## S1 Text. Effect of action-distractor conflict on learning

For the sake of full disclosure and transparency, we present here a summary of the estimated parameters obtained by m7 – a model with separate learning rates as function of the interaction between choice and distractor-action congruency, plus the distractor bias parameter in the decision rule. Since this model did not win over the others in the models comparison, due to the extra model complexity, these results are only suggestive, and should be interpreted with care. We still thought it important to share these results, since the only weak evidence we found for an effect on action vs. distractor conflict on learning would be a benefit to learning. This goes against the hypothesis that conflict generally carries a cost to learning, due to its aversive nature, and is the opposite of what we found for conflict between instructions and subjective values. Finally, although the hypotheses discussed here remain speculative, they may offer relevant ideas for future research.

As in the other models, the estimated distractor bias parameter was significantly different from 0 (average $\varphi$ = 0.17 ± 0.25, $t_{19}$ = 3.05, p = .007, d = 0.96). The estimated learning rates (**Fig A**) were submitted to a repeated-measures ANOVA (choice: free vs. instructed; distractor-action congruency: congruent vs. incongruent). This showed a significant main effect of choice ($F_{1, 19}$ = 15.56, $p$ < .001, $\eta_p^2$ = 0.45), as learning rates were lower in instructed than free choices. It additionally showed a significant main effect of distractor-action congruency ($F_{1, 19}$ = 14.98, $p$ = .001, $\eta_p^2$ = 0.44), with higher learning rates in incongruent than congruent trials. Yet, these main effects were qualified by a significant choice-by-congruency interaction ($F_{1, 19}$ = 5.94, $p$ = .02, $\eta_p^2$ = 0.24). Post-hoc tests revealed that learning rates were significantly higher in incongruent than congruent trials in instructed trials ($t_{19}$ = -4.85, $p$ < .001, $d$ = -1.08), but there was no significant effect of congruency in free trials ($t_{19}$ = -0.12, $p$ = .906, $d$ = -0.03). Moreover, learning rates were always lower in instructed than in free trials (free-congruent vs. instructed-congruent: $t_{19}$ = 4.28, $p$ < .001, $d$ = 0.96; free-incongruent vs. instructed-incongruent: $t_{19}$ = 3.14, $p$ = .005, $d$ = 0.70).

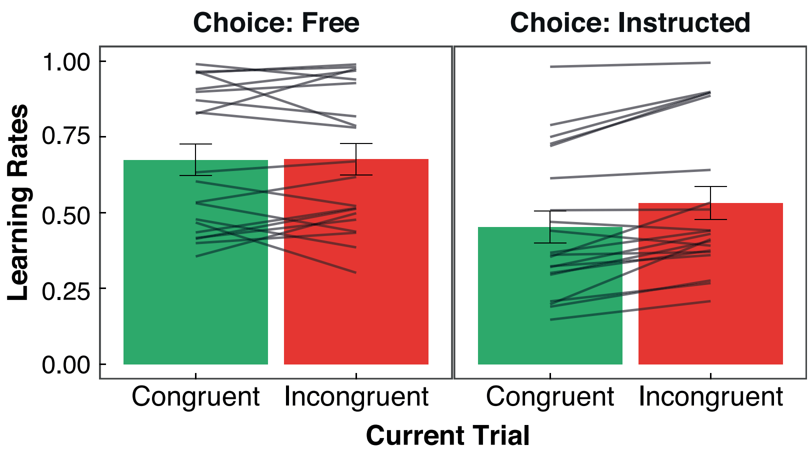


**Fig A. Estimated learning rates in m7.** Average learning rates ($\alpha$) in m7 as a function of choice and current trial distractor-action congruency. Error bars represent the standard error of the mean, and the lines represent each participant.

These results suggest that in instructed trials, the conflict triggered by incongruent distractors and targets might have led to an increase in learning rates, relative to no conflict. Yet, in free trials, action vs. distractor conflict did not influence learning. Recall that we found larger RTs cost in instructed than free trials due to conflict (see Results section). We hypothesised this was related to increased conflict costs associated with resolving both perceptual conflict (target vs. distractors), and at a response level (simultaneous activation of both responses), whereas in free choices, there was only conflict at the response level. This combined perceptual and response conflict in instructed trials may have required the deployment of more attentional resources to focus on the relevant stimuli, than in free trials. This might in turn result in enhanced attention at the time of the outcome, and thus in higher learning rates in incongruent than congruent trials.

It could have been argued that the absence of conflict effects on learning rates in free relative to instructed trials would be linked to differences in the effect sizes. RTs were slower in incongruent than congruent trials by an average of 66 ms in instructed trials, but only around 28 ms in free trials. This reduced effect might have thus been too weak to influence learning. However, the effects we found on learning rates due to conflict between instructions and subjective values (in m8) were associated with around 25 ms conflict costs on RTs (instructed low *minus* high value), similar to the cost of distractor-action conflict in free trials. Hence, the relatively smaller RTs costs in free trials cannot explain the absence of an effect on learning.

Therefore, the absence of effects of conflict on learning in free trials might rather be due to conflict being dealt with differently. Although free choices were still disrupted by incongruent distractors (evidenced by slower RTs, **Fig 2A**), such choices were likely driven by large differences in action value (as implied by our model, **Fig 3B**). We speculate that such chosen conflict might be subjectively experienced as different from imposed (or unavoidable) conflict. The extra effort might seem "justified" by the expected action values, rendering it less aversive, and cancelling out potential conflict costs on learning [1]. Furthermore, as participants might generally devote more attention to the task in free trials, due to a greater perceived relevance of information (as seen in the choice effect on learning rates), the attention at the time of the outcome might not be further modulated. In contrast, if participants are less engaged in the task in instructed trials, but are then obliged to pay attention to the task to successfully resolve conflict, this enhanced attention may then improve outcome processing, relative to the instructed-congruent trials.

We further suggest that having a choice in whether to experience conflict may partially explain why we did not find conflict costs on obtained rewards as previously reported with the Simon task [2], which also involves externally-triggered conflict. As mentioned in the introduction, during the learning phase of that study [2], participants had to respond according to stimuli, some of which were associated with conflict. Thus, the fact that conflict was unavoidable might have increased its subjective cost. Furthermore, their design closely mirrored other effort discounting tasks, wherein people learn how much effort is needed to obtain a reward and, subsequently, show a preference for low effort options [3]. In contrast, in our study, conflict with distractors was fully orthogonal to the learning task. This allowed us to investigate how learning might be dynamically influenced by an unpredictable, and task-irrelevant, experience of conflict, rather than offering the opportunity to learn to predict upcoming conflict. Future work is clearly needed to investigate the conditions under which conflict might discount rewards, or be its effects may be cancelled out by other mechanisms.

**References**


1. Schouppe N, Braem S, Houwer JD, Silvetti M, Verguts T, Ridderinkhof KR, et al. No pain, no gain: the affective valence of congruency conditions changes following a successful response. Cogn Affect Behav Neurosci. 2015;15: 251–261. doi:10.3758/s13415-014-0318-3

2. Cavanagh JF, Masters SE, Bath K, Frank MJ. Conflict acts as an implicit cost in reinforcement learning. Nat Commun. 2014;5: ncomms6394. doi:10.1038/ncomms6394

3. Vassena E, Silvetti M, Boehler CN, Achten E, Fias W, Verguts T. Overlapping Neural Systems Represent Cognitive Effort and Reward Anticipation. PLoS ONE. 2014;9: e91008. doi:10.1371/journal.pone.0091008

## S2 Text. The critical role of the distractor bias parameter

To further test the need to consider an effect of distractors at the decision stage, we assessed whether an alternative model without that parameter was able to reproduce the behaviour observed in our participants. For this, we compared the generative performance of the winning model (m8) to an alternative model (m8_alt), which differed only in that the decision rule followed the standard softmax rule (i.e. without the added $\varphi$, equivalent to $\varphi$ = 0; but with the remaining free parameters as in m8: [$\beta$, $\alpha_{Free} \neq \alpha_{Instructed\_High\ Value} \neq \alpha_{Instructed\_Low\ Value}$]). First, we fitted this model to the real data. Next, we used the average estimated parameter values to simulate data ($N$ = 100). **Fig B** shows the simulated data from the alternative model (m8_alt), overlaid over the real data and simulations from the winning model (m8, already displayed in **Fig 3A**). This shows that the alternative model, without the distractor bias parameter, fails to qualitatively reproduce the observed distractor bias effect, i.e. higher proportion of distractor congruent than incongruent choices, from the start of learning episodes. This bias is instead clearly observed in simulations of the winning model (m8), which included the distractor bias parameter.

Therefore, we believe the poor generative performance of the alternative model, which fails to capture an important behavioural effect, is sufficient cause for rejecting it as a suitable model [1], hence excluding other models without a distractor bias from our model space.

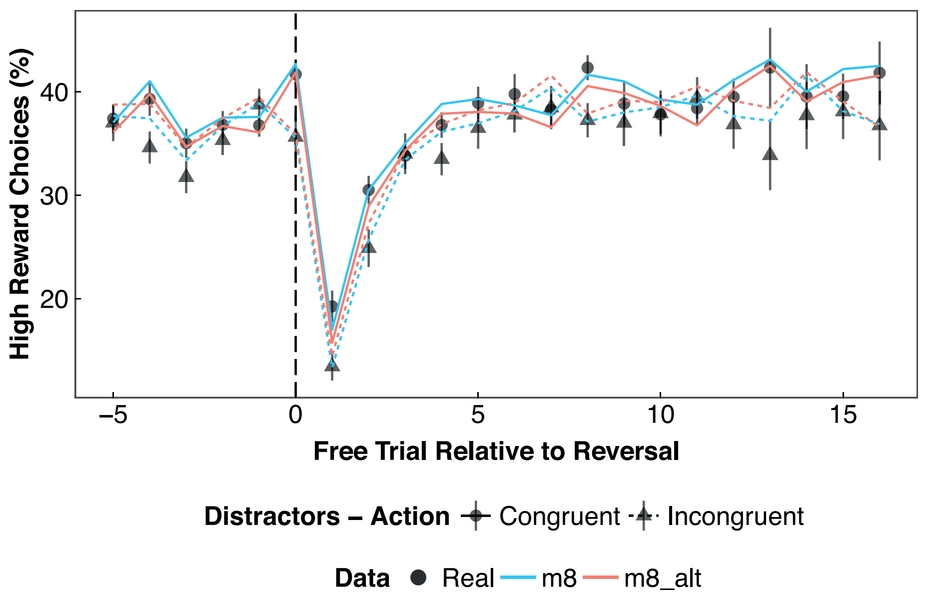


**Fig B. Model simulations.** Simulations for the winning model (m8) and a similar model without the distractor bias ($\varphi$) parameter as an alternative model (m8_alt). Dots and error bars represent real participants' data, while the lines represent the average of each model simulation.

### References

1. Palminteri S, Wyart V, Koechlin E. The Importance of Falsification in Computational Cognitive Modeling. Trends Cogn Sci. 2017; doi:10.1016/j.tics.2017.03.011

## S3 Text. Distractor bias parameter vs. behaviour correlations

To investigate the relation between conflict avoidance and conflict adaptation effects, we assessed the relation between the estimated distractor bias parameter (as an index conflict avoidance) and conflict adaptation on RTs. Conflict adaptation effects were calculated as the difference between conflict effects (I *minus* C) for previously *congruent* minus previously incongruent trials. Thus, larger conflict adaptation reflects a greater reduction in conflict effects following incongruent trials. Since similar conflict adaptation was observed for free and instructed trials, we averaged over choice conditions. This analysis revealed a significant positive correlation between the distractor bias parameter, i.e. conflict avoidance, and conflict adaptation effects on RTs (see **Fig C.i**, Pearson's correlation: $r$ = 0.54, $t_{18}$ = 2.72, $p$ = .014). That is, participants who were better able to adapt their behaviour to reduce conflict costs on RTs were also more likely to avoid conflict when unnecessary (i.e. in the absence of strong value differences).

These results should be interpreted with care, given our relatively small sample size. Nonetheless, they suggest that participants' sensitivity to conflict may be reflected in these two types of adaptive behaviours, rather than being a trade-off between them. It could have been hypothesised instead that participants who were worse at minimising RT costs would benefit most from avoiding conflict. Yet, this correlation implies that a common process of conflict monitoring and adaptation may underlie both types of behavioural responses. In fact, previous work has suggested that conflict signals can trigger both adjustments in cognitive control and conflict avoidance ([1–3] but see [4]).

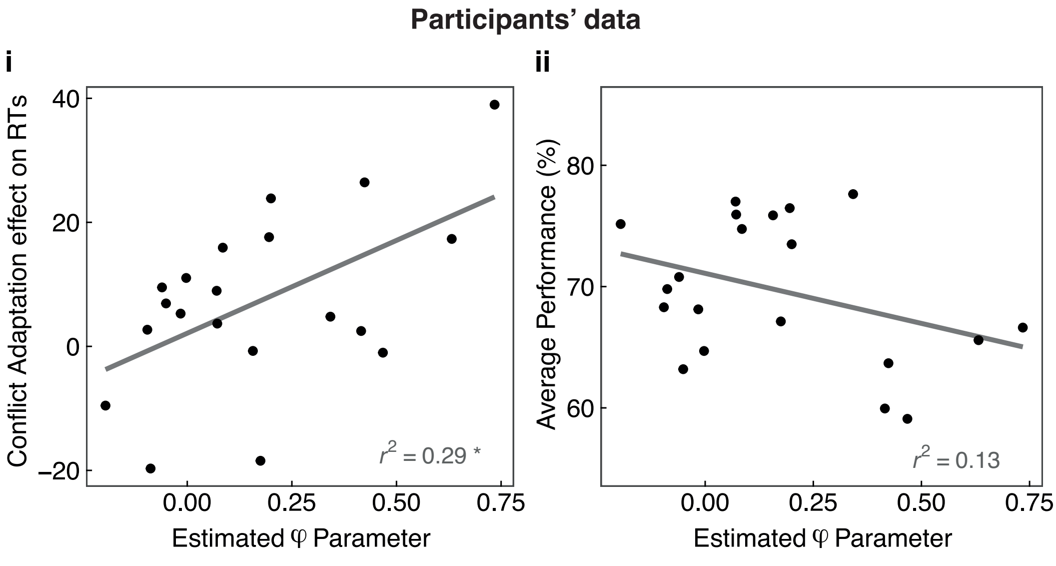


**Fig C. Relation between distractor bias parameter and participants' behaviour.** Correlations between the estimated distractor bias parameter ($\varphi$) and conflict adaptations effects on RTs (**i**), and average performance (**ii**).

Finally, it could have been hypothesised that having a larger choice bias might impair performance in the task, as participants choices might be too driven by the distractors rather than action values. Importantly, since the probably of left and right distractors was equate within each learning episode (i.e. between reversals), following the distractors' suggestion would be equally likely to be helpful vs. unhelpful to task performance (i.e. 50/50 chance). Nevertheless, we tested this hypothesis by assessing the correlation between the estimated distractor bias parameter and average task performance, which showed no significant correlation (**Fig C.ii**, Pearson's correlation: $r$ = -0.36, $t_{18}$ = -1.63, $p$ = .12).

The independence of distractor bias effects from average performance was further corroborated through model simulations. Virtual datasets ($N$ = 100) were simulated across a range of distractor bias ($\varphi$) values ([-2, 2], at intervals of 0.1; and constant $\beta$ = 2, $\alpha$ = 0.6). The simulated virtual choices were then used to calculate the average performance (i.e. average proportion of high reward choices, **Fig D.i**), as well as the percentage of distractor congruent choices (**Fig D.ii**), i.e. the consequence of the simulated distractor bias effect. These findings show that, across this broad range of $\varphi$ values ($\varphi$ estimated on participants' data varied less than 1 unit), changes in average performance were minimal, whereas they were associated with very large differences in the effect of distractor bias on free choices (i.e. proportion of distractor congruent vs. incongruent choices). This confirms that neither our task nor our model implies that the distractor bias would result in a significant impairment in task performance.

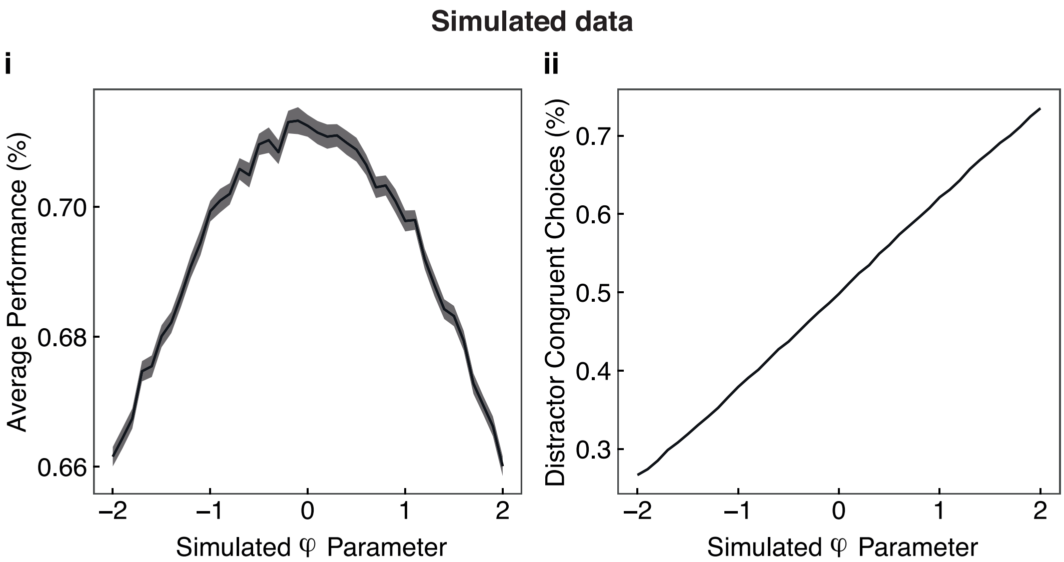


**Fig D. Effect of varying distractor bias parameter on simulated behaviour.**
Simulated data to assess the effect of varying the distractor bias parameter (range of [-2, 2], at intervals of 0.1; constant $\beta$ = 2, $\alpha$ = 0.6) on average performance (**i**), and on the percentage of distractor congruent choices (**ii**).

**References**

1. Botvinick MM. Conflict monitoring and decision making: Reconciling two perspectives on anterior cingulate function. Cogn Affect Behav Neurosci. 2007;7: 356–366. doi:10.3758/CABN.7.4.356

2. Dignath D, Kiesel A, Eder AB. Flexible conflict management: Conflict avoidance and conflict adjustment in reactive cognitive control. J Exp Psychol Learn Mem Cogn. 2015;41: 975–988. doi:10.1037/xlm0000089

3. Dreisbach G, Fischer R. If it's hard to read… try harder! Processing fluency as signal for effort adjustments. Psychol Res. 2011;75: 376–383. doi:10.1007/s00426-010-0319-y

4. Schouppe N, Ridderinkhof KR, Verguts T, Notebaert W. Context-specific control and context selection in conflict tasks. Acta Psychol (Amst). 2014;146: 63–66. doi:10.1016/j.actpsy.2013.11.010

## S4 Text. Effect of distractor vs. value conflicts in action selection in instructed trials

In instructed trials, participants could experience both (a) action-distractor conflict, and (b) instruction-value conflict. Although participants were instructed to avoid making errors, as these would reduce their final earnings, participants could have still thought that the errors might be less costly than losing 1 point (i.e. seeing the "-1" feedback, rather than the error "X"). Consequently, it remained possible that they might strategically choose to commit errors in instruction-value conflict trials (i.e. instructed-low value), to avoid the loss feedback.

To investigate these hypotheses, we compare the influence of two types of conflict in action selection in instructed trials, submitting reaction times (RTs) and error rates to a 2 x 2 repeated-measures ANOVA, as a function of current trial distractor-action congruency: congruent vs. incongruent, and of the expected value of the instructed action: high vs. low. As for the RT analysis shown in **Fig 3D**, trials were split into high vs. low value instructions based on simulating action values ($Q$ values) in each trial, for each participant (based on each participant's estimated parameters of m8).

***Reaction Times***

Analysis of RTs showed a significant main effect of congruency ($F_{1, 19}$ = 145.95, $p$ < .001, $\eta_p^2$ = 0.88), a significant main effect of action value ($F_{1, 19}$ = 28.65, $p$ < .001, $\eta_p^2$ = 0.60), and a significant interaction effect ($F_{1, 19}$ = 8.00, $p$ = .01, $\eta_p^2$ = 0.30). This shows that both factors influenced RTs. Moreover, as seen in **Fig E.i**, congruency had a larger effect on RTs (~66ms slower in incongruent vs. congruent trials), overall, than instructed action values (~25ms slower in low vs. high value trials). Post-hoc tests revealed that all pairwise comparisons were significantly different (all $ps$ < .001). The interaction effect revealed that the congruency effect (I-C) was significantly larger in high (~71ms) than in low (~61ms) value trials ($t_{19}$ = 2.83, $p$ = .011, $d$ = 0.63).

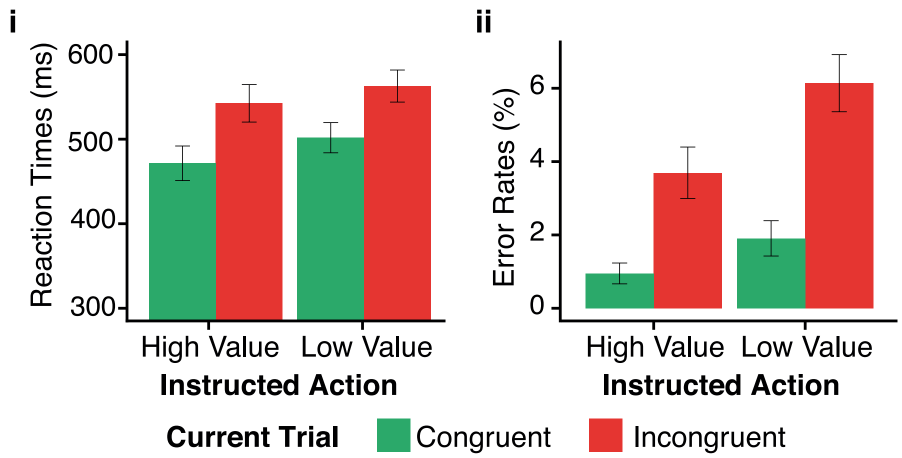


**Fig E. Action selection in instructed trials.** Average reaction times (**i**) and error rates (**ii**) in instructed trials as a function of distractor-action and instruction-value conflicts. Error bars represent the standard error of the mean.

***Error Rates***

Analysis of the error rates showed a significant main effect of congruency ($F_{1,19}$ = 33.19, $p$ < .001, $\eta_p^2$ = 0.64), a significant main effect of action value ($F_{1,19}$ = 16.02, $p$ < .001, $\eta_p^2$ = 0.46), and a marginal interaction effect ($F_{1,19}$ = 3.66, $p$ = .07, $\eta_p^2$ = 0.16). This shows that participants made more errors in incongruent trials, as well as in trials in which they were instructed to make the subjectively low value action. As seen in **Fig E.ii** congruency had a larger effect on errors (~3.5% more errors in incongruent trials), overall, than instructed action values (~1.7% more errors in low value trials).

As the interaction effect remained marginal, and we did not have specific *a priori* hypotheses about it, it was not tested further. Inspection of **Fig E.ii** points to a relatively larger congruency effect for low (~4.2%) than for high value (~2.7%) instructed actions. This could suggest that it was especially hard for participants to respond correctly when they experienced combined conflict between target and distractors, and between target instruction and internal values. This scenario (incongruent-low value) involves both the salient distractors and the internal values activating the same action (e.g. right), thus potentially requiring more cognitive control to suppress the inadequate activation in order to respond correctly (e.g. left), though this interpretation remains speculative.

***Summary***

In short, incongruent distractors were the predominant cause for disruption of action selection, as evidenced by increased error commission and slowing down RTs. This is consistent with participants generally trying to follow the instructions, but occasionally being too impulsive to suppress their internal motivations, rather than participants strategically committing errors when they disagreed with the instruction.

**S1 Table. Reduced model space comparison results.**

| **Models** | **AIC** ± SD | **Model Frequency** | **Exceedance Probability ($xp$)** |
|---|---|---|---|
| m1: Standard RL [$\beta$, $\alpha$] | 723.3 ± 191.4 | 0.14 | 0.02 |
| m2: [$\beta$, $\varphi$, $\alpha$] | 714.3 ± 183.5 | 0.10 | 0.00 |
| m3: [$\beta$, $\varphi$, $\alpha_C \neq \alpha_I$] | 715.2 ± 184.0 | 0.06 | 0.00 |
| **m4: [$\boldsymbol{\beta}$, $\boldsymbol{\varphi}$, $\boldsymbol{\alpha_{Free} \neq \alpha_{Instructed}}$]** | **704.7 ± 176.8** | **0.43** | **0.97** |
| m5: [$\beta$, $\varphi$, $\alpha_{Free\_C} \neq \alpha_{Free\_I} \neq \alpha_{Instructed}$] | 704.9 ± 178.4 | 0.09 | 0.00 |
| m6: [$\beta$, $\varphi$, $\alpha_{Free} \neq \alpha_{Instructed\_C} \neq \alpha_{Instructed\_I}$] | 705.2 ±177.4 | 0.06 | 0.00 |
| m7: [$\beta$, $\varphi$, $\alpha_{Free\_C} \neq \alpha_{Free\_I} \neq \alpha_{Instructed\_C} \neq \alpha_{Instructed\_I}$] | 705.4 ± 178.8 | 0.12 | 0.01 |

$\beta$ refers to choice temperature parameter, reflecting choice stochasticity. φ refers to the distractor bias parameter added to the decision rule. α refers to the learning rate, split by conditions. Winning model (m4) highlighted in bold. RL = Reinforcement Learning, C = Congruent, I = Incongruent. AIC = Akaike Information Criteria, SD = Standard Deviation.