 or its evoked representation to undergo pairings with the absence of outcome in one context and not the other. Therefore, because no such post-SOC training occurred, $S2_{SOC1}$ should have yielded responses similar to $S2_{SOC2}$ regardless of the context in which testing occurred. Admittedly, when testing occurs in a different context from the one in which SOC was learned, inter-context generalization may yield weaker second-order CRs (Hall & Honey, 1990). However, in that case, both $S2_{SOC1}$ and $S2_{SOC2}$ should have been affected the same way when tested in Contexts A and B. That is, responding to $S2_{SOC1}$ and $S2_{SOC2}$ should generalize indifferently to Contexts A and B because the two cues to which they were paired in Context C (i.e., $S1_{SOC1}$ and $S1_{SOC2}$, respectively) had received the same amount of reinforcement and nonreinforcement during Phase 1. The results showed that responses to $S2_{SOC1}$ and $S2_{SOC2}$ changed in opposite directions between Context C and Contexts A and B. That is, there was a significant interaction between cues ($S2_{SOC1}$ vs. $S2_{SOC2}$) and contexts (C vs. A), $F(1, 47) = 35.05$, $p < .001$, $d = 0.85$. Additionally, there was an interaction between cues ($S2_{SOC1}$ vs. $S2_{SOC2}$) and context (C vs. B), $F(1, 47) = 5.41$, $p < .025$, $d = 0.35$. Moreover, $S2_{SOC1}$ yielded higher scores than $S2_{SOC2}$ in Context A, $t(47) = 7.06$, $p < .001$, $d = 1.02$, whereas the reverse was observed in Context B in which $S2_{SOC2}$ yielded higher scores than $S2_{SOC1}$, $t(47) = 4.69$, $p < .001$, $d = 0.68$. Consistently, the interaction between context (A and B) and cues ($S2_{SOC1}$ and $S2_{SOC2}$) was significant, $F(1, 47) = 47.08$, $p < .001$, $d = 0.99$. As predicted, no appreciable differential effect of testing in Contexts A and B on responding to $S2_{S2alone}$, which was never paired with any S1, was observed, $t(47) = 0.10$, $p < .95$, $BF_{01} = 6.94$ (where BF_{01} is the Bayesian factor is in favor of the null hypothesis). Because the same S2, whether it was $S2_{SOC1}$ or $S2_{SOC2}$, yielded high responding or weak responding dependent upon the test context, it appears that responding to $S2_{SOC1}$ and $S2_{SOC2}$ was controlled by the contextually modulated associative status of $S1_{SOC1}$ and $S1_{SOC2}$, respectively, at the time of testing. This is contrary to what the mediated-extinction hypothesis predicted but fully concordant with the associative-chain account of SOC.

Finally, as expected from training, $S1_{SOC1}$ yielded higher scores in Context A than in Context B, $t(47) = 8.48$, $p < .001$, $d = 1.20$. The results were inverted for $S1_{SOC2}$, which yielded higher scores in Context B than in Context A, $t(47) = 9.68$, $p < .001$, $d = 1.39$. Moreover, $S2_{SOC1}$ mimicked $S1_{SOC1}$ by yielding higher scores in Context A than in Context B, $t(47) = 4.10$, $p < .001$, $d = 0.59$, and $S2_{SOC2}$ mimicked $S1_{SOC2}$ by yielding higher scores in Context B than in Context A, $t(47) = 8.58$, $p < .001$, $d = 1.24$. These latter results, together with those presented earlier, lend support to the associative-chain account of SOC because S2 elicited responses reflecting those to S1.

General discussion

Experiment 1 was expected to demonstrate that SOC with sequential S2–S1 pairings depends on a direct S2–outcome (or S2–CR) association that is impervious to extinction of S1 following the S2–S1 SOC pairings. However, Experiment 1 found that sequential SOC was attenuated by extinction of S1. Consequently, we examined the associative structure of sequential SOC using the data from Experiment 1 and then designed Experiment 2 as a more direct test of the mechanism by which extinction of S1 decreased (sequential) SOC.

Specifically, Experiment 1 tested the respective influence of extinguishing S1 or S2 (after they had been sequentially paired) on responding to the companion cue, S2 or S1, respectively. As we mentioned in the introduction, the learning of a direct association between S2 and the outcome (or CR) had previously been suggested to account for the absence of a decrease in the response to S2 when S1 was extinguished following sequential SOC. Assuming a direct S2–outcome

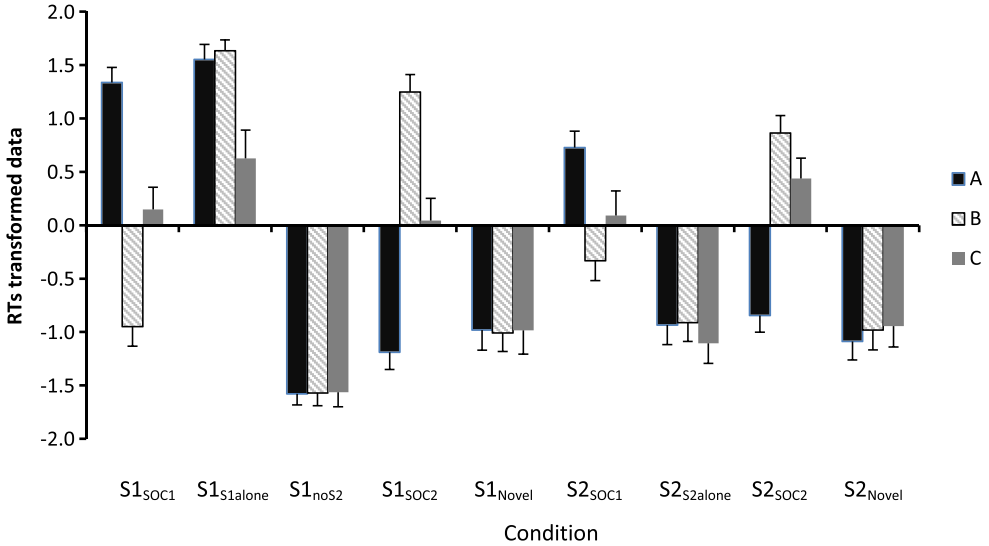

Fig. 3 Mean transformed RT and standard error as a function of cue for Experiment 2. S1s were always digits. S2s were always faces. See Table 1 for the roles of the various S1s and S2s

actually underlies SOC when S2 and S1 are paired sequentially, our mediated-extinction hypothesis asserts that this hypothetical S2–outcome direct association survives extinction of S1 due to the backward S2←S1 association not inducing extinction of S2 during extinction of S1. Thus, we expected little decrement in responding to S2 when S1 was extinguished. In contrast, we anticipated a strong influence of S2's extinction on responding to S1 because of the forward association from S2 to S1. That is, the forward S2→S1 association was expected to result in an effective pairing of the evoked representation of S1 and the absence of the outcome during extinction of S2, which could result in reduced responding to S1. The results did not support either of these expectations. Rather, extinction of S1 greatly reduced responding to S2 despite the assumed weak potential of S1 to activate S2's representation during Phase 3. Additionally, extinction of S2 failed to appreciably reduce responding to S1. Given the forward relationship of S2 to S1, it is here that we had the highest expectation of observing mediated extinction, but it did not materialize. Thus, the results of Experiment 1 appear to support the associative-chain account of SOC.

Although the failure of extinction of S2 to reduce responding to S1 in Experiment 1 argues against the mediated-extinction hypothesis, the argument here is indirect because S2 and S1 were distinctly different cues functionally and physically (i.e., second-order faces and first-order digits, respectively). In contrast, Experiment 2 provided a more direct test of the mediated-extinction hypothesis and the direct association account of SOC more generally. The procedures used in Experiment 2 categorically precluded the possibility of an evoked S2 during extinction of S1 because the S2–S1 pairings occurred after extinction of S1. Moreover, Experiment 2 assessed responding to S2 as a function of the associative status of S1, which was manipulated by test Contexts A and B serving as discriminative stimuli for different response potentials of S1. Because the S2–S1 pairings were presented after S1's discriminative training (i.e., excitatory conditioning in Context A and extinction in Context B), any decrease in responding to S2 in Context B relative to Context A (and Context C) would have to be attributed to activation of S1 at test. The results here, too, were consistent with the associative-chain account of SOC. Importantly, they showed that responses to the representation of S1 evoked by S2 were determined by the effective relationship that S1 had with the outcome in each test context. Alternatively stated, the S1 representation evoked by S2 predicted the outcome if and only if the actual S1 predicted the outcome in the test context. Notably, the reverse does not seem to be true. That is, Experiment 1 showed that potentially extinguishing the evoked S1 by extinguishing S2 had little to no effect on responding to S1.

It should be pointed out that preservation of responding to S2 following extinction of S1 that has been observed by some researchers in sequential SOC preparations (in contrast to what we observed in the present experiments) does not necessarily preclude SOC depending on an S2–S1–US associative chain at test. Because an extinguished response is often renewed when the cue is tested in a different context from the one in which it was extinguished (Bouton & Bolles, 1979; Bouton & King, 1983), it is possible that responding to S2 in those experiments was enabled by activation of a renewed S1–outcome association. In this situation, responding to S2 at test may have been the result of an evoked S1 representation, the response potential of which was renewed because the context of S1's evocation during a test of S2 (i.e., physical S2) differed from the context in which S1 was extinguished (i.e., absence of physical S2). Hence, the context of S1's extinction is the absence of S2, while the context of S1's evocation at test of S2 is S2. Therefore, the sometimes observed strong response to S2 outside the context

of nonreinforcement of S1 may reflect a mediated renewed response (i.e., S2→S1→renewed CR). This alternative explanation of why extinction of S1 sometimes fails to reduce responding to S2 (i.e., an associative chain involving a renewed S1 response cannot account for the results we observed in Experiment 1 because the test of $S2_{S1ext}$ yielded low scores (i.e., no mediated renewal was observed). But it can possibly explain the absence of modulation of S2's response reported by several authors, and it highlights the potential role of the context in controlling acquired responses.

In the associative-chain framework, presentations of S2 without S1 during extinction of S2 weakened the S2–S1 association necessary to observed responding to S2 that depended on activation of S1 at test. Thus, in Experiment 1, when S2 was extinguished, the test of S2 was expected to elicit a distinctly weaker response than the one S1 elicited (consistent with $S2_{S2ext}$ yielding negative scores and $S1_{S2ext}$ yielding positive scores) because of weakened evocation of S1 by the extinguished S2. In Experiment 2, learning that S1 was followed by no outcome in one context did not eradicate S1's potential to evoke the outcome representation in a different context. Thus, in the associative-chain framework, strong responding was expected to S1 and consequently S2 outside of the context in which S1 was nonreinforced. The present data support both of these expectations.

More speculatively, the current results are consistent with an optimal predictive-learning framework in that they suggest economy in the rules governing associations. Organisms adapt by learning to anticipate events that are likely to happen. Therefore, the more they take advantage of prior learning, the less they waste time and energy. The evoked representation of S1 in the presence of S2, following sequential learning, signals the organism of the likely appearance of S1. Experiment 1 shows that what was learned to follow S1 is now expected given the presence of S2, allowing the organism without direct S2–outcome experience to be prepared to receive or avoid S1's outcome. More importantly, if S1 stops predicting the outcome, all events which preceded S1 inherit extinction together with S1. Had the mediated-extinction hypothesis been validated, each cue having established a second-order association with the representation of the outcome would have had to undergo extinction itself through its evoked representation. Therefore, our data from Experiment 1 might be viewed as reflecting adaptive processes. The same principle characterizes what was observed in Experiment 2. Responding to S2 reflected context-dependent associations between S1 and what followed. Here again, once S2 predicted S1, it was the associative status of S1 that controlled responding to S2, sparing the organism having to learn individual meanings of second-order cues dependent upon the context.

In conclusion, the mediated-extinction hypothesis was proposed as a novel means of explaining the decremental effect of extinguishing S1 on responding to S2 in SOC. As it was expected to apply to S2–S1 simultaneous pairings and not S2–S1 sequential pairings because of the backward relationship of S2 to S1, its success could have reduced both types of SOC (i.e., simultaneous and sequential) to a single direct S2–US associative structure, which would have been pleasing in its parsimony. However, Experiment 1 and, more directly, Experiment 2 provided evidence against the mediated-extinction account of why post-SOC extinction of S1 only sometimes reduces responding to S2. Thus, although the mediated-extinction account in principle provided a potentially unifying account of the associative structure of all SOC, the present data argue strongly against it. As both of the present experiments examined sequential SOC, one might ask whether mediated extinction might at least contribute to the commonly observed reduction in responding to S2 that results from extinguishing S1. However, even if such extinction of the evoked representation of S2 contributes to this effect in simultaneous SOC, the present data demonstrate that the absence of mediated extinction in sequential SOC renders implausible the kind of direct associative structure of all SOC that we initially had entertained.

Acknowledgements This research was supported in part by NIMH Award 33881. We thank Crystal Casado, Tori Pena, Jessica Pino, Reid M. Portnoy, Benjamin M. Seitz, and Anna Tsvetkov for their comments on a prior version of this manuscript.

References

Barnet, R. C., Arnold, H. M., & Miller, R. R. (1991). Simultaneous conditioning demonstrated in 2nd-order conditioning: Evidence for similar associative structure in forward and simultaneous conditioning. *Learning and Motivation, 22,* 253–268.

Bouton, M. E., & Bolles, R. C. (1979). Contextual control of the extinction of conditioned fear. *Learning and Motivation, 10,* 445–466.

Bouton, M. E., & King, D. A. (1983). Contextual control of the extinction of conditioned fear: Tests for the associative value of the context. *Journal of Experimental Psychology: Animal Behavior Processes, 9,* 248–265.

Craddock, P., Molet, M., & Miller, R. R. (2012). Reaction time as a measure of human associative learning. *Behavioural Processes, 90*, 189–197.

Dickinson, A., & Burke, J. (1996). Within-compound associations mediate the retrospective revaluation of causality judgments. *Quarterly Journal of Experimental Psychology, 49B,* 60–80.

Hall, G. (1996).Learning about associatively activated stimulus representations: Implications for acquired equivalence and perceptual learning. *Animal Learning & Behavior, 24*, 233–255.

Hall, G., & Honey, R. C. (1990). Context-specific conditioning in the conditioned-emotional-response procedure. *Journal of Experimental Psychology: Animal Behavior Processes, 16*, 271–278.

Holland, P. C. (1981). Acquisition of representation-mediated conditioned food aversions. *Learning and Motivation, 12,* 1–18.

Holland, P. C. (1983). Representation-mediated overshadowing and potentiation of conditioned aversions. *Journal of Experimental Psychology: Animal Behavior Processes, 9,* 1–13.

Holland, P. C., & Forbes, D. T. (1982). Representation-mediated extinction of conditioned flavor aversions. *Learning and Motivation, 13*, 454–471.

Jara, E., Vila, J., & Maldonado, A. (2006). Second-order conditioning of human causal learning. *Learning and Motivation, 37*, 230–246.

Konorski, J. (1967). Integrative activity of the brain. Chicago: University of Chicago Press.

Minear, M., & Park, D. C. (2004). A lifespan database of adult facial stimuli. *Behavior Research Methods, Instruments, & Computers, 36*, 630–633.

Molet, M., Miguez, G., Cham, H. X., & Miller, R. R. (2012). When does integration of independently acquired temporal relationships take place? *Journal of Experimental Psychology: Animal Behavior Processes, 38*, 369–380.

Pavlov, I. P. (1927). Conditioned reflexes (G. V. Anrep, Trans.). London: Oxford University Press.

Polack, C. W., Molet, M., Miguez, G., & Miller, R. R. (2013). Associative structure of integrated temporal relationships. *Learning & Behavior 41*, 443–454.

Rescorla, R. A. (1982). Simultaneous second-order conditioning produces S-S learning in conditioned suppression. *Journal of Experimental Psychology: Animal Behavior Processes, 8,* 23–32.

Rizley, R. C., & Rescorla, R. A. (1972). Associations in second-order conditioning and sensory preconditioning. *Journal of Comparative and Physiological Psychology, 81,* 1–11.

Wagner, A. R. (1981). SOP: A model of automatic memory processing in animal behavior. In N. E. Spear & R. R. Miller (Eds.), Information processing in animals: Memory mechanisms (pp. 5–47). Hillsdale: Erlbaum.