Journal of Experimental Psychology: General
2016, Vol. 145, No. 9, e49–e71

0096-3445/16/$12.00 http://dx.doi.org/10.1037/xge0000200

The Elusive Nature of the Blocking Effect: 15 Failures to Replicate

Elisa Maes, Yannick Boddez,
and Joaquín Matías Alfei
KU Leuven

Angelos-Miltiadis Krypotos
Utrecht University

Rudi D'Hooge
KU Leuven

Jan De Houwer
Ghent University

Tom Beckers
KU Leuven and University of Amsterdam

With the discovery of the blocking effect, learning theory took a huge leap forward, because blocking provided a crucial clue that surprise is what drives learning. This in turn stimulated the development of novel association-formation theories of learning. Eventually, the ability to explain blocking became nothing short of a touchstone for the validity of any theory of learning, including propositional and other nonassociative theories. The abundance of publications reporting a blocking effect and the importance attributed to it suggest that it is a robust phenomenon. Yet, in the current article we report 15 failures to observe a blocking effect despite the use of procedures that are highly similar or identical to those used in published studies. Those failures raise doubts regarding the canonical nature of the blocking effect and call for a reevaluation of the central status of blocking in theories of learning. They may also illustrate how publication bias influences our perspective toward the robustness and reliability of seemingly established effects in the psychological literature.

Keywords: blocking, replicability, associative learning theory, prediction error, surprise

Ivan Pavlov's (1927) experiments on classical conditioning in dogs were a milestone in the study of elementary learning processes. The paradigm he introduced allowed researchers to investigate the psychological principles underlying associative learning in a relatively simple and highly controlled and systematic manner (Delamater & Lattal, 2014). In a classical conditioning experiment, presentations of a conditioned stimulus (conditional stimulus [CS]) are repeatedly followed by presentations of an unconditioned stimulus (US). As a result, the CS comes to elicit a conditioned response (CR) that it did not elicit before. Early psychological theories of learning (e.g., Bush & Mosteller, 1951) assumed that the mere co-occurrence of the CS and the US in space and time (i.e., spatiotemporal contiguity) was sufficient for this type of learning to occur. This idea was challenged by the observation of stimulus competition. It was Kamin who, with the discovery of the blocking effect (Kamin, 1969), suggested an alternative driving force for learning: surprise. The design and results of his blocking experiment are shown in Figure 1. The experiment consisted of three phases. In the first phase, the experimental group received several presentations of a noise followed by a footshock US (N+ training). In the second phase, both the experimental and control group received presentations of the noise compounded with a light followed by footshock (NL+ training). Thus, the difference between both groups was that the experimental group was conditioned to the noise before receiving compounded presentations of noise and light followed by shock, while the control group was not previously conditioned to the noise. In the third phase, Kamin tested conditioned responding to the light when it was presented alone. If contiguity is the sole determinant of learning, then no between-groups difference should have been observed, because both groups received an equal number of light-shock pairings. As

---

Elisa Maes, Yannick Boddez, and Joaquín Matías Alfei, Department of Psychology, KU Leuven; Angelos-Miltiadis Krypotos, Department of Clinical Psychology, Utrecht University; Rudi D'Hooge, Department of Psychology, KU Leuven; Jan De Houwer, Department of Experimental Clinical and Health Psychology, Ghent University; Tom Beckers, Department of Psychology, KU Leuven and Department of Clinical Psychology, University of Amsterdam.

All procedures performed were approved by the applicable ethics committees and in accordance with applicable guidelines and regulations. The research reported here and the preparation of the current article were supported by Research Grant G.0766.11N of the Fund for Scientific Research (FWO — Flanders) awarded to Tom Beckers, Rudi D'Hooge, and Jan De Houwer; InterUniversity Attraction Pole Grant P7/33 of the Belgian Science Policy Office awarded to Tom Beckers and Jan De Houwer; and KU Leuven Program Funding Grant PF/10/005 awarded to Tom Beckers. Part of the data, results, and ideas contained in this article have been reported at the First Joint Portuguese-Spanish Meeting on Comparative Psychology in Braga, Portugal, September 2014. The authors thank Mark Haselgrove for his comments on earlier versions of the article and Zsuzsanna Callaerts-Vegh, Dennis Garlick, Hannelore Goddeyn, Kenneth Leising, and Jared Wang for their assistance in data collection.

Correspondence concerning this article should be addressed to Tom Beckers, Tiensestraat 102 Box 3712, 3000 Leuven, Belgium. E-mail: tom.beckers@kuleuven.be

A

Group	Phase 1	Phase 2	Test
Experimental	16 N+	8 LN+	L?
Control		8 LN+	L?

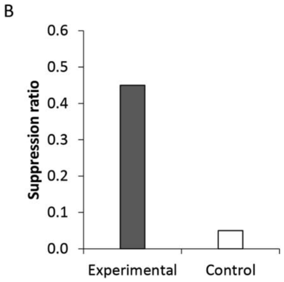

Figure 1. A. Design of Kamin's blocking experiment. The + represent the presentation of a footshock, N the presentation of a noise, and L the presentation of a light. B. Conditioned responding for the test trials for experimental and control group expressed as a suppression ratio. A suppression ratio of 0 corresponds to a strong conditioned response and a ratio of 0.5 corresponds to a complete lack of conditioned responding (after Kamin, 1969).

can be seen in Figure 1, Kamin found something else. Whereas the control group showed a strong CR to the light, the experimental group, which received conditioning to the noise before receiving pairings of the noise and light compound with shock, only showed a weak CR to the light. Prior conditioning with the noise appeared to have "blocked" conditioning to the light.

On the basis of this observation, Kamin suggested that surprise is critical for learning. In the experimental group, as a result of the first phase of training, the noise came to reliably predict the shock. Therefore, the shock was not surprising in the second phase, and learning about the light was said to be blocked. The role of surprise in learning has since been conceptualized in various ways (Holland, 1988). Perhaps most explicit was the formalization of surprise in the Rescorla-Wagner model of associative learning as prediction error (i.e., the discrepancy between expected and actual US occurrence; Kamin, 1969; Rescorla & Wagner, 1972). Since Rescorla and Wagner (1972) introduced the notion of prediction error, it has become a highly influential concept for the understanding of learning on behavioral, brain and even neuronal levels (e.g., Colombo, 2014; Corlett et al., 2004; den Ouden et al., 2009; Schultz & Dickinson, 2000; Steinberg et al., 2013; Tobler et al., 2006). Other theories have followed in its footsteps to suggest that surprise is necessary to engage association formation processes (e.g., Mackintosh, 1975; Pearce & Hall, 1980). Although most often explained in term of surprise, the blocking effect has also been explained by alternative theories of learning that do not ascribe a central role to prediction error in association formation, such as the comparator hypothesis (e.g., Miller & Matzel, 1988).

The capacity to explain blocking became nothing short of a touchstone for the validity of a theory of associative learning and until today, conditioning researchers use (variations of) the blocking effect to pit weaknesses and strengths of different theories of associative learning against each other (e.g., Blaisdell, Gunther, & Miller, 1999; Boddez, Baeyens, Hermans, & Beckers, 2011; Jones & Haselgrove, 2013; Mackintosh, 1971; Williams, 1996). Furthermore, blocking procedures have been used to distinguish between association-formation theories and rule-based or statistical theories of learning. For example, observations of the blocking effect in human causal learning tasks (Dickinson, Shanks, & Evenden, 1984) and social attribution tasks (van Overwalle & van Rooy, 2001) have been taken as support for the idea that associative processes play a crucial role in those phenomena (for a critical discussion of this argument see Boddez, De Houwer, & Beckers, in press).

The impact and importance of the blocking effect is further demonstrated by its omnipresence in the literature. Today, blocking has been reported using a wide variety of experimental procedures — for example appetitive and aversive learning protocols (e.g., Jennings & Kirkpatrick, 2006; Kamin, 1969), taste-aversion protocols (e.g., Willner, 1978), spatial learning (e.g., Rodrigo, Chamizo, McLaren, & Mackintosh, 1997), and human causal learning tasks (e.g., Dickinson et al., 1984) — and in a variety of species — including humans (e.g., Dickinson et al., 1984), rodents (e.g., Kamin, 1969), rabbits (e.g., Merchant & Moore, 1973), honey bees (e.g., Smith & Cobey, 1994), as well as mollusks (e.g., Sahley, Rudy, & Gelperin, 1981) and snails (e.g., Prados et al., 2013).

Some 10 years ago, we set out to establish a blocking procedure that would yield a robust blocking effect and could thus serve as a starting point to investigate the cognitive and neurobiological processes involved in blocking in rodents. Given our own previous successes in obtaining blocking effects (Beckers, Miller, De Houwer, & Urushihara, 2006; Wheeler, Beckers, & Miller, 2008), the abundance of publications reporting a blocking effect, and the importance attributed to it, we did not anticipate substantial problems in establishing such a procedure. Here, however, we report a series of 15 rodent experiments in which we tried but failed to obtain a robust blocking effect. Collectively, those 15 experiments represent the full record of all blocking studies in nonhuman animals executed or supervised by the last author since 2004.[1] In the first 14 of those experiments, various species, strains, experimental procedures, parameters, and set-ups were used. The procedures employed here were always based on published studies and while not identical to those previous reports in every detail, some came rather close to being exact replications (see Appendices A to E for a comparison between our protocols and the protocols used in published studies). Nonetheless, we either failed to obtain a blocking effect or when we found indications for a blocking effect were then unable to replicate that effect in follow-up experiments. Power analyses suggested that the absence of a blocking effect was not likely to be due to a lack of power (see Appendix P for details). Despite being relatively close replications, it cannot be excluded that we somehow deviated from previous studies that successfully demonstrated blocking in important ways. Therefore, we additionally performed a highly powered exact replication that adhered strictly to the protocol of a published study that did demonstrate a blocking effect; it did not yield a blocking effect either.

In all experiments described below, the experiments consisted of three phases (see Table 1). In the first phase (Elemental Training), animals in the blocking group received pairings of a Stimulus A with the US, while animals in the control group received an equal number of pairings of a different Stimulus B with the US. In the second phase (Compound Training), animals in both groups received presentations of stimulus compound AX followed by the US. In the third phase (Test), X was presented alone, without the

---

[1] Apart from two early studies for which crucial details could not be recovered (both of which were unsuccessful in producing blocking).

Table 1
General Design of the Experiments

Group	Elemental	Compound	Test
Experimental	A+	AX+	X
Control	B+	AX+	X

Note. The + represents the delivery of a US; A, B, and X represent different auditory and/or visual stimuli.

US. Importantly, while Kamin completely omitted Phase 1 training in the control group, we used a B control condition, which is regarded as the most appropriate of the control groups commonly used in blocking experiments (Arcediano, Escobar, & Matute, 2001; Taylor, Joseph, Balsam, & Bitterman, 2008; see the General Discussion for an elaborate discussion of the different control procedures used in blocking designs).

In what follows, the procedures and results of 15 experiments are described, subdivided in five series according to the species and strain of animals used and the general procedure applied. For each series, a general description of subjects, materials, and methods is provided first, followed by the specifics for each experiment and then the results. An overview of the stimuli used in all experiments can be found in Appendix F and an overview of the procedures in Appendix G. For comparability, we first present the results of classic frequentist analyses as typically reported in previous blocking studies. After the presentation of the five series of experiments and their results, we present the results of a Bayesian meta-analysis across all 15 studies. Then we turn to a discussion of what our consistent failure to obtain a solid blocking effect implies for theories of associative learning (be it association-formation theories, propositional theories or others) and for the reliability and replicability of psychological phenomena in general.

Series 1 (Experiments 1–4): Failure to Obtain Blocking in Aversive Conditioning in C57BL/6J Mice

Method

General overview. In four experiments, female and male mice were trained in a conditioned suppression procedure, in which interference with nose-poking for food pellets was measured to assess conditioned fear in food-deprived subjects. To our knowledge, no previous articles on blocking in mice had been published at the time those experiments took place (2004–2007). However, various studies (e.g., Jones & Gonzalez-Lima, 2001; Mackintosh, Dickinson, & Cotton, 1980) had reported successful blocking in rats with similar stimuli and a similar procedure as were employed in the present studies (see Appendix A for a comparison with those previous studies and Appendix F and G, respectively, for an overview of the stimuli and procedure used in this series).

Subjects. Subjects were experimentally naïve C57BL/6J mice obtained from our own breeding colony. The animals were housed in cages in a vivarium maintained on a 12-hr day/night cycle. The animals were allowed free access to water, whereas food availability was limited to 30 min per day following a progressive deprivation schedule initiated 1 week prior to the start of the study.

Experiment 1. Subjects were 24 male mice with body weights ranging from 19.1 g to 29.7 g before deprivation. During the experiment, three animals died of unknown cause. As a result, 11 mice remained in the experimental group and 10 mice in the control group.

Experiment 2. Subjects were 24 male mice with body weights ranging from 24.7 g to 30.0 g before deprivation. Both groups consisted of 12 mice.

Experiment 3. Subjects were 24 male mice with body weights ranging from 17.7 g to 21.7 g before deprivation. During the experiment, two animals died of unknown cause, leaving 11 mice in each groups.

Experiment 4. Subjects were 20 mice (15 females) with body weights ranging between 19.2 g to 24.7 g for females and between 26.7 g to 31.5 g for males at the beginning of the experiment. The experimental group consisted of 10 mice, all female. The subjects in this experiment were part of a larger cohort that also included mGLUR7 knock-out mice; the knock-out animal data are not reported here.

Apparatus. Four operant chambers (18 cm length × 18 cm width × 30 cm height; Coulbourn Instruments, Allentown, PA) were used in Experiments 1 and 2, while eight chambers were used in Experiments 3 and 4. In Experiment 4, those operant chambers were contained in isolation cubicles (Coulbourn Instruments, Allentown, PA); this was not the case in the other three experiments. All chambers had metal ceilings and side walls, and clear Plexiglas front and back walls. The floor was constructed of stainless steel grids (0.5 cm in diameter), through which a footshock could be delivered. In each chamber, there was a nose poke hole equipped with an infrared beam, which could be illuminated by means of a tri-light, and opposite to the nose poke hole was a recess. A food dispenser that could deliver 20-mg food pellets (Noyes precision pellets; Research Diets, New Brunswich, NJ) into the recess, which could be illuminated by a white light, was positioned on the opposite wall. Pellet delivery was indicated by the offset of the tri-light in the nose poke operandum and illumination of the food recess for 5 s. The enclosure could be illuminated by a house light. A speaker mounted on the wall was used to deliver tones with frequencies from 1,000 Hz to 3,500 Hz. All CSs were 30 s in duration (see Appendix F for an overview of the stimuli used in each experiment).

Experiment 1. The experiment was run in the dark. The flashing (0.5 s on/0.5 s off) and steady illumination of the house light served as Stimulus A and B, counterbalanced. A pulsing 3,500-Hz tone (200 ms on/200 ms off) served as Stimulus X. The US was a 0.5-s, 0.1-mA footshock.

Experiments 2. The house light remained on, unless it was flashing (0.2 s off/1.3 s on) to serve as Stimulus X. A steady 1,000-Hz tone and a complex, pulsing tone (1,500 Hz [0.1 s on/0.5 s off] and 2,500 Hz [0.5 s on/0.1 s off]) served as Stimulus A and B, counterbalanced. The US was the same as for Experiment 1.

Experiment 3. The stimulus parameters employed in this experiment were identical to Experiment 2, except for the US which was a 0.5-s, 0.2-mA footshock.

Experiment 4. The experiment was run in the dark. The flashing and steady illumination of the house light served as Stimulus A and B, as in Experiment 1. A tone served as Stimulus X. The US was a 0.5-s, 0.2-mA footshock.

Procedure. The experiments in this series were run in a mice lab at KU Leuven. The actual blocking training was preceded by a shaping phase to train the animals to nose pose for food. Each shaping session was 30 min long, while each training session was 25 min long. Appendix G provides an overview of the number of training days in each training phase for each experiment.

Shaping. Standard procedures were used to train the mice to nose poke for food pellets. A fixed time schedule, 120-s (FT 120-s), of noncontingent pellet delivery operated concurrently with a continuous reinforcement (CRF) schedule at the start of shaping; shaping ended on a variable interval (VI) 20-s (Experiment 1) or a VI 30-s (Experiments 2, 3, and 4) schedule. After shaping, the nose poke operandum was covered, except for Experiment 4, where pellets were delivered during all phases on a VI 30-s schedule.

Habituation. Animals in Experiment 3 were given 1 day of habituation training to X and animals in Experiment 4 were given 1 day of habituation training to A, B and X prior to the start of the experiment. No habituation to the CSs was provided in Experiments 1 and 2, but animals could habituate to the context during the shaping phase.

Phase 1: Elemental training. Experimental animals were exposed daily to three pairings of A with the US, with A and the US coterminating. The control animals received similar pairings of B with the US. The number of elemental training days differed for each experiment (see Table C1).

Phase 2: Compound training. All animals received daily three pairings of the AX compound with the US, with the compound and the US coterminating. Animals in Experiments 1, 2, and 3 received only one day of compound training. Animals in Experiment 4 received 4 days of compound training (see Table C1).

Reshaping. The nose poke holes were made accessible again (Experiments 1, 2, and 3), and all animals were retrained to nose poke for pellets.

Test. All animals were tested on X, which was presented four times during a 30-min session. Pellets were delivered on a VI 30-s schedule. No pellets were delivered during a 1-min period immediately preceding the CS (this preCS period was only 30 s in Experiment 4) and during the CS period.

Results

Data and analysis scripts for these and all following experiments are available at Open Science Framework (https://osf.io/fcwnr/?view_only=754693fa2907497a9ad8013a63813781). All statistical analyses were conducted with JASP version 0.7.1 (Love et al., 2015), which has a number of advantages over other statistical packages. First, JASP allows one to select a direction when conducting a one-tailed *t*-test, making the resulting *p*-values easier to interpret. Second, JASP allows one to calculate a Bayesian *t*-test (see Bayesian Analysis section below).

Mean suppression ratios were calculated as the mean number of nose pokes during presentations of X divided by the sum of this number and the mean number of responses per 30 s during the preCS period. Figure 2 depicts the mean suppression ratio across all presentations of X for experimental and control groups, for Experiments 1 to 4. The difference between the suppression ratio for the experimental group and the control group in Experiments 2 and 4 was numerically in line with a blocking effect; however, it failed to reach significance (smallest $p = .15$ [one-tailed]). No indications for a blocking effect were obtained in any of the other experiments in this series (see Appendix H for detailed statistics). To ascertain that the results were not due to differences in preCS responding between groups, we compared preCS responding before the first presentation of X between groups for all experiments. No baseline differences were observed in any of the four experiments (see Appendix I for detailed statistics). Power analyses on the basis of the effect sizes reported in the most similar published blocking studies (see Appendix P for details) suggested that the absence of a blocking effect was not due to a lack of power (estimated power >0.90 for all experiments).

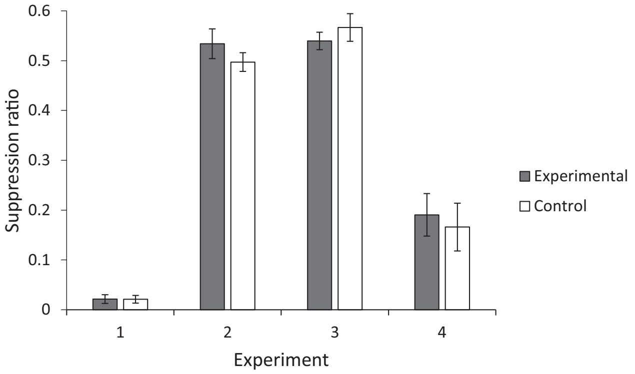

Figure 2. Mean suppression ratio at test across all presentations of X for experimental and control groups, for Experiments 1 to 4. Error bars represent standard errors of the means.

Series 2 (Experiments 5–6): Failure to Obtain Blocking in Aversive Conditioning in Long-Evans and Sprague-Dawley Rats

Method

General overview. In two experiments, female rats were trained in a conditioned suppression procedure, in which interference with lever-pressing for a sucrose solution was measured to assess conditioned fear in food-deprived subjects. The procedure of these experiments was highly similar to previously conducted studies that had demonstrated a blocking effect in which we were involved (Beckers et al., 2006; Wheeler et al., 2008; see Appendix B for a comparison with those previous studies and Appendix F and G, for an overview of the stimuli and procedure used in this series).

Subjects. Subjects were experimentally naïve, female rats obtained from Harlan (Indianapolis, IN). The animals were pair-housed in a vivarium maintained on a 14/10-hr dark/light cycle. The animals were allowed free access to water, whereas food availability was limited to 15 g per day per rat following a progressive deprivation schedule initiated 1 week prior to the start of the study. Experiments were conducted during the dark portion of the cycle. The subjects in these experiments were part of a larger experiment involving four groups. Animals in the other two groups were given subadditive pretraining before the start of blocking training, which has been reported to diminish blocking (Beckers et al., 2006), and were therefore not included in the analyses.

Experiment 5. Subjects were 24 (*ns* = 12) Long-Evans rats with body weights ranging from 225 g to 249 g before the start of the experiment.

Experiment 6. Subjects were 24 (*ns* = 12) Sprague–Dawley rats with body weights ranging from 225 g to 249 g before the start of the experiment.

Apparatus. Eight standard operant chambers (30 cm length × 25 cm width × 20 cm height; Med Associates, Georgia, VT) placed in sound- and light-resistant isolation cubicles were used (Med Associates, Georgia, VT). All chambers had Plexiglas front and back walls and ceilings, and aluminum side walls. The floors were constructed of stainless steel grids (0.5 cm in diameter), through which a 0.5-s, 0.5-mA footshock could be delivered. In each chamber, there was an operant lever, and adjacent to the lever was a recess. A water dipper could deliver 0.05 cc of a sucrose solution (20%) into a cup on the bottom of a recess. The enclosure was dimly illuminated by a house light. A diffuse light, placed on the opposite wall of the house light, was used to deliver a flashing light (0.25 s on/0.25 s off). The house light was turned off when the diffuse light was flashing. Two speakers, mounted on the outside walls of the chamber, were used to deliver a 3,000-Hz tone (70 dBA) and a white noise (70 dBA), respectively. A third speaker was used to deliver a 300-Hz tone (70 dBA) or a click train stimulus (four clicks/s, 70 dBA), which served as Stimuli A and B, counterbalanced. The 3,000-Hz tone and the flashing light served as Stimuli C and D, counterbalanced. The white noise served as Stimulus X. All CSs were 30 s in duration. For an overview of the stimuli used in each experiment, see Appendix F.

Procedure. The experiments in this series were run in a rat lab at the University of California, Los Angeles. The actual blocking training was preceded by a shaping phase to train the animals to lever press for the sucrose solution. Each session was 60 min long. Appendix G provides an overview of the number of training days in each training phase for each experiment.

Shaping. Standard procedures were used to train the rats to press the lever for the sucrose solution. A FT 120-s schedule operated concurrently with a CRF schedule at the start of shaping; shaping ended on a VI 30-s schedule.

Phase 1: Pretraining. During the 4 days of pretraining, all animals were exposed daily to two pairings of C with the US and one pairing of D with the US, with the stimulus and the US coterminating. The levers were retracted after shaping.

Phase 2: Elemental training. During the 3 days of elemental training, experimental animals were exposed daily to four pairings of A with the US, with A and the US coterminating. The control animals received similar pairings of B with the US.

Phase 3: Compound training. During a single compound training session, all animals were exposed to four pairings of the AX compound with the US, with the compound and the US coterminating.

Reshaping. The levers were inserted again and all animals were retrained to lever press for the sucrose solution.

Test. All animals were tested on X, which was presented four times, during a 30-min session. Sucrose solution was delivered on a VI 20-s schedule.

Results

Mean suppression ratios were calculated as above. Figure 3 presents the mean suppression ratio across all presentations of X for experimental and control groups, for Experiments 5 and 6. The difference between the suppression ratio for the experimental group and the control group in Experiment 5 was numerically in line with a blocking effect; however, it failed to reach significance, $t(22) = 1.45$, $p = .08$ (one-tailed). There was no indication whatsoever for a blocking effect in Experiment 6 (see Appendix H for detailed statistics). To ascertain that the results were not due to differences in preCS responding between groups, we compared preCS responding before the first presentation of X between groups for both experiments. No baseline differences were observed in any of the two experiments (see Appendix I for detailed statistics). Power analyses on the basis of the effect sizes reported in the most similar published blocking studies (see Appendix P for details) suggested that the absence of a blocking effect was not due to a lack of power (estimated power >.90 for both experiments).

Series 3 (Experiments 7–10): Failure to Obtain Blocking in Aversive Conditioning in Sprague-Dawley Rats

Method

General overview. In four experiments, female and male rats were trained in a conditioned suppression procedure, in which interference with lever-pressing for water was measured to assess conditioned fear in water-deprived subjects. The procedure of those experiments was again similar to previous reported studies that have demonstrated a blocking effect (Beckers et al., 2006; Blaisdell et al., 1999; Wheeler et al., 2008; see Appendix C for a comparison with those previous studies and Appendix F and G for an overview of the stimuli and procedure used in this series).

Subjects. Subjects were experimentally naïve, Sprague–Dawley rats obtained from Janvier (Le Genest-Saint-Isle, France). The animals were pair-housed in standard cages in a room with a 12-hr day-night cycle. The animals were allowed free access to food, whereas water availability was limited to 20 min per day

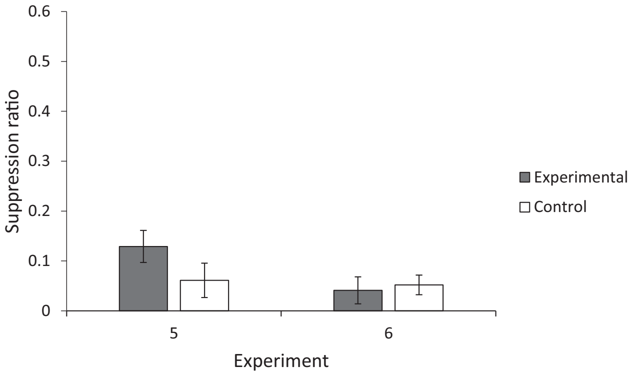

Figure 3. Mean suppression ratio at test across all presentations of X for experimental and control groups, for Experiments 5 and 6. Error bars represent standard errors of the means.

following a progressive deprivation schedule initiated one week prior to the start of the study.

Experiment 7. Subjects were eight (*ns* = 4) male rats with body weights ranging from 285 g to 310 g before deprivation.

Experiment 8. Subjects were eight (*ns* = 4) male rats with body weights ranging from 295 g to 330 g before deprivation.

Experiment 9. Subjects were eight (*ns* = 4) female rats with body weights ranging from 210 g to 225 g before deprivation.

Experiment 10. Subjects were eight (*ns* = 4) female rats with body weights ranging from 258 g to 270 g before deprivation.

Apparatus. Eight standard operant chambers (34 cm length × 33 cm width × 33 cm height; Coulbourn Instruments, Allentown, PA) housed in isolation cubicles (Med Associates Inc.) were used. All chambers had metal ceilings and side walls and clear Plexiglas front and back walls. The floors were constructed of stainless steel grids (0.5 cm in diameter), through which a 0.5-s footshock could be delivered. In each chamber, there was an operant lever, and adjacent to the lever was a recess. A water dipper could deliver 0.04 cc of water into a cup on the bottom of the recess. Water delivery was indicated by the onset of a white noise for 0.5 s and the illumination of the recess for 1 s. The enclosure was dimly illuminated by a house light. Two speakers were mounted on two different interior walls. One speaker was used to deliver a white noise (83 dBC in Experiments 7 and 8; 79 dBC in Experiments 9 and 10). The other speaker was used to deliver a 1,000-Hz (79 dBC) or 3,000-Hz tone (87 dBC), which served as Stimulus A and B, counterbalanced. A clicker mounted on each environmental chest was able to deliver a clicking sound (5 clicks/s) and served as Stimulus X (85 dBC in Experiment 7 and 80 dBC in Experiments 8, 9, and 10). All CSs were 30 s in duration. For an overview of the stimuli used in each experiment, see Appendix F.

Procedure. The experiments in this series were run in a rat lab at KU Leuven. The actual blocking training was preceded by a shaping phase to train the animals to lever press for water. Each session was 60 min long. Appendix G provides an overview of the number of training days in each training phase for each experiment.

Shaping. Standard procedures were used to train the rats to press the lever for water. A FT 120-s schedule operated concurrently with a CRF schedule at the start of shaping in Experiments 7, 8, and 9, while the FT 120-s schedule operated alone on the first day of shaping in Experiment 10; shaping ended on a VI 20-s schedule. For Experiments 7, 8, and 9, the levers were retracted after shaping.

Phase 1: Elemental training. During the 3 elemental training days, experimental animals were exposed daily to four pairings of A with the US, with A and the US coterminating. The control animals received similar pairings of B with the US.

Phase 2: Compound training. During a single compound training session, all animals received four pairings of the AX compound with the US, with the compound and the US coterminating.

Reshaping. The levers were inserted again (Experiment 7, 8, and 9) and the animals were retrained to lever press for water.

Test. All animals were tested on X, which was presented three times (Experiments 7 and 8) or four times (Experiments 9 and 10), during a 30-min session. Pellets were delivered on a VI 20-s schedule.

Results

Mean suppression ratios were calculated as above. Figure 4 depicts the mean suppression ratio across the first three presentations of X for experimental and control groups, for Experiment 7 to 10. The difference between the suppression ratio for the experimental group and the control group in Experiments 7 and 8 was numerically in line with a blocking effect; however, it failed to reach significance (smallest p = .10 [one-tailed]). However, it is clear from Figure 4 that not a hint of a blocking effect was observed in Experiments 9 and 10 (see Appendix H for detailed statistics). Therefore, a trend for a blocking effect was observed in two experiments, but we were not able to replicate the effect in two further experiments using similar procedures. To ascertain that the results were not due to differences in preCS responding between groups, we compared preCS responding before the first presentation of X between groups for all experiments. No baseline differences were observed in any of the four experiments (see Appendix I for detailed statistics). Despite the rather small sample sizes, power analyses on the basis of the effect sizes reported in the most similar published blocking studies (see Appendix P for details) suggested that the absence of a blocking effect was not likely to be due to a lack of power (estimated power >.70 for all experiments).

Series 4 (Experiments 11–14): Failure to Obtain Blocking in Appetitive Conditioning in Sprague-Dawley Rats

Method

General overview. In four experiments, female rats were trained in a Pavlovian approach procedure, in which magazine entries during stimulus presentation were measured as an index of food expectancy. The procedure of these experiments was again similar to previous reported studies that have demonstrated a blocking effect (Holland, 1999; Taylor et al., 2008; see Appendix D for a comparison with those previous studies and Appendix F and G, for an overview of the stimuli and procedure used in this series).

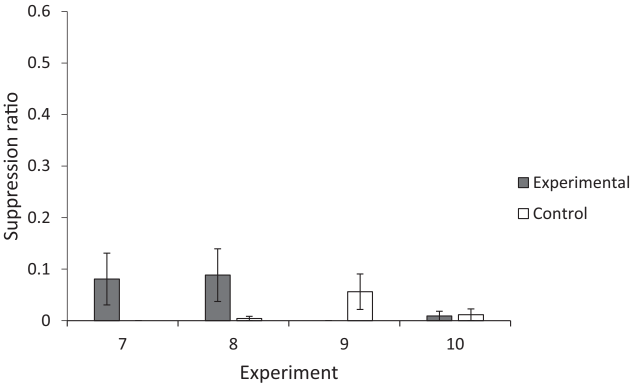

Figure 4. Mean suppression ratio at test across the first three presentations of X for experimental and control groups, for Experiments 7 to 10. Error bars represent standard errors of the means. Mean and *SD* of the control group in Experiments 7 were both 0.

Subjects. Subjects were experimentally naïve, female Sprague–Dawley rats obtained from Janvier (Le Genest-Saint-Isle, France). The animals were housed in a vivarium maintained on a 12-h day-night cycle. The animals were allowed free access to water, whereas food availability was limited to minimum 12 g per rat per day following a progressive deprivation schedule initiated 1 week prior to the start of the study.

Experiments 11 and 12. Subjects were 12 (*ns* = 6) female rats with body weights ranging from 206 g to 268 g for Experiment 11 and from 212 g to 248 g for Experiment 12, before deprivation. Animals were housed in groups of four.

Experiments 13 and 14. Subjects were 24 (*ns* = 12) female rats with body weights ranging from 160 g to 227 g before deprivation for Experiment 13 and from 185 g to 218 g for Experiment 14. Animals were housed in groups of six. The subjects in these experiments were part of a larger experiment involving four groups. Animals in the other two groups were given subadditive pretraining before the start of blocking training, which has been shown to diminish blocking (Beckers et al., 2006), and were therefore not included in the analyses.

Apparatus. Twelve standard operant chambers (34 cm length × 33 cm width × 33 cm height; Coulbourn Instruments, Allentown, PA) housed in sound- and light-resistant cubicles (Coulbourn Instruments, Allentown, PA) were used. All chambers had metal ceilings and side walls and clear Plexiglas front and back walls. The floors were constructed of stainless steel grids (0.5 cm in diameter). Each chamber was equipped with a food dispenser that could deliver 45-mg sucrose pellets (TestDiet, St. Louis, MO) into a recess, which could be illuminated by a white light. Pellet delivery was indicated by the illumination of the recess for 0.5 s. A photocell sensor placed in the recess was able to detect head entries. The enclosure was dimly illuminated by a house light. A light bulb, mounted on the opposite wall of the house light, was used to deliver a flashing light (0.1 s on/0.1 s off). The house light was turned off when the other light was flashing. Three speakers were mounted on two different interior walls. One was used to deliver a 1,000-Hz tone, a 7,000-Hz tone, or a 11,000-Hz tone. The second one was also used to deliver a 7,000-Hz tone. The third one delivered a white noise. A clicker was able to deliver a clicking sound (five clicks/s). All CSs were 10 s in duration. For an overview of the stimuli used in each experiment, see Appendix F.

Experiment 11. The clicker (72 dBC), the white noise (83 dBC), and a 1,000-Hz tone (90 dBC) served as A, B, and X, semicounterbalanced using the Latin square method.

Experiment 12. A pulsing 1,000-Hz tone (0.2 s on/0.2 s off, 73 dBC) and a pulsing 7,000-Hz tone (0.5 s on/0.1 s off, 68 dBC), both delivered through the same speaker, served as A and B, counterbalanced. The clicker (72 dBC) served as Stimulus X.

Experiment 13. The buzzer (77 dBC) and the flashing light were used as C and D, counterbalanced. A pulsing 1,000-Hz tone (0.2 s on/0.2 s off, 73 dBC) and a pulsing 7,000-Hz tone (0.5 s on/0.1 s off, 68 dBC), emerging from different speakers, served as A and B, counterbalanced. The clicker (72 dBC) served as Stimulus X.

Experiment 14. The buzzer and the flashing light were used as C and D, counterbalanced. A 1,000-Hz tone (73 dBC) and the clicker (72 dBC) served as A and B, counterbalanced. The white noise (69 dBC) served as Stimulus X and an 11,000-Hz tone (61 dBC) served as Stimulus T.

Procedure. The experiments in this series were run in a rat lab at KU Leuven. Each training session was 60 min long. Appendix G provides an overview of the number of training days in each training phase for each experiment.

Magazine training. All rats initially received 30 sucrose pellets during a 40-min session.

Pretraining. In Experiment 13 and 14, animals were exposed daily to 14 pairings of C with the US and seven pairing of D with the US, with the stimulus ending in delivery of a food pellet. No pretraining was given in Experiments 11 and 12.

Phase 1: Elemental training. In Experiments 11, 12, and 13, animals in the experimental group were exposed daily to 20 pairings of A with the US, with the stimulus ending in delivery of two food pellets. Animals in the control group received similar pairings of B with the US. In Experiment 14, animals in the experimental group were exposed daily to 16 pairings of A with the US and four unreinforced presentations of Stimulus T. The animals in the control group received 16 pairings of B with the US as well as four unreinforced presentations of Stimulus T. The number of elemental training days differed for each experiment (see Appendix G).

Phase 2: Compound training. During a single compound training session, all animals received 20 pairings of the AX compound with the US.

Test. All animals were tested on X, which was presented 10 (Experiments 11, 12, and 13, session duration 35 min) or 20 (Experiment 14, session duration 60 min) times.

Results

An elevation score for the number of head entries per trial was calculated as the mean number of head entries during each presentation of X minus the mean number of head entries during the 10-s period immediately preceding the CS. Figure 5 depicts the mean elevation score across the first 10 trials, for Experiments 11 to 14. The difference between the elevation score for the experimental group and the control group in Experiments 11, 12, and 13 was numerically in line with a blocking effect; however, it failed to reach significance (smallest $p = .19$ [one-tailed]). However, it is clear from Figure 5 that not a hint of a blocking effect was observed in Experiment 14 (see Appendix H for detailed statistics).

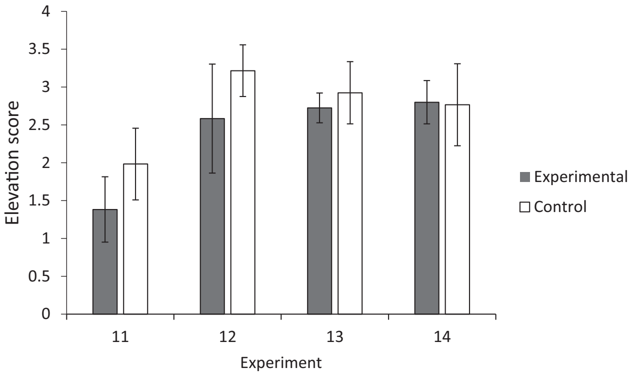

Figure 5. Mean elevation score across the first 10 presentations of X for experimental and control groups, for Experiments 11 to 14. Error bars represent standard errors of the means.

To ascertain that the results were not due to differences in preCS responding between groups, we compared preCS responding before the first presentation of X between groups for all experiments. No baseline differences were observed in any of the four experiments (see Appendix I for detailed statistics). Power analyses on the basis of the effect sizes reported in the most similar published blocking studies (see Appendix P for details) suggest that the absence of a blocking effect was not due to a lack of power (estimated power = .70 for Experiments 11 and 12 and >0.90 for Experiment 13 and 14).

Series 5 (Experiment 15): Failure to Obtain Blocking in Appetitive Conditioning in CRL:CD Rats in a Preregistered Exact Replication

Method

General overview. In light of our consistent failure to obtain blocking using a variety of procedures that were inspired by but not fully identical to previous reports, in a final experiment we conducted a highly powered exact replication of a blocking effect reported in the literature. In consideration of the restrictions imposed by the equipment we had available and the strains of rats readily available to us, we decided to replicate Conditions 1 and 2 (a regular blocking condition and its control) of Taylor, Joseph, Balsam, and Bitterman (2008, Experiment 3). In this experiment, male rats were trained in a Pavlovian approach procedure, in which magazine entries during stimulus presentation were measured as an index of food expectancy. The protocol was screened and approved by the acting editor of the journal prior to the start of the study, and preregistered online at https://osf.io/f3uxm/?view_only=fe87f3cd67234810a5dc25e7cdb5377f. More detailed information about the procedure is also provided there; see Appendix F and G, respectively, for an overview of the stimuli and procedure used in this experiment. The sample size of the study was determined such as to yield an estimated power of more than .90 based on the effect size reported by Taylor et al. (2008) (see Appendix P for details).

Subjects. Subjects were 60 (*ns* = 30) experimentally naïve, male CRL:CD rats obtained from Charles River Laboratories (Saint Germain Nuelles, France). The animals were housed in a vivarium maintained on a 12-hr day-night cycle. The animals were allowed free access to water, whereas food availability was limited to minimum 1 hr per day following a progressive deprivation schedule initiated 1 week prior to the start of the study.

Apparatus. The same 12 operant chambers and cubicles were used as for the previous series. Each chamber was equipped with a food dispenser that could deliver two 45-mg nonpurified grain-based pellets (TestDiet, St. Louis, MO) into a recess. A photocell sensor placed in the recess was able to detect head entries. The enclosure was dimly illuminated by a red house light. A light bulb, mounted above and to the left of the recess, was used to present Stimulus X. Two speakers, mounted on two different interior walls, were used to present a 1,000-Hz tone (80 dbC) and a white noise (80 dBC), which served as Stimuli A and B, counterbalanced. All CSs were 12 s in duration. For an overview of the stimuli used, see Appendix F.

Procedure. The experiments in this series were run in a rat lab at KU Leuven. Each training session was 35 min long. Appendix G provides an overview of the number of training days in each training phase.

Magazine training. Reinforcement was delivered on a variable time 45-s schedule for 3 days.

Phase 1: Elemental training. During the 15 days of elemental training, all animals in the experimental group were exposed daily to 10 pairings of A with the US, with the stimulus ending in delivery of two food pellets. Animals in the control group received similar pairings of B with the US.

Phase 2: Compound training. During the 5 compound training days, all animals received 10 pairings of the AX compound with the US, with similar specifics as for elemental training.

Test. Three test sessions were held on successive days, with X presented on the first day, A on the second, and B on the third day. Each test session included 10 trials, on each of which the stimulus to be tested was presented without reinforcement for 12 s. After conducting those test sessions, we decided to add a fourth day of testing, which was not conducted by Taylor et al. (2008) and not included in our preregistered replication protocol, to evaluate whether a difference in responding might emerge over further testing. During this final test session X was presented using the same parameters as for the previous test sessions.

Acquisition X. Given the lack of conditioned responding to X during test, we decided to evaluate whether X was able to support conditioning at all. All animals received 10 pairings of X with the US, with similar parameters as for the elemental training of A and B. This additional acquisition phase was also not conducted by Taylor et al. (2008) and not included in our preregistered replication protocol.

Results

Mean elevation scores per trial were calculated for X, A, and B as the mean number of head entries during each presentation of the CS (X, A, or B) minus the mean number of head entries during the 12-s period immediately preceding the CS. Figure 6 depicts the mean elevation scores across the 10 presentations of each CS during the three test sessions for experimental and control groups. It is clear from Figure 6 that not a hint of a blocking effect was observed, despite the fact that, as expected and in line with Taylor et al. (2008), elevation scores to A were higher in the experimental

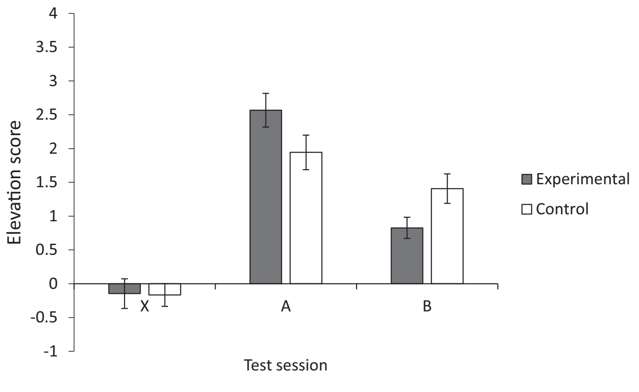

Figure 6. Mean elevation scores across the 10 presentations of each CS during the three test sessions for experimental and control groups, for Experiment 15. Error bars represent standard errors of the means.

group than in the control group and elevation scores to B were lower in the experimental group than the control group (see Appendix H for detailed statistics). To ascertain that the results were not due to differences in preCS responding between groups, we compared preCS responding before the first presentation of X, A, and B between groups. No baseline differences were observed in any of the three test sessions (see Appendix I for detailed statistics). Whereas with a Cohen's *d* of 0.81, the blocking effect reported by Taylor et al. (2008) was somewhat smaller in size than that of many other reported blocking effects, the much larger than average sample size ensured a power of over .90 to detect such effect (see Appendix P for details).

We did not observe a significant difference between the experimental and control group in the additional test session for X either ($M_{\text{experimental}} = -0.00$, $SD_{\text{experimental}} = 0.19$, $M_{\text{control}} = -0.00$, $SD_{\text{control}} = 0.14$), $t(58) < 0.01$, $p = .50$, $d < 0.01$, $BF_{10} = 0.26$.

In light of the surprising lack of blocking, we ran an additional training session, which was not included in our preregistered replication protocol, to check whether animals did notice X and were able to learn about X. After 1 day of acquisition with X, mean number of head entries during X ($M = 2.48$; $SD = 1.44$) was higher than mean number of head entries during the preX interval ($M = 1.93$; $SD = 1.28$), $t(59) = -3.76$, $p < .01$, $d = -0.49$, $BF_{10} = 62$. Mean elevation score across all trials did not differ between the experimental ($M = 0.58$; $SD = 1.20$) and control group ($M = 0.51$; $SD = 1.06$), $t(58) = 0.24$, $p = .81$, $d = 0.06$, $BF_{10} = 0.27$.

Altogether, the results suggest that X, although perfectly capable of supporting conditioning, was overshadowed to the same extent by A and B (despite A eliciting more responding in the blocking than in the control group). Overshadowing is the observation that conditioned responding to an elementally tested CS (i.e., the overshadowed cue; in this case X) is weaker when it was trained in compound with another CS (i.e., the overshadowing cue; in this case A), as compared with when it was trained elementally (Pavlov, 1927, p. 141). So, although we observed reduced responding to X this was not a consequence of the A + training provided to the animals in the experimental group and hence, not a true blocking effect.

Bayesian Analysis

Frequentist statistical techniques do not allow to infer the absence of an effect (Wagenmakers, 2007), making it difficult to draw conclusions regarding the lack of blocking from the analyses presented so far. In Bayesian hypothesis testing, however, a Bayes Factor (BF) can be calculated that quantifies the strength of the relative statistical evidence for two rivaling hypotheses. A BF quantifies the relative probability of the data under, for example, the null versus the alternative hypothesis (Dienes, 2011; Gallistel, 2009; Morey, 2015; Rouder, Speckman, Sun, Morey, & Iverson, 2009). If a BF of about 1 is obtained, there is no evidence in favor of either one of the hypotheses; the more the BF exceeds 1, the more evidence is obtained for the data under the hypothesis in the nominator, compared with the hypothesis in the denominator. The reverse holds for BFs below 1. According to Jeffreys (1961), BFs above 3 can be regarded to provide substantial evidence in favor of the hypothesis that is in the nominator (or, conversely, values below 0.33 provide substantial evidence for the hypothesis in the denominator). In order to calculate the BF one needs to specify a prior distribution of the probability of the different effect sizes under each hypothesis before considering the data (Dienes, 2011).

We calculated BF_{10} (i.e., the probability of the data given the null hypothesis divided by the probability of the data given the alternative hypothesis) for all experiments reported in the current article using JAPS 0.7.1 (Love et al., 2015) and assuming a default prior distribution (Ly, Verhagen, & Wagenmakers, in press).[2] None of the experiments yielded a BF_{10} above 3 (see Appendix H for detailed statistics), indicating that none of the experiments provided substantial evidence for the alternative hypothesis — that is for the presence of a blocking effect. Four experiments (Experiments 3, 6, 9, and 15) provided substantial evidence for the absence of a blocking effect (BF_{10} below 0.33). The remaining experiments yielded at best anecdotal evidence for either of the hypotheses, with BFs between 0.33 and 3.

To evaluate the overall evidence provided by the data, we next computed a meta-analytic BF (MABF) using the BayesFactor package (Morey & Rouder, 2015) in R (R Core Team, 2015). A MABF can be interpreted in much the same way as a regular BF, that is it expresses the relative probability of observing the data under the two competing hypotheses after observing the data across *all* the reported experiments (see Rouder, & Morey, 2011 for more details on MABFs). The computed $MABF_{10}$ was equal to 0.13, providing substantial evidence for the null hypothesis over the alternative hypothesis. Of importance, the MABF provided stronger support for the null hypothesis than any of the individual experiments.

One might argue that the observation of blocking in Experiments 2, 3, and 15 was hampered by a floor effect — if the control group is hardly responding to X, lower responding in the experimental group cannot be expected. To exclude that the results of the Bayesian meta-analysis were driven mainly by the latter experiments, we repeated the Bayesian meta-analysis without them. A $MABF_{10}$ of 0.16 was obtained, indicating that even when excluding the potential influence of floor effects, we find substantial evidence in favor of the null hypothesis.

One might similarly argue that the observation of blocking in Experiments 1 and 6 to 10 was hampered by a ceiling effect. It is not clear that that is a valid argument, because a strong response in the control group is to be expected (animals in this group are assumed to learn about X) and the stronger the response in the control group, the more room there should be for observing a reduction in that response in the experimental group. The observation that also in the experimental group conditioned responding is very strong in those experiments suggests that the animals learned about X in those groups as well, which — by its very definition — constitutes the absence of a blocking effect. Nevertheless, we also conducted a meta-analysis that included Experi-

---

[2] We performed robustness analyses to evaluate to what extent our findings were influenced by our choice of a default prior distribution on the effect size (Wagenmakers, Wetzels, Borsboom, & van der Maas, 2011, Online Appendix). The results of those analyses are available on https://osf.io/fcwnr/?view_only=754693fa2907497a9ad8013a63813781. With wider priors (i.e., assigning a higher prior probability to effect sizes further removed from zero), the relative evidence in favor of the null hypothesis further increased.

ments 4, 5, and 11–14 only (thereby omitting all experiments that could possibly suffer from floor or ceiling effects) and still obtained substantial evidence in favor of the null hypothesis ($MABF_{10}$ of 0.20).

General Discussion

The purpose of the series of experiments reported here was to establish a blocking procedure that would produce robust effects and could thus serve as a starting point to investigate the cognitive processes involved in blocking. The abundance of publications reporting a blocking effect and the importance attributed to it, suggest that it is a robust phenomenon. Yet in 15 experiments in which we used procedures similar or identical to previously published studies that demonstrated a blocking effect, not a single significant blocking effect was obtained using one-tailed tests and a rejection criterion of $p < .05$. Power analyses moreover indicate that it is unlikely that the consistent absence of a blocking effect can be attributed to a lack of power (see Appendix P for details). Even when we replicated a published report in great detail, no indication of a blocking effect was observed, despite an estimated power of well above .90. A trend toward significance ($p < .10$) was observed in two experiments (Experiments 5 and 8), but Bayesian analysis suggests that even in those studies, evidence for blocking was no more than anecdotal, and in subsequent studies, we were never able to replicate those near-significant effects. For four experiments (Experiments 3, 6, 9, and 15), Bayesian analysis provided substantial evidence for the null hypothesis. Moreover, a meta-analytic BF provided substantial evidence in favor of the null hypothesis across all 15 experiments.

Importantly, the failures to generate a blocking effect reported in the current article were not limited to one specific lab or experimental set-up: The first series was run in a mice lab at KU Leuven, the second series was run in a rat lab at the University of California, Los Angeles, and the third, fourth, and fifth series were run in a rat lab at KU Leuven. Nor were they limited to one specific procedure: We tried conditioned suppression of nose poking and lever pressing as well as an appetitive procedure. Although we are not the first to report problems in obtaining a blocking effect (Taylor et al., 2008; see further), the current report is by far the most extensive series of failures to find blocking in the literature. While from the abundance of successful studies published, it appears undisputable that blocking is a genuine and important phenomenon, our results do raise doubts regarding the canonical nature of the blocking effect. The current series of failures suggests that blocking is a highly parameter-dependent phenomenon. This is especially highlighted by the fact that we failed to replicate the near-significant effects of Experiments 5 and 8 in highly similar follow-up studies. In a final attempt to obtain blocking (Experiment 15), we followed a published report to the letter, yet a blocking effect was once again not observed. Admittedly, the protocol employed in this experiment might differ from the one by Taylor et al. (2008) in parameters that were not mentioned in their report because they were deemed irrelevant. Yet such an interpretation would imply not only that blocking is a highly parameter-dependent phenomenon, but also that we lack knowledge on which parameters are significant and deserve reporting. In what follows, we will discuss in detail the validity of our results and the implications of our findings for psychological science in general and the associative learning field in particular.

It is one thing to observe that we failed to find evidence for a blocking effect across a series of 15 experiments, but quite another to determine what the cause for those failures might be. Despite the fact that we adhered to protocols and procedures described in previous reports in which a blocking effect was demonstrated, we may have somehow failed to fulfill crucial boundary conditions. Theoretical accounts for blocking may offer clues regarding potential boundary conditions. Many association-formation models of associative learning postulate that surprise is essential for learning to occur (e.g., Mackintosh, 1975; Pearce & Hall, 1980; Rescorla & Wagner, 1972; see also Introduction). If the presence or absence of a US on a given trial is surprising, the associative strength of the stimuli presented on that trial will change (increase if a US is surprisingly present, decrease if a US is surprisingly absent). Associative strength in turn determines the strength of the CR (possibly in combination with other factors; e.g., Rescorla, 2001). Therefore, in order for X to acquire less associative strength in the experimental group than in the control group and blocking to be observed, two factors are crucial according to those models: (a) the US should not be surprising at the beginning of the compound phase in the experimental group, because only then will X be prevented from gaining associative strength; and (b) the US should be surprising at the beginning of the compound phase in the control group, because only then will X be able to gain associative strength. In order for those two conditions to be met (a) the blocking Stimulus A should gain significant associative strength over the course of elemental training (otherwise the first condition will not be met); and (b) generalization of associative strength from B to A should be limited such that the US is not fully predicted on the first presentations of AX in the control group, allowing X (and A) to gain more associative strength during compound training in the control group than in the experimental group.

To check whether those boundary conditions were met in our experiments, we analyzed the training data of Experiments 4 and 10 to 15 (in the other experiments, Pavlovian training was performed offline, so training data are not available). The available data from the elemental phase show that conditioned responding to the elemental stimulus (A or B; expressed as suppression ratios for Experiments 4 and 10 and elevation scores for Experiments 11 to 15) was significantly higher by the end of the phase than at the start (see Appendix J for detailed statistics).[3] Moreover, responding seemed to be at asymptote at the end of elemental training, indicated by (a) the absence of a linear trend over trials on the last day of elemental training; and (b) the absence of a difference in conditioned responding between the first and last trial of the last day of elemental training (see Appendix K and L for detailed statistics). The procedures used in experiments that were conducted offline were similar to the procedures of one or more of the other experiments. Therefore, we can be relatively confident that the first hypothetical boundary condition for generating a blocking effect that can be derived from theoretical models of associative

---

[3] For Experiment 4 only session level data were available. So, rather than comparing CRs to the first and the last presentation of A, CRs during the first session were compared with CRs during the last session.

learning (i.e., that the US was predicted by the blocking Stimulus A at the end of elemental phase) was met.[4] To check whether the second hypothetical boundary condition was met, one can look at the difference in responding between the experimental and control condition on the first AX presentation of the compound phase. The rationale behind this comparison is that if there is more responding to the first AX presentation in the experimental group than in the control group, there cannot have been full generalization from B to A, thus leaving more room for X to gain associative strength on AX trials in the control group than in the experimental group. Of course, this is an imperfect test of generalization, because responding to AX in the control group can be influenced by a number of factors other than conditioned generalization, such as an orienting response to AX (AX is more novel in the control group than in the experimental group). Arguably, then, this test might overestimate the degree of generalization from B to A. A better alternative would be to compare responding to A and B directly in both groups, as we did in Experiment 15. In this experiment, experimental and control animals differed in responding to A (higher responding in the experimental than in the control group) as well as B (higher responding in the control than in the experimental group), yet no difference in conditioned responding to the first AX presentation was observed between groups. That observation supports the idea that comparing responding to the first AX presentation might overestimate generalization. Moreover, at least in Experiment 15, we have strong evidence against a generalization account of our results.

The question remains whether B to A generalization could account for the lack of blocking in the other experiments. Two observations are relevant for this question. First, across Experiments 10 to 15 conditioned responding (expressed as suppression ratios for Experiment 10 and elevation scores for Experiments 11 to 15) on the first trial of the compound phase was numerically higher in the experimental group than in the control group in all but one of the experiments (Experiment 12), although the differences between the groups were never significant (see Appendix M for statistical details; relevant data were not available for the other experiments). Note, however, that this trend was observed even though responding on the first AX trial probably overestimates generalization. Overall, those results suggest, at the very least, that generalization was not complete. Second, conditioned responding to X was substantial. If generalization from B to A was strong in some of our experiments, the absence of a blocking effect in those experiments might be the result of X not being able to gain associative strength in the control group (because X would be blocked by A as the result of generalization from preceding B+ training). In that case, low conditioned responding to X should be observed at test in both groups because A would block learning of the X-US relation in both conditions. However, across Experiments 10 to 15 conditioned responding (expressed as suppression ratios for Experiment 10 and elevation scores for Experiments 11 to 15) for the first presentation of X at test was never significantly less than conditioned responding to the last presentation of A for either of the groups (see Appendix N for statistical details; relevant data were not available for the other experiments). In fact, conditioned responding to X was often even numerically higher than final conditioned responding to A (exceptions here are Experiments 12 and 15). Hence, whereas a generalization account would predict blocking due to an absence of X-US learning in both the control and blocking groups, we observed strong evidence for X-US learning in both groups. So, all in all, it seems unlikely that the lack of blocking that we observed was due to excessive generalization from B to A. Moreover, given that we observed a difference in responding to both A and B between groups in Experiment 15, we can be confident that at least in this experiment the failure to observe blocking was not a consequence of excessive generalization from B to A.

It is important to note that the two potential boundary conditions discussed above follow naturally from the assumptions entailed by many association-formation models. However, given that most published reports of a blocking effect do not contain training data, there is no way of knowing whether those boundary conditions were actually met in previous research and thus whether there is empirical evidence for those boundary conditions. For instance, it is not clear from earlier research whether a difference in conditioned responding to AX between the experimental and control groups at the start of training is necessary or even helpful to observe blocking. It would be important to establish those (or any other) boundary conditions empirically because they are not a theoretical necessity. That is, some theoretical accounts of the blocking effect (e.g., Beckers et al., 2006; Miller & Matzel, 1988) do not yield the same boundary conditions. In general, little has been said in the literature about boundary conditions for blocking. Our results are important if only for that reason, because they clearly suggest that the blocking effect is indeed dependent on (a variety of) boundary conditions, the exact nature of which is yet to be determined.

The question remains then how to reconcile the relative abundance of demonstrations of blocking in the literature with our consistent failure to observe the effect under various conditions. We believe that at least part of that conundrum may be linked to the use of suboptimal control conditions in a number of published reports, that is, control conditions that do not allow to conclude that the observed between-groups difference is actually a true blocking effect. We are not the first to raise this issue (Arcediano, Escobar, & Matute, 2001). Kamin, and others after him, simply omitted elemental training altogether in the control group (e.g., Allen, Padilla, Myers, & Gluck, 2002; Feldman, 1975; Kamin, 1969; Solomon, 1977). However, the difference between groups in exposure to the US in such a design is highly problematic; weaker conditioned responding to X in the experimental group than in the control group can then simply reflect differences in habituation to the US or any other sort of mere exposure effect (Prados et al., 2013). To avoid this problem, researchers have been using unsignalled US presentations in the control group, in different ways (random presentations of A and the US during the elemental phase, backward AB+/A+ training or simply replacing the elemental training by presentations of the US; e.g., Holland & Gallagher, 1993; Parker, 1986; Wagner, 1969). The problem is that responding during test has been proven to be enhanced in such control groups relative to conditions that receive compound AX+ training only, perhaps due to contextual reinforcement during training

---

[4] In Experiment 4, suppression ratios did not exceed 0.15 from the second day onward. This suggests that, although elemental training was shorter in Experiments 1, 2, and 3, the blocking cue A would have asymptotically predicted the presence of the US at the end of elemental training in those experiments as well.

(Taylor et al., 2008). Other researchers have replaced elemental training by discrimination training (A+/B− in the experimental group and B+/A− in the control group; e.g., Dopson, Pearce, & Haselgrove, 2009). However, such designs empirically conflate blocking and reduced overshadowing: Higher responding to X in the control group than in the experimental group may reflect enhanced responding in the former rather than reduced responding in the latter, relative to mere compound training. That is, a reduction of overshadowing may be observed if the overshadowing cue is preexposed without reinforcement (e.g., De Houwer, Beckers, & Glautier, 2002). In a within-subjects variant of this paradigm, A+/B− training is followed by AX+/BY+ training (e.g., Rescorla, 1999; Waelti, Dickinson, & Schultz, 2001). Here as well, the observation of a difference between X and Y can be the result of a true blocking effect (reduced responding to X due to A+ training), a reduction of overshadowing (heightened responding to Y due to B− training), or a combination of both. Of importance, existing theories of learning may be differentially compatible with each of those phenomena. For instance, the Rescorla-Wagner model readily predicts true blocking but has no mechanism to account for reduced overshadowing. Therefore, if apparent blocking effects in the literature are driven in whole or in part by the use of suboptimal control groups, it may affect how well they support various theories of learning.

The control procedure used in the current experiments (B+ training followed by AX+ training) equates experience with the US between groups and avoids the possibility that between-groups differences reflect reduced overshadowing rather than true blocking. It has been claimed, however, that an apparent blocking effect in such a procedure might reflect heightened responding to X in the control group rather than diminished responding to X in the blocking group, on the assumption that more generalization to X takes place from n B pairings, with n the number of elemental training trials, and m A pairings, with m the number of compound training trials, (as the control group receives) than from $n + m$ A pairings (Blaser et al., 2006). This possibility was refuted when Taylor et al. (2008) observed equal responding to X in an X-absent "blocking" group (which received mere elemental pairings of A with the US) compared with an X-absent "control" group (which first received elemental pairings of B with the US and then elemental pairings of A with the US). In conclusion, the control procedure used in the current experiments is to be regarded as the most appropriate of the control groups commonly used in blocking experiments (Arcediano, Escobar, & Matute, 2001; Taylor et al., 2008).

To reiterate, we do not want to dispute that the blocking effect exists. As indicated in the introduction, some of us had performed or been involved with successful blocking studies in rodents before embarking on the present series of failures. The full set of all the blocking experiments in rodents that any of the authors of the current article have ever executed, supervised or otherwise been involved with consists of the experiments reported here plus the experiments reported by Beckers et al. (2006) and Wheeler et al. (2008). This set can therefore be taken as an unbiased data-set, devoid of publication bias. If we repeat the Bayesian meta-analysis reported above but now including the 11 blocking effects contained in those earlier reports (excluding a few experimental and control groups that were specifically designed not to yield a blocking effect, such as the subadditive pretraining groups in Beckers et al., 2006, Experiment 1; for details, see Appendix O), we obtain a MABF$_{10}$ >100, which clearly indicates that blocking is a real phenomenon (see https://osf.io/fcwnr/?view_only=754693fa2907497a9ad8013a63813781 for more details).

The results presented here do suggest that a true blocking effect is more difficult to obtain than one might assume from the literature and that we lack insight into its boundary conditions. An imprecise use of the term "blocking," which conflates a number of other effects, perhaps in combination with some degree of publication bias may have led to an overestimation of how robust and reliable true blocking is and an underestimation of the importance of potential boundary conditions for its observation. Indeed, given the canonical nature of the blocking effect in textbooks of learning, many researchers may have been dissuaded from publishing failures to obtain blocking. Some may have been dissuaded from pursuing the effect upon failure. Others may have continued until they obtained a robust and reliable effect, but have likely neglected to systematically examine and document relevant variables and boundary conditions for obtaining the effect, in light of the general view that the blocking effect is not subject to such conditions. As a result, we may be left with a biased perspective regarding the universality of blocking. This situation is perhaps not much different from what appears to be the case for other seemingly established phenomena in the psychological literature. Indeed, we believe that blocking is not unique in being over promoted. Other phenomena within and beyond the domain of associative learning may be more elusive than their status suggests as well. Consider, as an example, social priming effects, that is, observations that activation of relevant stereotypes, attitudes, traits and goals in one context can influence the behavior in another context unconsciously. While evidence has been reported for a variety of such effects (e.g., individuals can be implicitly primed to walk faster or slower (e.g., Bargh, Chen, & Burrows, 1996) or to behave more or less intelligently (Dijksterhuis & van Knippenberg, 1998), some researchers have recently started to question the mere existence of social priming (e.g., Shanks et al., 2013). Rather than attempting to address this issue as a yes-or-no question, other researchers have started to focus on the conditions under which a social priming effect can be observed and the mechanisms that mediate its occurrence (e.g., Bargh, 2006; Doyen, Klein, Pichon, & Cleeremans, 2012; Gomes & McCullough, 2015; Shariff & Norenzayan, 2015), an avenue, we argue, that would be fruitful for the blocking effect as well.

In closing, we have no doubt that true blocking exists, that many existing reports of the effect are genuine, and that many of those reports may be replicable in the hands of the very researchers that obtained them. However, the misuse of the term blocking and publication bias might have mislead us in our perception of how robust, reliable and general of a phenomenon blocking is. Our results suggest that several boundary conditions might need to be fulfilled in order to observe blocking. We therefore argue that blocking, rather than being a touchstone for our theories of elementary learning, should be the subject of further investigation. Some studies have begun to explore procedural variables that are important for blocking to occur (Arcediano, Escobar, & Miller, 2004; Blaser et al., 2006; Feldman, 1975; Janisewicz & Baxter, 2003; LoLordo, Jacobs, & Foree, 1982; Pineño et al., 2005; Taylor et al., 2008). However, it is clear that many moderators are unknown, as the current series of failures illustrates. In tandem with a functional approach that is directed at uncovering the moderators, further research may aim to decipher the dynamic cognitive processes that might underlie blocking (Boddez, Haesen, Baeyens, & Beckers, 2014). Through a combined increase in procedural and process level knowledge (see De Houwer, 2011, for a theoretical discussion on the benefits of combining functional and cognitive approaches to psychology), blocking will perhaps become a less central effect in theories of learning, but hopefully also a less elusive and more amenable one. Meanwhile, the above should serve as a cautionary tale that the canonical status of a phenomenon in psychological science and its widespread inclusion in handbooks of psychology should not be taken as a proxy for its empirical reliability and robustness.

References

Allen, M. T., Padilla, Y., Myers, C. E., & Gluck, M. A. (2002). Selective hippocampal lesions disrupt a novel cue effect but fail to eliminate blocking in rabbit eyeblink conditioning. *Cognitive, Affective & Behavioral Neuroscience, 2,* 318–328. http://dx.doi.org/10.3758/CABN.2.4.318

Arcediano, F., Escobar, M., & Matute, H. (2001). Reversal from blocking in humans as a result of posttraining extinction of the blocking stimulus. *Animal Learning & Behavior, 29,* 354–366. http://dx.doi.org/10.3758/BF03192901

Arcediano, F., Escobar, M., & Miller, R. R. (2004). Is stimulus competition an acquisition deficit or a performance deficit? *Psychonomic Bulletin & Review, 11,* 1105–1110. http://dx.doi.org/10.3758/BF03196744

Bargh, J. A. (2006). What have we been priming all these years? On the development, mechanisms, and ecology of nonconscious social behavior. *European Journal of Social Psychology, 36,* 147–168. http://dx.doi.org/10.1002/ejsp.336

Bargh, J. A., Chen, M., & Burrows, L. (1996). Automaticity of social behavior: Direct effects of trait construct and stereotype-activation on action. *Journal of Personality and Social Psychology, 71,* 230–244. http://dx.doi.org/10.1037/0022-3514.71.2.230

Beckers, T., Miller, R. R., De Houwer, J., & Urushihara, K. (2006). Reasoning rats: Forward blocking in Pavlovian animal conditioning is sensitive to constraints of causal inference. *Journal of Experimental Psychology: General, 135,* 92–102. http://dx.doi.org/10.1037/0096-3445.135.1.92

Blaisdell, A. P., Gunther, L. M., & Miller, R. R. (1999). Recovery from blocking achieved by extinguishing the blocking CS. *Animal Learning & Behavior, 27,* 63–76. http://dx.doi.org/10.3758/BF03199432

Blaser, R. E., Couvillon, P., & Bitterman, M. E. (2006). Blocking and pseudoblocking: New control experiments with honeybees. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 59,* 68–76. http://dx.doi.org/10.1080/17470210500242938

Boddez, Y., Baeyens, F., Hermans, D., & Beckers, T. (2011). The hide-and-seek of retrospective revaluation: Recovery from blocking is context dependent in human causal learning. *Journal of Experimental Psychology: Animal Behavior Processes, 37,* 230–240. http://dx.doi.org/10.1037/a0021460

Boddez, Y., De Houwer, J., & Beckers, T. (in press). The inferential reasoning theory of causal learning: Towards a multi-process propositional account. In M. Waldmann (Ed.), *Oxford handbook of causal reasoning*. Oxford, UK: Oxford University Press.

Boddez, Y., Haesen, K., Baeyens, F., & Beckers, T. (2014). Selectivity in associative learning: A cognitive stage framework for blocking and cue competition phenomena. *Frontiers in Psychology, 5,* 1305. http://dx.doi.org/10.3389/fpsyg.2014.01305

Bush, R. R., & Mosteller, F. (1951). A mathematical model for simple learning. *Psychological Review, 58,* 313–323. http://dx.doi.org/10.1037/h0054388

Colombo, M. (2014). Deep and beautiful. The reward prediction error hypothesis of dopamine. *Studies in History and Philosophy of Biological and Biomedical Sciences, 45,* 57–67. http://dx.doi.org/10.1016/j.shpsc.2013.10.006

Corlett, P. R., Aitken, M. R. F., Dickinson, A., Shanks, D. R., Honey, G. D., Honey, R. A., . . . Fletcher, P. C. (2004). Prediction error during retrospective revaluation of causal associations in humans: FMRI evidence in favor of an associative model of learning. *Neuron, 44,* 877–888.

De Houwer, J., Beckers, T., & Glautier, S. (2002). Outcome and cue properties modulate blocking. *The Quarterly Journal of Experimental Psychology, 55,* 965–985. http://dx.doi.org/10.1080/02724980143000578

Delamater, A. R., & Lattal, K. M. (2014). The study of associative learning: Mapping from psychological to neural levels of analysis. *Neurobiology of Learning and Memory, 108,* 1–4. http://dx.doi.org/10.1016/j.nlm.2013.12.006

den Ouden, H. E. M., Friston, K. J., Daw, N. D., McIntosh, A. R., & Stephan, K. E. (2009). A dual role for prediction error in associative learning. *Cerebral Cortex, 19,* 1175–1185. http://dx.doi.org/10.1093/cercor/bhn161

Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *The Quarterly Journal of Experimental Psychology, 36,* 29–50. http://www.tandfonline.com/doi/abs/10.1080/14640748408401502

Dienes, Z. (2011). Bayesian versus orthodox statistics: Which side are you on? *Perspectives on Psychological Science, 6,* 274–290. http://dx.doi.org/10.1177/1745691611406920

Dijksterhuis, A., & van Knippenberg, A. (1998). The relation between perception and behavior, or how to win a game of trivial pursuit. *Journal of Personality and Social Psychology, 74,* 865–877. http://dx.doi.org/10.1037/0022-3514.74.4.865

Dopson, J. C., Pearce, J. M., & Haselgrove, M. (2009). Failure of retrospective revaluation to influence blocking. *Journal of Experimental Psychology: Animal Behavior Processes, 35,* 473–484. http://dx.doi.org/10.1037/a0014907

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PLoS ONE, 7,* e29081. http://dx.doi.org/10.1371/journal.pone.0029081

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,* 175–191. http://dx.doi.org/10.3758/BF03193146

Feldman, J. M. (1975). Blocking as a function of added cue intensity. *Animal Learning & Behavior, 3,* 98–102. http://dx.doi.org/10.3758/BF03209108

Gallistel, C. R. (2009). The importance of proving the null. *Psychological Review, 116,* 439–453. http://dx.doi.org/10.1037/a0015251

Gomes, C. M., & McCullough, M. E. (2015). The effects of implicit religious primes on dictator game allocations: A preregistered replication experiment. *Journal of Experimental Psychology: General, 144,* e94–e104. http://dx.doi.org/10.1037/xge0000027

Holland, P. C. (1988). Excitation and inhibition in unblocking. *Journal of Experimental Psychology: Animal Behavior Processes, 14,* 261–279. http://dx.doi.org/10.1037/0097-7403.14.3.261

Holland, P. C. (1999). Overshadowing and blocking as acquisition deficits: No recovery after extinction of overshadowing or blocking cues. *The Quarterly Journal of Experimental Psychology, 52,* 307–333. http://dx.doi.org/10.1080/027249999393022

Holland, P. C., & Gallagher, M. (1993). Effects of amygdala central nucleus lesions on blocking and unblocking. *Behavioral Neuroscience, 107,* 235–245. http://dx.doi.org/10.1037/0735-7044.107.2.235

Houwer, J. D. (2011). Why the cognitive approach in psychology would profit from a functional approach and vice versa. *Perspectives on Psychological Science, 6,* 202–209. http://dx.doi.org/10.1177/1745691611400238

Janisewicz, A. M., & Baxter, M. G. (2003). Transfer effects and conditional learning in rats with selective lesions of medial septal/diagonal band cholinergic neurons. *Behavioral Neuroscience, 117,* 1342–1352. http://dx.doi.org/10.1037/0735-7044.117.6.1342

Jeffreys, H. (1961). *The theory of probability*. Oxford, UK: Oxford University Press.

Jennings, D., & Kirkpatrick, K. (2006). Interval duration effects on blocking in appetitive conditioning. *Behavioural Processes, 71,* 318–329. http://dx.doi.org/10.1016/j.beproc.2005.11.007

Jones, D., & Gonzalez-Lima, F. (2001). Mapping Pavlovian conditioning effects on the brain: Blocking, contiguity, and excitatory effects. *Journal of Neurophysiology, 86,* 809–823.

Jones, P. M., & Haselgrove, M. (2013). Blocking and associability change. *Journal of Experimental Psychology: Animal Behavior Processes, 39,* 249–258. http://dx.doi.org/10.1037/a0032254

Kamin, L. J. (1969). Predictability, surprise, attention, and conditioning. In B. A. Cambell & R. M. Church (Eds.), *Punishment and aversive behavior* (pp. 279–296) New York, NY: Appleton-Century-Crofts.

LoLordo, V. M., Jacobs, W. J., & Foree, D. D. (1982). Failure to block control by a relevant stimulus. *Animal Learning & Behavior, 10,* 183–192. http://dx.doi.org/10.3758/BF03212268

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., . . . Wagenmakers, E. J. (2015). JASP (Version 0.7.1) [Computer software]. Retrieved from https://jasp-stats.org/download/

Ly, A., Verhagen, A. J., & Wagenmakers, E.-J. (in press). Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology. *Journal of Mathematical Psychology*.

Mackintosh, N. J. (1971). An analysis of overshadowing and blocking. *The Quarterly Journal of Experimental Psychology, 23,* 118–125. http://dx.doi.org/10.1080/00335557143000121

Mackintosh, N. J. (1975). A theory of attention: Variations in the associability of stimuli with reinforcement. *Psychological Review, 82,* 276–298. http://dx.doi.org/10.1037/h0076778

Mackintosh, N. J., Dickinson, A., & Cotton, M. M. (1980). Surprise and blocking: Effects of the number of compound trials. *Animal Learning & Behavior, 8,* 387–391. http://dx.doi.org/10.3758/BF03199622

Merchant, H. G., III, & Moore, J. W. (1973). Blocking of the rabbit's conditioned nictitating membrane response in Kamin's two-stage paradigm. *Journal of Experimental Psychology, 101,* 155–158. http://dx.doi.org/10.1037/h0035784

Miller, R. R., & Matzel, L. D. (1988). The comparator hypothesis: A response rule for the expression of associations. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 22, pp. 51–92). San Diego, CA: Academic Press. http://dx.doi.org/10.1016/S0079-7421(08)60038-9

Morey, R. D. (2015). *JASP and Bayes: A fresh way to do statistics* [PowerPoint slides]. Retrieved from https://static.jasp-stats.org/presentations/August2015/RM/#1

Morey, R. D., & Rouder, J. N. (2015). *BayesFactor: Computation of Bayes factors for common designs. R package version 0.9.12–2*. Retrieved from http://CRAN.R-project.org/package=BayesFactor

Parker, L. A. (1986). Drug-specific blocking of lithium-, amphetamine-, and apomorphine-induced conditioned flavor avoidance. *Learning and Motivation, 17,* 91–113. http://dx.doi.org/10.1016/0023-9690(86)90022-6

Pavlov, I. P. (1927). *Conditioned reflexes, an investigation of the physiological activity of the cerebral cortex*. London, UK: Oxford University Press.

Pearce, J. M., & Hall, G. (1980). A model for Pavlovian learning: Variations in the effectiveness of conditioned but not of unconditioned stimuli. *Psychological Review, 87,* 532–552. http://dx.doi.org/10.1037/0033-295X.87.6.532

Pineño, O., Urushihara, K., & Miller, R. R. (2005). Spontaneous recovery from forward and backward blocking. *Journal of Experimental Psychology: Animal Behavior Processes, 31,* 172–183. http://dx.doi.org/10.1037/0097-7403.31.2.172

Prados, J., Alvarez, B., Acebes, F., Loy, I., Sansa, J., & Moreno-Fernández, M. M. (2013). Blocking in rats, humans and snails using a within-subjects design. *Behavioural Processes, 100,* 23–31.

R Core Team. (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org/

Rescorla, R. A. (1999). Learning about qualitatively different outcomes during a blocking procedure. *Animal Learning & Behavior, 27,* 140–151. http://dx.doi.org/10.3758/BF03199671

Rescorla, R. A. (2001). Are associative changes in acquisition and extinction negatively accelerated? *Journal of Experimental Psychology: Animal Behavior Processes, 27,* 307–315. http://dx.doi.org/10.1037/0097-7403.27.4.307

Rescorla, R. A., & Wagner, A. R. (1972). A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black & W. F. Prokasy (Eds.), *Classical conditioning II: current research and theory* (pp. 64–99) New York, NY: Appleton-Century-Crofts.

Rodrigo, T., Chamizo, V. D., McLaren, I. P. L., & Mackintosh, N. J. (1997). Blocking in the spatial domain. *Journal of Experimental Psychology: Animal Behavior Processes, 23,* 110–118. http://dx.doi.org/10.1037/0097-7403.23.1.110

Rouder, J. N., & Morey, R. D. (2011). A Bayes factor meta-analysis of Bem's ESP claim. *Psychonomic Bulletin & Review, 18,* 682–689. http://dx.doi.org/10.3758/s13423-011-0088-7

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16,* 225–237. http://dx.doi.org/10.3758/PBR.16.2.225

Sahley, C., Rudy, J., & Gelperin, A. (1981). An analysis of associative learning in a terrestrial mollusc. *Journal of Comparative Physiology*. A, Neuroethology, Sensory, Neural, and Behavioral Physiology, *144,* 1–8. http://dx.doi.org/10.1007/BF00612791

Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience, 23,* 473–500. http://dx.doi.org/10.1146/annurev.neuro.23.1.473

Shanks, D. R., Newell, B. R., Lee, E. H., Balakrishnan, D., Ekelund, L., Cenac, Z., . . . Moore, C. (2013). Priming intelligent behavior: An elusive phenomenon. *PLoS ONE, 8,* e56515–e56515. http://dx.doi.org/10.1371/journal.pone.0056515

Shariff, A. F., & Norenzayan, A. (2015). A question of reliability or of boundary conditions? Comment on Gomes and McCullough. *Journal of Experimental Psychology: General, 144,* e105–e106. http://dx.doi.org/10.1037/xge0000111

Smith, B. H., & Cobey, S. (1994). The olfactory memory of the honeybee Apis mellifera. II. Blocking between odorants in binary mixtures. *The Journal of Experimental Biology, 195,* 91–108. http://jeb.biologists.org/content/195/1/91.short

Solomon, P. R. (1977). Role of the hippocampus in blocking and conditioned inhibition of the rabbit's nictitating membrane response. *Journal of Comparative and Physiological Psychology, 91,* 407–417. http://dx.doi.org/10.1037/h0077330

Steinberg, E. E., Keiflin, R., Boivin, J. R., Witten, I. B., Deisseroth, K., & Janak, P. H. (2013). A causal link between prediction errors, dopamine neurons and learning. *Nature Neuroscience, 16,* 966–973. http://dx.doi.org/10.1038/nn.3413

Taylor, K. M., Joseph, V. T., Balsam, P. D., & Bitterman, M. E. (2008). Target-absent controls in blocking experiments with rats. *Learning & Behavior, 36,* 145–148. http://dx.doi.org/10.3758/LB.36.2.145

Thalheimer, W., & Cook, S. (2002). *How to calculate effect sizes from published research articles: A simplified methodology*. Retrieved from http://work-learning.com/effect_sizes.htm

Tobler, P. N., O'doherty, J. P., Dolan, R. J., & Schultz, W. (2006). Human neural learning depends on reward prediction errors in the blocking paradigm. *Journal of Neurophysiology, 95,* 301–310. http://dx.doi.org/10.1152/jn.00762.2005

van Overwalle, F., & van Rooy, D. (2001). How one cause discounts or augments another: A Connectionist account of causal competition. *Personality and Social Psychology Bulletin, 27,* 1613–1626. http://dx.doi.org/10.1177/01461672012712005

Waelti, P., Dickinson, A., & Schultz, W. (2001). Dopamine responses comply with basic assumptions of formal learning theory. *Nature, 412,* 43–48. http://dx.doi.org/10.1038/35083500

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review, 14,* 779–804. http://dx.doi.org/10.3758/BF03194105

Wagenmakers, E.-J., Wetzels, R., Borsboom, D., & van der Maas, H. L. J. (2011). Online appendix for "Why psychologists must change the way they analyze their data: the case of psi:" A robustness analysis. *Journal of Personality and Social Psychology, 100,* 426–432. http://dx.doi.org/10.1037/a0022790

Wagner, A. R. (1969). Stimulus validity and stimulus selection in associative learning. In N. J. Mackintosh & W. K. Honig (Eds.), *Fundamental issues in associative learning* (pp. 90–122). Halifax, Nova Scotia, Canada: Dalhousie University Press.

Wheeler, D. S., Beckers, T., & Miller, R. R. (2008). The effect of subadditive pretraining on blocking: Limits on generalization. *Learning & Behavior, 36,* 341–351. http://dx.doi.org/10.3758/LB.36.4.341

Williams, B. A. (1996). Evidence that blocking is due to associative deficit: Blocking history affects the degree of subsequent associative competition. *Psychonomic Bulletin & Review, 3,* 71–74. http://dx.doi.org/10.3758/BF03210742

Willner, J. A. (1978). Blocking of a taste aversion by prior pairings of exteroceptive stimuli with illness. *Learning and Motivation, 9,* 125–140. http://dx.doi.org/10.1016/0023-9690(78)90016-4

Appendix A

Comparison of Experimental Design and Procedure Employed in Series 1 to Similar Published Studies

Features	Experiments 1–4	Jones & Gonzalez-Lima, 2001	Mackintosh, Dickinson, & Cotton, 1980
Animals	C57BL/6J mice	Male Long-Evans black-hooded rats	Female hooded Lister rats
Control condition	B control	Unpaired CS and US presentations	B control
N blocking	At least 10	6	12
N control	At least 10	7	12
Procedure	Aversive	Aversive	Aversive
Dependent variable	SR of nose pokes	SR of drinking behavior and freezing	SR of licking response
US	.1–.2 mA; see Table F	.75-s, .5-mA	.5-s, .75-mA
A and B	Tones and lights; see Table F	A: Two flashing white lights B: Not used	A: Overhead light; B: Flashing light (not counterbalanced)
X	Tone or light; see Table F	Low-frequency FM tone	1,800-Hz tone
Duration CS	10 s	15 s	60 s
# elemental pairings	Between 6 and 54; see Table G	16	4
# compound pairings	Between 3 and 18; see Table G	12	2
# test trials	3	3	5

Note. SR = Suppression ratio.

(*Appendices continue*)

Appendix B

Comparison of Experimental Design and Procedure Employed in Series 2 to Similar Published Studies

Features	Experiments 5–6	Beckers et al., 2006	Wheeler, Beckers, & Miller, 2008
Animals	Female Long-Evans (Experiment 5)/Female Sprague-Dawley (Experiment 6)	Male Sprague-Dawley	Male Sprague-Dawley
Control condition	B control	B control	B control
N blocking	12	12	12
N control	12	12	12
Procedure	Aversive	Aversive	Aversive
Dependent variable	SR of lever pressing	SR of lever pressing	SR of lever pressing
US	.5-s, .5-mA	.5-s, .7-mA	.5-s, .7-mA
A and B	300-Hz tone/Clicker	300-Hz tone/1,900-Hz tone	Low complex tone/1,900-Hz tone
X	White noise	Clicker	Clicker
Duration CS	30 s	30 s	30 s
# elemental pairings	12	12	12
# compound pairings	4	4	4
# test trials	4	4	4
Pretraining procedure	2 C+/D+	C+/D+/E+ or 2 C+/DE+	2 C+/D+
# pretraining pairings	12	12	12

Note. SR = Suppression ratio.

Appendix C

Comparison of Experimental Design and Procedure Employed in Series 3 to Similar Published Studies

Features	Experiments 7–10	Blaisdell et al., 1999
Animals	Male (Experiments 7 & 8) and female (Experiments 9 & 10) Sprague-Dawley rats	Male and female Sprague-Dawley rats
Control condition	B control	B control
N blocking	4	6
N control	4	6
Procedure	Aversive	Aversive
Dependent variable	SR of lever pressing	Mean times to lick for 5 cumulative seconds in the presence of X
US	.5-s, .7-mA	0.5-s, 1.0-mA shock
A and B	1,000-Hz tone/3,000-Hz tone	Complex tone (3,000 and 3,200 Hz)/White noise
X	Clicker	Clicker
Duration CS	30 s	10 s
# elemental pairings	12	12
# compound pairings	4	4
# test trials	3 or 4	Presentation of X for a maximum of 15 min

Note. SR = Suppression ratio; The procedures from Beckers et al. (2006) and Wheeler et al. (2008), summarized in Table B, are very similar as well.

(*Appendices continue*)

Appendix D

Comparison of Experimental Design and Procedure Employed in Series 4 to Similar Published Studies

Features	Experiments 11–14	Holland, 1999, Experiment 6	Taylor et al., 2008
Animals	Female Sprague-Dawley rats	Female Sprague-Dawley	Male CRL:CD rats
Control condition	B control	B control	B control
N blocking	6 (Experiments 11 & 12)/12 (Experiments 13–14)	8	14
N control	6 (Experiments 11 & 12)/12 (Experiments 13–14)	8	14
Procedure	Appetitive	Appetitive	Appetitive
Dependent variable	Elevation of head entries	Food cup behavior	Number of head entries
US	Sucrose pellet	food pellet	2 food pellets
A and B	Tone/clicker/White noise; see Table F	A: Noise; B: Clicker; not counterbalanced	1,000-Hz tone/White noise
X	Tone/clicker/White noise; see Table F	Light	Bright light
Duration CS	10 s	10 s	12 s
# elemental pairings	Between 60 and 100; see Table F	64	150
# compound pairings	20	32	50
# test trials	10	8	10

Note. SR = Suppression ratio.

Appendix E

Comparison of Experimental Design and Procedure Employed in Series 5 to Similar Published Studies

Features	Experiment 15	Taylor et al., 2008
Animals	Male CRL:CD rats	Male CRL:CD rats
Control condition	B control	B control
N blocking	30	14
N control	30	14
Procedure	Appetitive	Appetitive
Dependent variable	Elevation of head entries	Number of head entries
US	2 food pellets	2 food pellets
A and B	1,000-Hz tone/White noise	1,000-Hz tone/White noise
X	Bright light	Bright light
Duration CS	12 s	12 s
# elemental pairings	150	150
# compound pairings	50	50
# test trials	10	10

(*Appendices continue*)

Appendix F

Stimuli Used in Experiments 1 to 15

Experiment	C/D	A/B	X	T	US
1		Flashing light/Steady light	Pulsing 3,500-Hz tone		.1-mA shock
2		1,000 Hz-tone/Pulsing tone (1,500 Hz and 2,500 Hz)	Flashing light		.1-mA shock
3		1,000-Hz tone/Pulsing tone (1,500 Hz and 2,500 Hz)	Flashing light		.2-mA shock
4		Flashing light/Steady light	Tone		.2-mA shock
5–6	3,000-Hz tone/Flashing light	300-Hz tone/Clicker	White noise		.5-mA shock
7–10		1,000-Hz tone/3,000-Hz tone	Clicker		.7-mA shock
11		Clicker /White noise/1,000-Hz tone		Sucrose pellet	
12		Pulsing 1,000-Hz/Pulsing 7,000-Hz tone	Clicker		Sucrose pellet
13	Buzzer/Flashing light	Pulsing 1,000-Hz/Pulsing 7.000-Hz tone	Clicker		Sucrose pellet
14	Buzzer/Flashing light	1,000-Hz tone/Clicker	White noise	11,000-Hz tone	Sucrose pellet
15		1,000-Hz tone/White noise	Bright light		2 grain-based pellets

Appendix G

Total Number of Trials in Each Phase

Experiment	Habituation to CS	Pretraining	Phase 1: Elemental	Phase 2: Compound
1			6 (2)	3 (1)
2			18 (6[a])	3 (1)
3	3 (1)[b]		16 (4)	3 (1)
4	3 (1)[c]		54 (18)	12 (4)
5–6		12 (4)[d]	12 (3)	4 (1)
7–10			12 (3)	4 (1)
11			60 (3)	20 (1)
12			80 (4)	20 (1)
13		105 (5)[d]	80 (4)	20 (1)
14		252 (12)[d]	100 (5)	20 (1)
15			150 (10)	50

Note. Number of days in each phase between brackets.

[a] Rats received 2 days of reshaping between the fifth and sixth day of elemental training. [b] Habituation to X. [c] Habituation to A, B, and X. [d] Pretraining consisted out of C+ and D+ presentations with twice as many C as D trials.

(*Appendices continue*)

Appendix H

Descriptives and Results of One-Tailed Independent Parametrical and Bayesian *t*-Tests Conducted on Test Data (Obtained With JASP 0.7.1, Love et al., 2015)

Experiment	Measure	$N_{\text{experimental}}$	N_{control}	$M_{\text{experimental}}$	M_{control}	$SD_{\text{experimental}}$	SD_{control}	*t*-value	*df*	*p*-value	*d*	BF_{10}
1	SR	11	10	.021	.021	.029	.024	.037	19	.485	.016	.401
2	SR	12	12	.534	.497	.103	.065	1.048	22	.153	.428	.899
3	SR	11	11	.540	.567	.058	.092	−.819	20	.789	−.349	.244
4	SR	10	10	.190	.173	.135	.137	.287	18	.389	.128	.486
5	SR	12	12	.129	.061	.111	.119	1.448	22	.081	.591	1.396
6	SR	12	12	.041	.052	.094	.068	−.330	22	.628	−.135	.303
7	SR	4	4	.081	.000	.100	.000	1.610	3	.103	1.138	1.749
8	SR	4	4	.088	.004	.102	.008	1.645	3.04	.099	1.163	1.801
9	SR	4	4	.000	.056	.000	.069	−1.628	3	.899	−1.151	.283
10	SR	4	4	.009	.011	.018	.023	−.162	6	.562	−.115	.477
11	ES	6	6	1.383	1.983	1.057	1.158	−.937	10	.185	−.541	.939
12	ES	6	6	2.583	3.217	1.763	.835	−.795	10	.223	−.459	.831
13	ES	12	12	2.725	2.925	.681	1.422	−.439	15.79	.333	−.179	.516
14	ES	12	12	2.800	2.767	.991	1.874	.054	16.71	.521	.022	.360
15:X	ES	30	30	−.147	−.167	1.198	.917	.073	58	.529	.019	.249
15:A	ES	30	30	2.567	1.943	1.362	1.403	1.746	58	.043	.451	1.756
15:B	ES	30	30	.827	1.407	.854	1.190	−2.169	58	.017	−.560	3.553

Note. SR = Suppression ratio; ES = Elevation score; *d* = Cohen's *d*.

Appendix I

Descriptives and Results of Two-Tailed Independent Parametrical and Bayesian *t*-Tests Conducted With JASP 0.7.1 (Love et al., 2015) on Responding During First preCS Interval at Test

Experiment	Measure	$N_{\text{experimental}}$	N_{control}	$M_{\text{experimental}}$	M_{control}	$SD_{\text{experimental}}$	SD_{control}	*t*-value	*df*	*p*-value	*d*	BF_{10}
1	NP	11	10	20.910	22.500	8.348	9.009	−.420	19	.679	−.184	.417
2	NP	12	12	29.500	44.083	16.11	24.92	−1.702	22	.103	−.695	1.038
3	NP	11	11	39.364	32.546	17.862	13.663	1.006	20	.327	.429	.551
4°	NP	10	10	5.013	7.942	5.447	5.976	−1.145	18	.267	−.512	.627
5	LP	12	12	11.333	9.750	5.211	4.224	.818	22	.422	.334	.476
6	LP	12	12	26.00	27.083	14.013	11.526	−.207	22	.838	−.084	.379
7	LP	4	4	17.250	14.750	5.909	8.057	.500	6	.635	.354	.564
8	LP	4	4	26.750	38.000	16.153	36.579	−.563	6	.594	−.398	.575
9	LP	4	4	18.500	17.250	11.846	9.845	.162	6	.876	.115	.527
10	LP	4	4	33.000	23.750	7.703	13.401	1.197	6	.276	.846	.776
11	HE	6	6	.833	1.000	1.602	2.000	−.159	10	.877	−.092	.471
12	HE	6	6	1.333	.667	1.966	.816	.767	10	.461	.443	.563
13	HE	12	12	.500	.250	1.000	.866	.655	22	.519	.267	.436
14	HE	12	12	1.500	1.833	2.316	2.552	−.335	22	.741	−.137	.389
15:X	HE	30	30	1.733	2.433	2.477	2.909	−1.004	58	.320	−.259	.400
15:A	HE	30	30	1.300	1.767	2.136	2.487	−.780	58	.439	−.201	.339
15:B	HE	30	30	1.600	.100	2.541	.403	3.194	30.46	.003	.825	15.64

Note. NP = Nose pokes; LP = Lever presses; HE = Head entries; *d* = Cohen's *d*.
° For Experiment 4, the means and standard deviations are calculated over the test session because trial-level information was not available.

(*Appendices continue*)

Appendix J

Descriptives and Results of Pairwise Parametrical and Bayesian *t*-Tests Conducted With JASP 0.7.1 (Love et al., 2015) Comparing Conditioned Responding to First and Last Presentation of A or B in Elemental Phase

Experiment	Measure	*N*	M_{first}	M_{last}	SD_{first}	SD_{last}	*t*-value	*df*	*p*-value	*d*	BF_{10}
4°	SR	20	.436	.067	.176	.096	8.443	19	$<.001$	1.888	>100
10	SR	8	.135	.000	.137	.000	2.791	7	.027	.987	3.035
11	ES	12	−.333	4.167	1.497	2.552	−6.051	11	$<.001$	−1.747	>100
12	ES	12	−1.083	2.917	1.975	4.461	−2.561	11	.026	−.739	2.716
13	ES	24	−.708	3.333	2.074	2.973	−6.276	23	$<.001$	−1.281	>100
14	ES	24	.958	3.458	3.099	3.551	−3.149	23	.004	−.643	9.533
15	ES	60	−.733	5.083	3.550	4.931	−8.137	59	$<.001$	−1.050	>100

Note. SR = Suppression ratio; ES = Elevation score; *d* = Cohen's *d*.
° For Experiment 4, the means and standard deviations are calculated over the entire first and last session because trial-level information was not available.

Appendix K

Results of Repeated Measures ANOVA With Trial as Within-Subjects Variable on Last Day of Elemental Training

Experiment	Measure	*N*	*F*-value	*df*	*p*-value	$\eta^2_{partial}$
10	SR	8	3.050	1, 7	.124	.303
11	ES	12	2.430	1, 11	.147	.181
12	ES	12	.379	1, 11	.551	.033
13	ES	24	.026	1, 23	.874	.001
14	ES	24	4.971	1, 23	.036	.178
15	ES	60	.239	1, 59	.633	.004

Note. This analysis was conducted using SPSS; it is not possible to conduct a Bayesian Repeated Measures ANOVA with a within-subjects variable that consists out of more than three levels with JASP 0.7.1 (Love et al., 2015).

Appendix L

Descriptives and Results of Pairwise Parametrical and Bayesian *t*-Tests Conducted With JASP 0.7.1 (Love et al., 2015) Comparing Conditioned Responding to First and Last Presentation of A or B During the Last Elemental Training Session

Experiment	Measure	*N*	M_{first}	M_{last}	SD_{first}	SD_{last}	*t*-value	*df*	*p*-value	*d*	BF_{10}
10	SR	8	.013	.000	.035	.000	1.000	7	.351	.354	.500
11	ES	12	4.000	4.167	2.412	2.552	−.133	11	.897	−.038	.290
12	ES	12	3.750	2.917	3.864	4.461	.621	11	.547	.179	.339
13	ES	24	4.792	3.333	3.230	2.973	1.680	23	.106	.343	.728
14	ES	24	3.500	3.458	3.799	3.551	.066	23	.948	.013	.215
15	ES	60	4.033	5.083	4.933	4.931	−1.593	59	.117	−.206	.465

Note. SR = Suppression ratio; ES = Elevation score; *d* = Cohen's *d*.

(*Appendices continue*)

Appendix M

Descriptives and Results of *t*-tests Comparing Conditioned Responding to the First Presentation of AX in the Compound Phase Between Experimental and Control Groups

Experiment	Measure	*N*	$M_{\text{experimental}}$	M_{control}	$SD_{\text{experimental}}$	SD_{control}	*t*-value	*df*	*p*-value	*d*	BF_{10}
10	SR	8	.000	.125	.000	.250	−1.000	3	.391	−.707	.695
11	ES	12	3.000	2.667	2.280	4.274	.169	10	.870	.097	.472
12	ES	12	−.167	4.500	7.441	2.588	−1.451	10	.177	−.838	.879
13	ES	24	3.583	3.250	2.065	3.108	.309	22	.760	.126	.387
14	ES	24	4.167	1.583	4.366	4.166	1.483	22	.152	.605	.817
15	ES	60	−1.900	−2.233	3.397	3.757	.360	58	.720	.093	.277

Note. SR = Suppression ratio; ES = Elevation score; *d* = Cohen's *d*.

Appendix N

Descriptives and Results of Pairwise *t*-Tests Comparing Conditioned Responding to the Last Presentation of A or B in the Elemental Phase and the First Presentation of X at Test

Group	Measure	*N*	$M_{\text{A/B}}$	M_{X}	$SD_{\text{A/B}}$	SD_{X}	*t*-value	*df*	*p*-value	*d*	BF_{10}
Experimental	SR	4	.000	.000	.000	.000		No statistics computed			
Control	SR	4	.000	.000	.000	.000		No statistics computed			
Experimental	SR	6	3.833	2.500	3.189	3.834	.491	5	.644	.200	.412
Control	ES	6	4.500	4.167	1.975	5.115	.202	5	.848	.083	.380
Experimental	ES	6	3.333	5.667	5.279	3.724	−.863	5	.428	−.352	.499
Control	ES	6	2.500	5.500	3.937	2.950	−3.503	5	.017	−1.430	4.589
Experimental	ES	12	3.583	4.667	3.579	2.229	−.780	11	.452	−.225	.372
Control	ES	12	3.083	5.083	2.353	3.988	−1.641	11	.129	−.474	.823
Experimental	ES	12	2.333	3.417	3.676	3.288	−.646	11	.531	−.187	.344
Control	ES	12	4.583	3.417	3.175	3.397	1.096	11	.297	.316	.473
Experimental	ES	30	5.867	−.333	5.818	2.845	6.060	29	<.001	1.106	>100
Control	ES	30	4.300	−.700	3.789	3.334	5.287	29	<.001	.965	>100

Note. SR = Suppression ratio; ES = Elevation score.

(*Appendices continue*)

Appendix O

Overview of the Additional Blocking Conditions (Experimental Versus Control Groups) Included in the Extended Bayesian Meta-Analysis Along With the Experiments Reported Here

Experimental condition	$M_{\text{experimental}}$	M_{control}	$SD_{\text{experimental}}$	SD_{control}	N^1_{total}	t-value
Beckers et al., 2006, Experiment 1, Irrelevant element condition	.200	.056	.083	.083	24	4.243
Beckers et al., 2006, Experiment 1, Irrelevant compound condition	.269	.036	.156	.042	24	5.003
Beckers et al., 2006, Experiment 2, Additive condition	.139	.040	.156	.048	24	2.101
Beckers et al., 2006, Experiment 2, Irrelevant element condition	.100	.089	.118	.100	24	.246
Beckers et al., 2006, Experiment 2, Irrelevant compound condition	.160	.049	.139	.111	24	2.167
Beckers et al., 2006, Experiment 3, Submaximal high condition	.145	.035	.097	.038	24	3.657
Beckers et al., 2006, Experiment 3, Submaximal low condition	.202	.091	.111	.104	24	2.531
Wheeler et al., 2008, Experiment 1, Irrelevant no shift condition[2]	.451	.100	.135	.087	24	7.577
Wheeler et al., 2008, Experiment 1, Subadditive shift condition[2]	.259	.050	.156	.083	24	4.098
Wheeler et al., 2008, Experiment 2 Long condition[2]	.163	.033	.135	.052	24	3.111
Wheeler et al., 2008, Experiment 3, Difference condition[2]	.239	.042	.142	.073	24	4.277

[1] For all conditions the number of animals in the experimental and control group was equal. [2] Data for the first two presentations of X at test. Means and standard deviations were estimated based on the reported figures.

Appendix P

Power Analyses

The power analyses reported below are based on the effect sizes obtained in similar studies (see Appendices A to E). None of those studies reported the effect size based on the local error term. In case *t*- or *F*-values for the comparison between the relevant groups (blocking and control) using the local error term were reported, Cohen's *d* was estimated using those reported values based on the formula of Thalheimer and Cook (2002). If *t*- or *F*-values for this comparison were not reported, estimations of the means and standard deviations from the reported figures were used to derive Cohen's *d*, again using the appropriate formula from Thalheimer and Cook (2002). For each series, an overall effect size, calculated as a weighted mean based on sample size, was then estimated. After calculating the overall effect size, power analyses were conducted with G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007).

Series 1

Effect Size

Jones and Gonzalez-Lima (2001). Cohen's $d = 1.83$. Calculation of effect size was based on the F-value comparing freezing in the tone-blocked group with the tone-excitor group. Note that the effect size is possibly inflated due to use of an improper control group (unpaired control).

Mackintosh, Dickinson, and Cotton (1980). Insufficient data reported to calculate an effect size.

Power Analyses

Experiment 1: $\beta = 0.992$ (for one-tailed t-test and assuming $d = 1.83$).

(*Appendices continue*)

Experiment 2: $\beta = 0.996$ (for one-tailed *t*-test and assuming $d = 1.83$).

Experiment 3: $\beta = 0.994$ (for one-tailed *t*-test and assuming $d = 1.83$).

Experiment 4: $\beta = 0.989$ (for one-tailed *t*-test and assuming $d = 1.83$).

Series 2

Effect Size

Beckers et al. (2006). Cohen's $d = 1.81$ for irrelevant elemental pretraining and Cohen's $d = 2.13$ for irrelevant compound pretraining (note that those effect sizes differ slightly from the effect sizes reported by Beckers et al. (2006) because the reported effect sizes were based on the global error term).

Wheeler et al. (2008). Cohen's $d = 3.23$ effect size based on estimations of means and standard deviations of the irrelevant-control-no-shift group and irrelevant-blocking-no-shift group in the first block.

Weighted Effect Size

Cohen's $d_{\text{series 2}} = 2.93$, effect size weighted based on sample size (see Table B) for the three effect sizes reported above.

Power Analysis

Experiment 5–6: $\beta > 0.99$ (for one-tailed *t*-test and assuming $d = 2.39$).

Series 3

Effect Size

Blaisdell et al. (1999). Cohen's $d = 1.79$, based on estimations of means and standard deviations for the two-phase blocking and control group.

Note that the procedures of Beckers et al. (2006) and Wheeler et al. (2008) were also very similar to the ones used in Series 3. Including the effect sizes of those studies in the power analysis would result in a higher power.

Power Analysis

Experiment 7–10: $\beta = 0.72$ (for one-tailed *t*-test and assuming $d = 1.79$).

Series 4

Effect Size

Holland (1999, Experiment 6). Cohen's $d = 2.30$, based on estimations of means and standard deviations of the no-extinction blocking and no-extinction overshadowing group (= B control) and assuming that error bars represent standard error of the mean (in case error bars would represent standard deviations the effect size and thus the estimated power would be larger).

Taylor et al. (2008). Cohen's $d = 0.81$, based on result of *t*-test for comparing blocking and control group.

Weighted Effect Size

Cohen's $d_{\text{series 4}} = 1.35$, effect size weighted based on sample size (see Table D) for the two effect sizes reported above.

Power Analyses

Experiment 11–12: $\beta = 0.70$ (for one-tailed *t*-test and assuming $d = 1.35$).

Experiment 13–14: $\beta = 0.94$ (for one-tailed *t*-test and assuming $d = 1.35$).

Series 5

Effect Size

Taylor et al. (2008). Cohen's $d = 0.81$, based on result of *t*-test for comparing blocking and control group.

Power Analysis

Experiment 15: $\beta = 0.93$ (for one-tailed *t*-test and assuming $d = 0.81$).

Received December 16, 2015
Revision received May 19, 2016
Accepted May 23, 2016 ■