Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

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

# Cue competition in mental state inference: Blocking effects in learning to interpret the behaviors of others

André Mata *, André Vaz, Tomás A. Palma

*CICPSI, Faculdade de Psicologia, Universidade de Lisboa, Portugal*

ARTICLE INFO

Editor: Lasana Harris

*Keywords:*
Mental state inference
Blocking
Cue competition
Associative learning

ABSTRACT

In the literature on learning, one of the most robust cue competition effects is blocking: the previous learning of a cue-outcome association prevents learning that other cues predict the same outcome if those cues are presented together with the first cue. In this research, we investigated blocking effects in mental state inference. Participants learned to diagnose the internal states of a target person based on the behaviors he displayed. Blocking effects were observed across several studies, such that, when participants had previously learned that a certain behavior predicted a certain internal state, they later failed to learn about the predictive value of other behaviors that were paired with the original behavioral cue. Implications are discussed for cognitive models of learning and cue competition, as well as for the social psychology of mental state inference.

## 1. Introduction

As people get to know other people, they learn how to interpret subtle cues in their behavior to infer what they are thinking or feeling. Imagine that Jack gets to know Jill at work, and with repeated interaction he finds out that, whenever Jill is nervous or anxious about something (e.g., missing a deadline), she fidgets with her watch. As they get better acquainted with each other, they start to interact in a more leisurely context (e.g., they meet for drinks after work), where Jill's behavior is less constrained and she is able to manifest other cues. In that context, whenever Jill is nervous or anxious about something, she fidgets with her watch *and* bites her nails. The question is: given that the first cue is present, will Jack pick up on the second cue, or will he fail to learn its diagnostic value? The second cue is equally predictive of the outcome, but it may become blocked by the previous learning of the association between the first cue (fidgeting with watch) and the internal state (anxious), such that, if the second cue is presented without the first one, he may fail to make the correct inference. Thus, his initial learning history may interfere with the successful learning of other equally valid cues.

This paper explores blocking effects in mental state inference. In particular, it focuses on blocking in learning to infer the internal states of another person (e.g., their mood, their motivations) from observing their behaviors, as in the example above. There is a long tradition of psychological research on how people make lay inferences about the minds of others (e.g., Ambady & Rosenthal, 1992; Ames, 2004; Baron-Cohen, Leslie, & Frith, 1985; Funder, 1995; Jones & Davis, 1965; Winter & Uleman, 1984; Wu & Mitchell, 2019). Much of that research has focused on individual differences related to the perceiver or the target of the inferences (for a review, see Funder, 2012). The present research explores a key situational factor: the order in which behavior-mental state associations are learned.

### 1.1. Blocking effects in social psychology

Cue competition effects are well-established in the learning literature. Among them, one of the most robust effects in both human and non-human animal learning is blocking. Blocking occurs when the previous learning of an association between a certain cue and an outcome interferes with the subsequent learning of an association between another cue and the same outcome. Specifically, in a typical blocking paradigm (e.g., Aitken, Larkin, & Dickinson, 2001; Dickinson, Hall, & Mackintosh, 1976; Dickinson, Shanks, & Evenden, 1984), subjects first learn to associate a certain cue (A) with a specific outcome (X). In a second phase, the cue is then presented alongside a second cue (B), both predicting the same outcome as before (X), rendering the second cue redundant. In a final test phase, when the second cue is presented in the absence of the first one, it is thought to hold less predictive power.

Blocking has been widely researched in both non-human and human learning, but much fewer studies have explored blocking effects in social

* Corresponding author at: Faculdade de Psicologia, Alameda da Universidade, 1649-013 Lisboa, Portugal.
*E-mail address:* andremata@psicologia.ulisboa.pt (A. Mata).

https://doi.org/10.1016/j.jesp.2021.104122
Received 26 July 2020; Received in revised form 22 February 2021; Accepted 27 February 2021

psychology. Here are a few notable exceptions:

Cramer, Weiss, Steigleder, and Balling (1985) explored blocking in interpersonal attraction. In their studies, the attraction that participants felt toward a person (because that person showed constant agreement with the participants' opinions) was blocked when that person was paired with another person who participants already found attractive.

Sanbonmatsu, Akimoto, and Gibson (1994) used an associative account of blocking to explain how stereotypes influence judgments of the causality of stereotype-irrelevant factors. When participants had to explain why other students failed or succeeded in a course, the previous activation of the covariation between the outcome and stereotype-relevant factors reduced participants' search for additional predictive cues, thus blocking the perception of the covariations of other, stereotype-irrelevant, factors.

Collins, Percy, Smith, and Kruschke (2011) examined blocking in the context of how people integrate information from social (advice from others) and nonsocial sources (direct experience) to make decisions, such as choosing stocks, diagnosing diseases, and predicting political affiliations. They found that when participants had to make decisions based on social sources alone, they followed the information acquired in an early training phase and ignored as redundant the information obtained in a late training phase (i.e., blocking). However, when participants received information from both social and nonsocial sources (e.g., advice about the political affiliation of targets in the initial training phase and targets' stances on political issues in the late training phase, or vice versa), blocking was not observed.

Finally, in a more recent study, FeldmanHall, Dunsmoor, Kroes, Lackovic, and Phelps (2017) investigated whether blocking also plays a role when humans learn about others' social values in a group setting. In their study, participants initially played a series of social-economic games where they could learn whether game-dictators were characteristically altruistic or selfish based on their monetary splits. In a subsequent learning episode, participants were again exposed to the same dictators, but now the dictators' altruistic or selfish monetary splits were made collectively with a partner whom participants had not encountered before. Then, participants played a trust game in which they had to decide how much of their own money to entrust to the dictators as well as to their partners. They found that participants entrusted more money to altruistic than selfish dictators who were encountered alone in the first learning episode. However, there was no difference in how much money they entrusted to the dictators' partners paired with either altruistic or selfish outcomes in the second learning episode. That is, the presence of previously known altruistic and selfish dictators blocked the partners from acquiring positive or negative social value, as demonstrated by participants' decisions to treat them as if they had not been associated with altruistic or selfish outcomes.

While all of these studies explore blocking in a broad sense, only some of them follow the standard blocking paradigm, whereby there is a learning phase where an outcome is associated with a cue, followed by a second moment where participants fail to associate a new cue with the same outcome. This is the case with the studies by Cramer et al. (1985) and FeldmanHall et al. (2017). However, the other two studies do not conform strictly to the standard blocking paradigm. In Sanbonmatsu et al.'s (1994) experiments, there was no explicit learning phase where each cue was associated with a certain outcome. Rather, all types of cues were presented simultaneously, followed by the outcome. The blocking came not from previous learning and introduction of new cues, but instead by making some cues more salient than others, which retroactively hid the covariations between the outcome and the blocked cues. And although Collins et al. (2011) did follow a more conventional blocking paradigm, the argument could be made that social and nonsocial cues did not compete for association with the same type of outcomes. While nonsocial cues were associated with specific outcomes – i.e., a symptom predicts a particular disease –, social cues, on the other hand, were associated with veracity, independent of outcome – i.e., an advisor was learned to be trustworthy or not, regardless of which advice they gave.

More importantly, previous studies have largely ignored blocking effects in mental state inference. We believe that cue competition phenomena have implications for how people learn about others and how they make sense of their behaviors. Social psychology knows a great deal about the effects of order (e.g., Anderson, 1965; Asch, 1946; Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993) and redundancy (e.g., Alves & Mata, 2019; Unkelbach, Fiedler, & Freytag, 2007; Weaver, Garcia, Schwarz, & Miller, 2007), but cue competition effects have not (to the best of our knowledge) been considered in the study of mental state inference (with the exception of FeldmanHall et al.'s, 2017 work on how people infer social values in the context of economic games). The present studies explore cue competition effects (specifically, blocking) in a fundamental task of lay social psychology: how people interpret external cues of others to infer their internal states – in particular, how they use other people's behaviors to infer their thoughts and feelings. Accurate mentalizing is at the core of competent social inference, but cue competition may interfere with it.

### 1.2. Overview

In all three experiments, participants learn to infer certain mental states of a target person (e.g., *optimistic*) based on several behavioral cues (e.g., *leans back in chair*). In the first phase, behaviors are presented individually, participants guess the correspondent mental state, and are provided with feedback about whether their inference is correct. In the second phase, behaviors (some of them which had already been presented before, and other new ones) are presented in pairs. Participants once again infer the corresponding mental state and receive feedback. In the test phase, the same behaviors from the second phase are presented, arranged in pairs that provide multiple tests of the blocking effect, and participants are once again asked to infer mental states. Each experiment uses a different procedure (see Table 1 below), in order to establish the robustness of the effect.

### 1.3. Sample and open practices

Sensitivity power analyses (G*Power 3.1.9.6; Faul, Erdfelder, Lang, & Buchner, 2007), assuming a power of .80 and an alpha level of .05 (two-tailed), revealed that the sample sizes used in the present experiments were sensitive to detect the following effect sizes for the main statistical effect of interest (i.e., blocking): Experiment 1 was able to detect a minimum effect size of $\eta_p^2 = .05$; Experiment 2 was able to detect a minimum effect size of $\eta_p^2 = .02$; Experiment 3 was able to detect a minimum effect size of $\eta_p^2 = .09$.

The data for each experiment were collected in one shot without prior statistical analysis. We report all data exclusions (if any), all

**Table 1**
Procedure of Experiments 1–3.

| | Phase 1 | | Phase 2 | | Test Phase | |
|---|---|---|---|---|---|---|
| Experiment 1 | No. trials | | No. trials | | No. trials | |
| | 12 | A → X | 12 | AB → X | 6 | AB → ? |
| | 12 | K → Y | 12 | CD → Y | 6 | AC → ? |
| | 6 | E → X | 6 | EF → X | 6 | AD → ? |
| | 6 | E → Y | 6 | EF → Y | 6 | BC → ? |
| | | | | | 6 | BD → ? |
| | | | | | 6 | CD → ? |
| Experiment 2 | 12 | A → X | 12 | AB → X | 4 | All cues combined |
| | 12 | C → Y | 12 | CD → Y | | |
| | 12 | E → Z | 12 | EF → Z | | |
| Experiment 3 | 10 | A → X | 10 | AB → X | 4 | AE → ? |
| | 10 | C → Y | 10 | CD → Y | 4 | AF → ? |
| | 10 | I → U | 10 | EF → Z | 4 | AG → ? |
| | 10 | J → V | 10 | GH → W | 4 | AH → ? |
| | | | | | 4 | BE → ? |
| | | | | | 4 | … |

manipulations, and all measures. Data and analysis scripts are publicly available at the Open Science Framework website (https://osf.io/ysbgp/?view_only=d561dc9d71054c9584542719b470fa8d). The Ethics Committee of the authors' institution approved these experiments.

## 2. Experiment 1

In the present experiments, participants first learn the associations between specific behavioral cues (e.g., scratching nose, fidgeting with ring) and internal states (e.g., lying, feeling anxious), before introducing additional behavioral cues. In the end, participants will be presented pairs of cues predictive of different states (i.e., competing cues), and asked to guess which state the target is in. If initially learned associations block the learning of subsequent cues, then these blocked cues should have diminished predictive power compared to competing non-blocked cues, and participants will more likely guess the states predicted by the non-blocked cues.

### 2.1. Method

#### 2.1.1. Participants

Seventy-two undergraduates participated in this study ($M_{age}$ = 22.81, $SD = 4.72$; 44 female), and were compensated with gift vouchers.

#### 2.1.2. Materials

The materials for Experiment 1 were pre-tested. Thirty-three undergraduates (additional to the 72 participating in the main study) completed this pretest. Of these, one did not complete the survey, and one gave the same answer for all stimuli; these participants were excluded, and a final sample of 31 participants was considered.

One hundred combinations of 10 behavioral cues (e.g., talking fast; coughing) and 10 emotional/attitudinal motivational states (e.g., lying; being optimistic), previously generated by the authors, were used. These behavioral cues and states can be seen in the appendix.

Participants rated the likelihood of a person being in a certain state, given that he displayed a certain behavior, on a 9-point scale, from 1 – not likely to 9 – very likely. All combinations of behavior and state were rated by each participant. The order of the combinations was randomized into two different orders, and participants were randomly attributed to either order.

From the 100 pretested combinations of behaviors and states, we selected pairs with a low average likelihood rating (seeking to use material low in a priori association), and where the average likelihood ratings within each pair of cues did not differ significantly from each other. In total, 9 different groups of cues and states were created. These groups were similar in structure, differing only in the specific behavioral cues and internal states. For all groups, previous association levels were low and balanced so that no cue predicted better either one of the two states, within that group.

#### 2.1.3. Procedure

Each participant performed the task with only one group of behaviors and states, and the procedure was the same for all groups: there were two learning phases, followed by a testing phase (see Table 1).

Throughout all studies, we shall refer to cues expected to have caused blocking as *early cues* (signaling that they first appeared in phase 1), and to cues expected to have been blocked as *late cues* (signaling that they first appeared in phase 2, paired with an early cue). Cues that were neither expected to have been blocked nor to have blocked some other cue (i.e., cues that were presented either always alone or whose pair was not an early cue) will be referred to as *control cues*.

In the first phase, participants were told that Peter performed a certain behavior (A, K, or E), and asked what was Peter more likely feeling (X or Y). After each answer, the correct response was displayed. Each cue appeared 12 times (trials). For cue A, the correct answer was X; for cue K, the correct answer was Y; and for cue E, X was the correct answer for half the trials, and Y was the correct answer for the other half.

In the second phase, early cue A was always paired with late cue B, cue E paired with cue F, and cue K was replaced with the pair of control cues CD. Once again, there were 12 trials of AB, with X as the correct answer, 12 trials of CD, with Y as the correct answer, and 12 trials of EF, half with X as the correct answer, and half with Y.

Finally, in the testing phase, only cues A, B, C and D were presented. All cues were combined with each other, with 6 trials per pair: AB, AC, AD, BC, BD, and CD. In this phase, no feedback was provided concerning the correct response; each answer was immediately followed by a new trial. Within all phases, the order of the trials was randomized, and, in all paired trials, the order of both cues was counterbalanced.

### 2.2. Results

For each pair of cues, in the test phase, we computed the proportion of trials in which the participant selected X (out of the 6 total per pair of cues). Because the only options were X and Y, the proportion of Y is equal to 1 - X, and is therefore redundant. In the following experiments, with more than 2 choices, this proportion was computed for each choice, in each pair of cues. To test for a blocking effect (and generally compare the associative strength of any two variables), two analyses were considered: a direct and an indirect one.

#### 2.2.1. Direct analysis

This analysis matches early cue A and late cue B with control cues C and D, and tests whether these have a different pull, that is, whether the proportion of X is larger when C and D are matched with A, compared to when they are matched with B. Indeed, a repeated-measures ANOVA showed a significant main effect of blocking, $F(1, 71) = 11.31$, $p = .001$, $\eta^2_p = .14$, suggesting that B had a weaker pull toward X than did A (Fig. 1 displays mean proportions in the direct analysis across Studies 1–3). The main effect of pairs of cues (BC-AC vs. BD-AD) and the interaction effect were not significant, $Fs < 1$.

#### 2.2.2. Indirect analysis

It is possible that the differences observed are not due to the late cue having a weaker pull than the control cues (i.e., to it having been blocked), but rather to differences between the early cue and the control ones. If B was blocked, then the proportion of X in both BC and BD should be significantly below .50. If, however, the differences are due to A having more weight than the control cues, then the proportion of X in AC and AD should be significantly higher than .50. One-sample *t*-tests confirm the latter: for AC ($M = .64$) and AD ($M = .61$), the proportion of X was significantly greater than .50: $t(71) = 3.21$ $p = .002$, $d = 0.38$; and $t(71) = 2.66$ $p = .010$, $d = 0.31$, respectively. For BC ($M = .47$) and BD ($M = .44$), however, the proportion of X did not significantly differ from .50: $t(71) = -0.75$ $p = .472$, $d = -0.09$; and $t(71) = -1.43$ $p = .185$, $d = -0.16$, respectively.

### 2.3. Discussion

When compared against the early cue (direct analysis), the late cue had a significantly weaker association to its respective outcome. When compared against the control cues (indirect analysis), however, results suggest that the late cue had the same associative strength as the non-blocked control cues.

## 3. Experiment 2

The results of Experiment 1 point to potential shortcomings in the procedure: while the late cue was indeed weaker in predicting the outcome than the early cue, it did not significantly differ from the control cues in its predictive power. However, it is still possible that cue B was blocked (as evidenced by it having less weight than A); the lack of

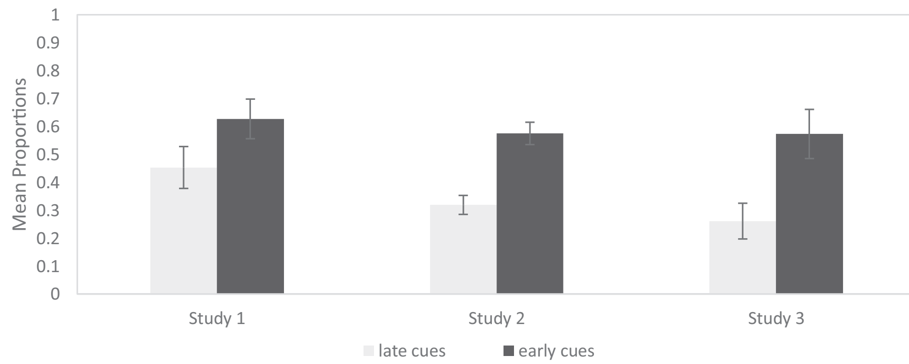


**Fig. 1.** Mean proportions and 95% CI of late vs. respective early cues' predicted outcome, when paired with a competing cue predictive of a different outcome (Direct Analysis), across studies.

differences between cue B (the late cue) and cues C and D (the control cues) may be because the learning of the association between cues C and D and outcome Y was impaired, as outcome Y was predicted by cue K in phase 1. Since outcome Y was already associated to cue K, this may have prevented a new association to cues C and D. To control for this possibility, Experiment 2 replaced cue K, in phase 1, with cue C (see Table 1).

Additionally, rather than having a bad predictor cue in phase 1 (cue E), Experiment 2 included a third possible state that was reliably predicted by cue E. Having 3 options makes the task more complex and provides one additional pair of cues for the analysis.

Moreover, the testing phase tested all possible combinations of the 6 cues (A, B, C, D, E and F). The relevant pairs were the ones that combined an early cue (A, C or E) with a late cue (B, D or F) that predicted a different state than the one predicted by the early one: AD, AF, CB, CF, EB and ED.

### 3.1. Method

#### 3.1.1. Participants

Fifty undergraduates participated in this study ($M_{age}$ = 21.16, $SD$ = 6.37; 37 female), and were compensated with gift vouchers.

#### 3.1.2. Materials

In this experiment, rather than using pre-selected pairs of cue-state, all 10 cues and 10 emotional states were randomly combined, enabling a richer set of stimuli.

#### 3.1.3. Procedure

There were two learning phases, followed by a testing phase (see Table 1). For each participant, a randomly selected set of cues/emotional states was used.

Once again, in the first phase participants were presented with a single behavior at a time (A, C or E—*early cues*). Unlike in Experiment 1, in Experiment 2 there were 3 possible emotional states: X, Y and Z. There were 12 trials for each cue, and the correct answer was provided after each trial. For cue A, the correct answer was X; for cue C, the correct answer was Y, and for cue E, the correct answer was Z.

In the second phase, each of the early cues (A, C, and E) was paired with a late cue (B D, and F, respectively), keeping the same emotional states as the correct answers: X for AB, Y for CD, and Z for EF. Once again, each pair of cues appeared in 12 trials, with the correct answer always given after each trial.

Finally, in the testing phase, all possible pairs of the 6 cues (A, B, C, D, E and F) were presented, 4 trials each, totaling 60 trials (4 trials × 15 combinations). In this phase, no corrective feedback was given.

### 3.2. Results

Once again, the proportions of each answer, per pair of cues in the test phase, were computed.

#### 3.2.1. Direct analysis

In this analysis, both cues in a pair (e.g., A and B) were compared against each other through a third *comparator* cue (e.g., C), by testing for differences in the proportion of the main cues' outcome. For example, if B was blocked by A, then AC should have a significantly higher proportion of X, compared to BC. A repeated-measures ANOVA tested for these differences for all three pairs of cues (AB, CD, and EF), matching them with all other possible cues (e.g., AB were tested using C, D, E, and F as comparators). Additionally, the strength of the comparator was also accounted for. Having an early cue as the comparator (e.g., C as a comparator for AB: AC vs. BC) should decrease the proportion of the main cue's outcome (e.g., X), compared to having a late cue as comparator (e.g., D as a comparator for AB: AD vs. BD), because the former should pull more strongly away from the main cue's outcome. Thus, if there are indeed differences between the early cues and the late cues, then not only will there be a main effect of blocking – i.e., the early cue pulls more toward its outcome than the late cue –, but also a main effect of the strength of the comparator cue – i.e., the early comparator cue pulls more away from the other outcome than the late comparator cue.

As expected, there was both a main effect of blocking, $F(1, 49) = 55.86$, $p < .001$, $\eta^2_p = .53$, and a main effect of comparator strength, $F(1, 49) = 56.75$, $p < .001$, $\eta^2_p = .54$. Together, these results indicate that there were differences between early and late cues, such that early cues both pulled more in the direction of their predicted outcome, and away from the other outcomes, compared to late cues (see Fig. 1).

Additionally, and as in Experiment 1, there was no main effect of pairs, $F(5, 245) = 1.19$, $p = .314$, $\eta^2_p = .02$, suggesting that the pattern was consistent across all pairs of cues, and only a significant interaction effect of blocking by comparator strength emerged, $F(1, 49) = 4.25$, $p = .045$, $\eta^2_p = .08$, though the effects reported above hold for all pairs.

#### 3.2.2. Indirect analysis

This analysis compares whether, within the same pair of competing cues, the proportion of one outcome is greater than the other. To give an example, AD should have a significantly higher proportion of X (predicted by A) than Y (predicted by D), if D was blocked. There were 6 relevant pairs of cues (AD, AF, CB, CF, EB, and EF) that combined an early cue with a late cue with different outcomes. A repeated-measures ANOVA revealed a main effect of blocking, $F(1, 49) = 58.21$, $p < .001$, $\eta^2_p = .54$, such that, for all 6 pairs, there was a greater proportion of the early cue predicted-outcome, when compared to the proportion of the late cue predicted-outcome. The main effect of pairs of cues and the interaction effect were not significant, $Fs < 1$.

### 3.3. Discussion

As with Study 1, comparing the late cues with their respective early cues (i.e., that were in the same phase-2 pair; direct analysis) shows a significant difference in the proportion of times participants chose their outcome. Specifically, the corresponding outcomes predicted by the cues were chosen less often for the late cues than for the early cues, suggesting decreased associative strength for the late cues.

When late cues were compared against different early cues (i.e., predictive of a different outcome, in the same test phase pair; indirect analysis), once again the late cues predicted a significantly lower proportion of their respective outcome, compared with the early cues.

## 4. Experiment 3

Because the differences in associative strength in Experiment 2 may have resulted from the early cue (e.g., A) having had more trials than the late cue (e.g., D), and not because the late cue was actually blocked, the testing phase in Experiment 3 combined the late cues with non-blocked (control) cues that were first introduced in phase 2 and, thus, had the same number of trials.

On the other hand, as in Experiment 1, the previous association with a different cue could affect the association of the cues that first appeared in phase 2 with their respective outcomes. To prevent that, while also keeping the same number of states in both phases, phase 1 included 2 cues (I and J) and 2 states (U and V) that were replaced with pairs of cues EF and GH, and states Z and W, in phase 2. This way, the control cues introduced in phase 2 (EF and GH) had the same number of trials as the late cues (B and D), without the possibility of their learning having suffered interference from previous learning.

The testing phase combined cues A, B, C and D with cues E, F, G and H. Combining B and D with E, F, G and H, enables a comparison between late and control cues with the same number of trials. Because there were more cues/pairs of cues, the number of trials for each cue, in phases 1 and 2, was reduced to 10.

### 4.1. Method

#### 4.1.1. Participants

Thirty undergraduates participated in this study ($M_{age} = 26.30$, $SD = 10.85$; 18 females, 1 did not disclose), and were compensated with gift vouchers.

#### 4.1.2. Materials

Same as in Experiment 2.

#### 4.1.3. Procedure

There were two learning phases, followed by a testing phase (see Table 1). For each participant, a randomly selected set of cues and states was used.

In phase 1, participants were presented with a single behavior at a time, followed by the correct answer. There were 4 possible behavioral cues: A, C, I and J; and 4 possible state outcomes: X, Y, U and V. Table 1 shows the correct state for each cue. There were 10 trials per cue.

In phase 2, early cues A and C were paired with late cues B and D, respectively, cue I and state U were replaced with control cues EF and state Z, and cue J and state V were replaced with control cues GH and state W. Once again, there were 10 trials for each combination.

In the test phase, each of the early and late cues A, B, C and D was combined with control cues E, F, G and H. There were 4 trials for each pair, totaling 64 trials (4 trials × 16 combinations). No corrective feedback was given after each trial.

### 4.2. Results

As in the previous studies, the proportions of each answer, per pair of cues in the test phase, were computed.

#### 4.2.1. Direct analysis

Unlike in Experiment 2, in the present experiments we did not match the late cues with other early/late cues in the test phase. Rather, the late cues were matched with control cues (i.e., the comparator cues were control cues). As such, no effect of comparator strength was tested. Replicating previous results, only a main effect of blocking was significant, $F(1, 29) = 26.39$, $p < .001$, $\eta^2_p = .48$. No significant main effect of pairs or interaction effect was observed, $Fs < 1$.

#### 4.2.2. Indirect analysis

A repeated-measures ANOVA revealed a main effect of blocking, $F(1, 29) = 11.66$, $p = .002$, $\eta^2_p = 0.29$, such that, for every pair of late (B or D) and control cues (E, F, G, and H), the proportions of the outcomes associated with the control cues (outcomes Z or W) were significantly greater than the proportions of the outcomes associated with the late cues (outcomes X and Y). For example, the proportion of Z was higher than the proportion of X in BE. There was neither a significant main effect of pairs of cues, $F(7, 203) = 1.07$, $p = .387$ $\eta^2_p = .04$, nor a significant interaction effect, $F(7, 203) = 1.02$, $p = .421$, $\eta^2_p = .03$.

### 4.3. Discussion

Similarly to the previous studies, a comparison between the late and the early cues (direct analysis) shows that early-control cue pairs predicted a significantly higher proportion of the target cue's outcome than did the corresponding late-control cue pairs. That is, late cues had a weaker association with their outcome.

Unlike Study 2 (which had no control cues) and Study 1 (which showed no late/control cue differences), comparing late cues against control cues (indirect analysis) also revealed a blocking effect, whereby the outcome predicted by the control cue was chosen more often than the outcome predicted by the late cue, for all the late-control cue pairs. In other words, late cues had a weaker association with their outcome than did control cues that had been presented the same number of trials.

## 5. General discussion

This research shows that the same cue competition phenomena that have been observed in other forms of learning also hold for mental state inference, when people try to make sense of other people's behaviors and learn how those behaviors predict their internal states. Across three experiments with various procedural versions of the blocking paradigm, learning that a certain behavior predicted a certain internal state (a certain mood or motivation) blocked the subsequent learning that other behaviors also predicted the same internal state. These results add to a large body of evidence showing that cue competition is a robust and pervasive phenomenon in human cognition (FeldmanHall & Dunsmoor, 2019).

Thus, the learning history (i.e., the order in which people learn to associate certain behavioral cues with the corresponding internal states) is of crucial importance. This is particularly relevant when people first learn about the meaning of the behaviors of a person in a certain context, and then see that person behave in another context. To the extent that different situations bring about different behavioral expressions (e.g., behavior might differ in a formal workplace setting vs. an informal leisure setting), the opportunities to learn about the cue value of certain behaviors might differ across contexts. However, usually, the process of getting acquainted with someone is context-bound. That is, people first tend to interact with a new acquaintance in a certain context (e.g., their workplace, as in the introductory example), and so they might be less sensitive to other diagnostic cues in a different context. This is yet another way (along with the widely studied correspondence bias; Gilbert & Malone, 1995) in which blindness to contextual influences can interfere with sound mental state inference.

### 5.1. Implications for mental state inference

This research has implications for the psychology of mental state inference. One of the conditions that have been suggested as essential for accurate social inference is that the behavioral cue must be available to, and detected by, the observer (Funder, 1995). Cue competition effects might render some cues invisible. However, this is not because of the more obvious reasons that were identified in previous research: "because the judge may be inattentive, unperceptive, or distracted, or the behavior itself (e.g., a momentary facial expression), although available in principle, might be extremely difficult to see" (Funder, 1995, p. 658), but rather because learning order and redundancy effects render it unnoticeable. When authors (e.g., Funder, 1995) have discussed the issue of the visibility of traits, they argue that some traits are more judgeable than others as they are more externally visible (e.g., sociability) than others (e.g., daydreaming). However, the visibility of internal states in the present research was determined not by whether some cues were manifested more than others but rather because the learning history of some cues blocked learning of subsequent cues that would otherwise have been as visible as the former.

Those previous models of mental state inference have discussed not only what makes for a good trait but also what makes for a good target. With regard to this, those models predict that "people with a higher level of behavioral activity, especially social activity, should be easier to judge than those who are less active, because more active people naturally have more behaviors that then become available as potential clues about their personality" (Funder, 1995, p. 661). However, cue competition presents a case where accurate mental state inference is hindered (not helped) by the availability of multiple cues. Thus, the study of cue competition adds complexity and nuance to the psychology of mentalizing, as it suggests that the ability to accurately infer the internal states of others might depend on factors other than the visibility, quantity, and consistency of behavior.

Finally, yet another aspect that makes the present research different from the vast majority of research on mental state inference is that, in previous studies (e.g., on perspective-taking, theory of mind, and trait inference), there are accurate responses based on the universal meaning of certain cues (e.g., the meaning of facial expressions in the Reading-the-Mind-in-the-Eyes Test; Baron-Cohen, Jolliffe, Mortimore, & Robertson, 1997), whereas in the present research the meaning of behaviors was idiosyncratic and had to be learned for each target. That is, whereas previous studies used clues that were universally diagnostic (i.e., behaviors or expressions that in most people signal the same trait or disposition), our work explores the ability to learn idiosyncratic associations about a specific target. Thus, the present research focuses on how people really get to know a specific other person and their unique behavioral signatures, and not just whether some people are better at reading others, or whether some people are easier to read than others.

### 5.2. Implications for learning models

While this research seeks mainly to contribute to the understanding of social cognition, it also has potential implications for cognitive models of learning and cue competition.

The main theoretical accounts of blocking can be broadly divided into two classes: associative accounts and propositional (or inferential) accounts. Associative accounts share the assumption that learning is accomplished by the trial-by-trial formation of unqualified associations in memory (e.g., Dickinson, 1980; Mackintosh, 1983; Pearce & Hall, 1980; Rescorla & Wagner, 1972). For example, the influential Rescorla-Wagner model (Rescorla & Wagner, 1972) posits that the formation of associations between a cue (A) and an outcome (X) depends on the extent to which the outcome is unexpected on a certain trial. Thus, in a blocking procedure, the A->X association is formed in phase 1 because the outcome is initially unexpected. In phase 2, when cue A and a new cue B are presented with the same outcome (AB->X), the outcome is no longer unexpected given the presence of cue A and, therefore, the association between cue B and the outcome is not formed.

Alternatively, propositional accounts argue that blocking (and cue competition effects in general) arises from the interplay between propositional knowledge and higher-order reasoning, and not from the formation of simple associations (De Houwer, 2009; De Houwer, Beckers, & Vandorpe, 2005; Lovibond, 2003; Lovibond, Been, Mitchell, Bouton, & Frohardt, 2003; Waldmann & Holyoak, 1992). For example, participants might infer that if cue A alone causes outcome X to occur with a certain intensity, and if Cue A and B together cause the outcome to occur with the same intensity (AB->X), then B is not likely causing X, otherwise X should have occurred with higher intensity (for a review, see Shanks, 2010).

Causal-model theory (Waldmann, 2000; Waldmann & Holyoak, 1992) diverges from standard associative models in stating that the causal status of the cues (cause or effect) is a key moderating factor in cue competition. In particular, causal-model theory predicts that blocking occurs in predictive learning (i.e., when learning to predict an effect from multiple possible causes), but not in diagnostic learning (i.e., when learning to diagnose a cause based on multiple possible effects). Presumably, participants in our studies interpreted the behavioral cues as effects and the to-be-inferred internal state as the cause, in which case causal-model theory would not predict cue competition. However, we observed large cue competition effects in our diagnostic task, and thus these results seem to favor associative theories over causal-model theory and propositional theories in general.

### 5.3. Future research

Future studies can explore generalization effects and potential moderators of those effects. For instance, what someone learns about person X might block the learning of other cues in person Y, and the degree to which that generalization occurs might depend on the similarity and group belonging of those people (FeldmanHall et al., 2018). Thus, there might be effects of what is learned about one person that carry over to the way in which the behavior of another person is interpreted.

Moreover, it would be interesting to test cue competition phenomena in a live interaction setting. A recent review in this area called attention to "the importance of asking perceivers to interpret spontaneous target behavior" (Wu, Sheppard, & Mitchell, 2019, p. 3). These authors denounced that the vast majority of tasks used in research on mental state inference "usually require the participant to attribute a mental state to a hypothetical target (social agent) or a posed actor based on a fictitious situation, an abstract social animation, or based on a static image in the absence of situational context (e.g., photographs of the eyes). Though these process-based tasks are helpful to probe the cognitive underpinnings of mindreading, they say rather less about how accurately participants routinely make everyday interpretations of other minds when observing subtle and spontaneous behavior" (p. 3). Fortunately, new methods are enabling the study of mental state inference in spontaneous situations outside the lab (e.g., Mehl, Gosling, & Pennebaker, 2006).

Another direction still is to study whether cue competition effects can extend to the whole chain of retrodictive inference (Wu et al., 2019). Specifically, it is possible to extend the study of blocking from inferences about proximal causes (a certain behavior is caused by a certain internal state) to inferences about distal causes (a certain behavior is caused by a certain internal state, which in turn was caused by an external event in the world; Pillai, Sheppard, & Mitchell, 2012). Thus, social perceivers might not just stop at inferring internal states, as such distal inferences (i.e., inferring what event triggered the internal state that eventually originated the behavior) can be equally valuable in terms of making sense of a person and learning to predict her behavior.

These are but a few examples of the great potential to learn novel insights about mental state inference by considering cue competition phenomena and associative theories of learning in general. This is in accordance with recent research showing how associative learning models can grant a novel understanding of various social psychological phenomena and how complex behavior and cognition can arise from the same elementary principles of learning that are shared by a wide range of species (FeldmanHall & Dunsmoor, 2019).

## Acknowledgments

This research was supported by Portuguese national funds through FCT—Fundação para a Ciência e a Tecnologia, I.P. (DL57/2016/CP [1439]/CT[3], PTDC/PSI-GER/7592/2020).

## Appendix A. Appendix

| Materials | |
|---|---|
| **Behavioral cues** | **Internal states** |
| Blinks | Tired |
| Scratches nose | Anxious |
| Plays with hair | Optimistic |
| Leans back in chair | Annoyed |
| Coughs | Bored |
| Taps foot | Sad |
| Fidgets with ring | Distracted |
| Talks fast | Restless |
| Cleans glasses | Lying |
| Crosses arms | Feeling attracted |

## References


Aitken, M. R. F., Larkin, J. W., & Dickinson, A. (2001). Re-examination of the role of within-compound associations in the retrospective revaluation of causal judgements. *The Quarterly Journal of Experimental Psychology Section B: Comparative and Physiological Psychology, 54*(1), 27–51. https://doi.org/10.1080/713932745.

Alves, H., & Mata, A. (2019). The redundancy in cumulative information and how it biases impressions. *Journal of Personality and Social Psychology, 117*(6), 1035–1060. https://doi.org/10.1037/pspa0000169.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin, 111*, 256–274. https://doi.org/10.1037/0033-2909.111.2.256.

Ames, D. R. (2004). Strategies for social inference: A similarity contingency model of projection and stereotyping in attribute prevalence estimates. *Journal of Personality and Social Psychology, 87*(5), 573–585. https://doi.org/10.1037/0022-3514.87.5.573.

Anderson, N. H. (1965). Primacy effects in personality impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology, 2*(1), 1–9. https://doi.org/10.1037/h0021966.

Asch, S. E. (1946). Forming impressions of personality. *Journal of Abnormal and Social Psychology, 41*, 258–290. https://doi.org/10.1037/h0055756.

Baron-Cohen, S., Jolliffe, T., Mortimore, C., & Robertson, M. (1997). Another advanced test of theory of mind: Evidence from very high functioning adults with autism or Asperger syndrome. *Journal of Child Psychology and Psychiatry, 38*(7), 813–822. https://doi.org/10.1111/j.1469-7610.1997.tb01599.x.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind". *Cognition, 21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8.

Collins, E. C., Percy, E. J., Smith, E. R., & Kruschke, J. K. (2011). Integrating advice and experience: Learning and decision making with social and nonsocial cues. *Journal of Personality and Social Psychology, 100*(6), 967–982. https://doi.org/10.1037/a0022982.

Cramer, R. E., Weiss, R. F., Steigleder, M. K., & Balling, S. S. (1985). Attraction in context: Acquisition and blocking of person-directed action. *Journal of Personality and Social Psychology, 49*(5), 1221–1230. https://doi.org/10.1037/0022-3514.49.5.1221.

De Houwer, J. (2009). The propositional approach to associative learning as an alternative for association formation models. *Learning & Behavior, 37*(1), 1–20. https://doi.org/10.3758/LB.37.1.1.

De Houwer, J., Beckers, T., & Vandorpe, S. N. (2005). Evidence for the role of higher order reasoning processes in cue competition and other learning phenomena. *Learning & Behavior, 33*, 239–249. https://doi.org/10.3758/BF03196066.

Dickinson, A. (1980). *Contemporary animal learning theory*. London England: Cambridge University Press.

Dickinson, A., Hall, G., & Mackintosh, N. J. (1976). Surprise and the attenuation of blocking. *Journal of Experimental Psychology: Animal Behavior Processes, 2*(4), 313–322. https://doi.org/10.1037/0097-7403.2.4.313.

Dickinson, A., Shanks, D., & Evenden, J. (1984). Judgement of act-outcome contingency: The role of selective attribution. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology, 36*(1), 29–50. https://doi.org/10.1080/14640748408401502.

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. https://doi.org/10.3758/BF03193146.

FeldmanHall, O., & Dunsmoor, J. E. (2019). Viewing adaptive social choice through the lens of associative learning. *Perspectives on Psychological Science, 14*(2), 175–196. https://doi.org/10.1177/1745691618792261.

FeldmanHall, O., Dunsmoor, J. E., Kroes, M. C., Lackovic, S., & Phelps, E. A. (2017). Associative learning of social value in dynamic groups. *Psychological Science, 28*(8), 1160–1170. https://doi.org/10.1177/0956797617706394.

FeldmanHall, O., Dunsmoor, J. E., Tompary, A., Hunter, L. E., Todorov, A., & Phelps, E. A. (2018). Stimulus generalization as a mechanism for learning to trust. *Proceedings of the National Academy of Sciences, 115*(7), E1690–E1697. https://doi.org/10.1073/pnas.1715227115.

Funder, D. C. (1995). On the accuracy of personality judgment: A realistic approach. *Psychological Review, 102*(4), 652–670. https://doi.org/10.1037/0033-295X.102.4.652.

Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science, 21*, 177–182. https://doi.org/10.1177/0963721412445309.

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*(1), 21–38. https://doi.org/10.1037/0033-2909.117.1.21.

Jones, E. E., & Davis, K. E. (1965). From acts to dispositions the attribution process in person perception. In *Advances in experimental social psychology (vol. 2)* (pp. 219–266). Academic Press. https://doi.org/10.1016/S0065-2601(08)60107-0.

Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When more pain is preferred to less: Adding a better end. *Psychological Science, 4*(6), 401–405. https://doi.org/10.1111/j.1467-9280.1993.tb00589.x.

Lovibond, P. F. (2003). Causal beliefs and conditioned responses: Retrospective revaluation induced by experience and by instruction. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*(1), 97–106. https://doi.org/10.1037/0278-7393.29.1.97.

Lovibond, P. F., Been, S. L., Mitchell, C. J., Bouton, M. E., & Frohardt, R. (2003). Forward and backward blocking of causal judgment is enhanced by additivity of effect magnitude. *Memory & Cognition, 31*(1), 133–142. https://doi.org/10.3758/BF03196088.

Mackintosh, N. J. (1983). *Conditioning and associative learning*. New York: Oxford University Press.

Mehl, M. R., Gosling, S. D., & Pennebaker, J. W. (2006). Personality in its natural habitat: Manifestations and implicit folk theories of personality in daily life. *Journal of Personality and Social Psychology, 90*(5), 862–877. https://doi.org/10.1037/0022-3514.90.5.862.

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*(6), 532–552. https://doi.org/10.1037/0033-295X.87.6.532.

Pillai, D., Sheppard, E., & Mitchell, P. (2012). Can people guess what happened to others from their reactions? *PLoS One, 7*(11), Article e49859. https://doi.org/10.1371/journal.pone.0049859.

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: Appleton- Century-Crofts.

Sanbonmatsu, D. M., Akimoto, S. A., & Gibson, B. D. (1994). Stereotype-based blocking in social explanation. *Personality and Social Psychology Bulletin, 20*(1), 71–81. https://doi.org/10.1177/0146167294201007.

Shanks, D. R. (2010). Learning: From association to cognition. *Annual Review of Psychology, 61*, 273–301. https://doi.org/10.1146/annurev.psych.093008.100519.

Unkelbach, C., Fiedler, K., & Freytag, P. (2007). Information repetition in evaluative judgments: Easy to monitor, hard to control. *Organizational Behavior and Human Decision Processes, 103*(1), 37–52. https://doi.org/10.1016/j.obhdp.2006.12.002.

Waldmann, M. R. (2000). Competition among causes but not effects in predictive and diagnostic learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 26*(1), 53–76. https://doi.org/10.1037/0278-7393.26.1.53.

Waldmann, M. R., & Holyoak, K. J. (1992). Predictive and diagnostic learning within causal models: Asymmetries in cue competition. *Journal of Experimental Psychology: General, 121*(2), 222–236. https://doi.org/10.1037/0096-3445.121.2.222.

Weaver, K., Garcia, S. M., Schwarz, N., & Miller, D. T. (2007). Inferring the popularity of an opinion from its familiarity: A repetitive voice can sound like a chorus. *Journal of Personality and Social Psychology, 92*(5), 821–833. https://doi.org/10.1037/0022-3514.92.5.821.

Winter, L., & Uleman, J. S. (1984). When are social judgments made? Evidence for the spontaneousness of trait inferences. *Journal of Personality and Social Psychology, 47*(2), 237–252. https://doi.org/10.1037/0022-3514.47.2.237.

Wu, W., & Mitchell, P. (2019). Accurate inferences of others' thoughts depend on where they stand on the empathic trait continuum. *Personality and Individual Differences, 148*, 110–116. https://doi.org/10.1016/j.paid.2019.05.025.

Wu, W., Sheppard, E., & Mitchell, P. (2019). Interpreting signals in other people's behavior to sense things about them and to infer things about their world. *Social and Personality Psychology Compass, 13*(12), Article e12513. https://doi.org/10.1111/spc3.12513.