Causal Transfer of Specific Attentional Control-states

Imagine your morning commute: the traffic, the complicated streets, the sea of angry, harassed drivers. How do you manage these demands on your attention? One possibility is that people approach a new situation with a heightened sense of attention, because similar environments have proven difficult in the past. In order for this mechanism to work, people would need to bind together a specific attentional control-state (e.g., heightened focus of attention) and their memory of the demanding environment. This type of context-control-learning has indeed been demonstrated in a number of studies: probabilistic cues (e.g., stimulus location or color) can implicitly facilitate the retrieval of context-appropriate attentional control-states (e.g., high attentional focus)[1,2]. For example, an intersection with several car accidents monthly might get bound to a high attentional control-state, while the residential street outside your house would conversely require relatively less attentional focus. Contextual cues can thus guide strategic adjustment to control demands. Such stimulus-control-learning is highly adaptive, combining the speed of automatic processing with the flexibility of cognitive control[3].

Yet greater flexibility could be achieved if learned control-states were transferred across associated stimuli or contexts. Indeed, in a recent study, we found that, like reward[4], stimulus-control associations are transferrable: When stimulus A predicts stimulus B, and stimulus B is subsequently paired with a specific control-state, participants transfer B’s attentional control settings to A[5]. This suggests that learned control-states and their associated cognitive strategies can be generalized across contexts. Here, in two experiments, we tested whether this stimulus-to-stimulus transfer learning of cognitive control depends on *how* the associations are learned.

The basic task[5] consisted of a stimulus-stimulus (S-S) association phase, a stimulus-control (S-C) association phase, and a stimulus-control transfer (S-CT) phase (Figure 1). In the

S-S phase, face or house (S1) images repeatedly preceded the presentation of scene (S2) images, in order to foster S1-S2 paired-associates in memory. Participants were asked to categorize the S2 images according to scene-specific response mappings. Next, in the S-C phase, S2 images preceded color-word stimuli in a Stroop task. Crucially, S2 images served as implicit probabilistic cues: half were predictive of congruent trials (low control-demand) and the other half were predictive of incongruent trials (high control-demand). Finally, in the S-CT phase, S1 images, instead of S2 images, preceded the Stroop stimuli, but did not have any predictive relationship with congruency. This allowed us to test whether any S2 control-demand associations acquired during the S-C phase transferred to their S1 paired-associates. Control transfer learning would be evident if S1 images associated with S2 high control-demand images produced a smaller congruency effect in the transfer phase compared to S1 images associated with S2 low control-demand images.

In Experiment 1, S1 images, rather than S2 images, predicted Stroop congruency, and S2 images assessed transfer in the S-CT phase. In our previous work[5], we linked stimulus pairs, and stimuli and attentional control-states, in a causal chain (S1 → S2 → control-state) structure so that each association is reactivated to facilitate transfer along the chain. However, learned stimulus-control associations may also transfer across linked cues through knowledge of the covariational information provided by a common-cause (S1 → S2, S1 → control-state) structure (Figure 2). If participants learn stimulus-control associations through a common-cause structure, this indicates that transfer effects are observed implicitly through associative learning: here, the S1 images act as temporal causes that predict S2 image presentation and control-state formation, so if associative transfer via the S2 transfer probes is observed, this occurs independently of causal learning. However, if participants cannot learn these associations through a common-cause structure, this implies both that causal learning plays a significant role in the transfer of control-states and that strategy or reasoning adjustments are possible, because causal relationships are consciously observable and prone to manipulation.

In Experiment 2, the S-C phase preceded the S-S phase. One possible mechanism for *how* control states are transferred through a causal chain is through mediated learning (Figure 2). As with reward[4], hippocampal interactions with the striatum might facilitate transfer such that S1 images accrue model-based value downstream through the S1-S2 paired association. However, this may also occur through direct control-learning, whereby S1 images accrue immediate model-free value, because S2 images evoke a representation of S1 images during S-C conditioning[6]. Here, with the S-C conditioning phase first, transfer cannot occur through the acquisition of model-free value. This has direct implications for the brain regions involved in control-state transfer as well as the role that dopamine may play in learning control.

With 44 participants per experiment, we found that control-state transfer occurred only through a causal chain structure and likely through mediated learning (Figure 3). In the S-S phase, we replicated our main effect of cue validity[5]: validly cued S2 images were categorized more quickly than invalidly cued scenes across experiments ($p = 0.034$, $p = 0.028$). Control-learning was not observed in Experiment 1 (run x cue x congruency: $p = 0.606$), but was observed in Experiment 2 ($p = 0.032$). Both Experiments show no transfer effects ($p > 0.860$).

This study addresses the important question of how mental task structure can affect stimulus-control-learning. While transfer of stimulus-reward associations has been shown to occur[5], a hotly debated question is the extent to which cognitive control is "like" reward and thus subject to associative learning[3,7]. Here, we have shown that control-states are transferred through causal mechanisms and that transfer cannot occur through model-free value accumulation. This suggests that individuals can strategically modulate how they respond to attentionally demanding environments and that one potential difference between reward and control-learning is the role that dopamine plays. Because dopamine prediction errors are thought to underlie model-free learning, our results suggest that control-learning is not accompanied by the release of dopamine.

The implications of this research are also crucial for mental health. An individual can only generalize cognitive strategies over similar contexts if they first learn to associate certain cues or contexts with the need for control. However, many clinical conditions, such as schizophrenia or dementia, are characterized by context-inappropriate behavior, including the inability to deviate from routine or learn from feedback. These are typically attributed to failures of control, but might instead arise from a deficit in control-learning. In our control-state transfer paradigm, this is evident in two ways: first, if individuals fail to learn control-state associations, and second, if control-states transfer but persist even when the contextual cues are no longer helpful. A better understanding of the interaction between control and memory mechanisms will lay the foundations for understanding potential control-learning failure modes – and possible therapeutic approaches – in psychiatric and neurodegenerative disorders.

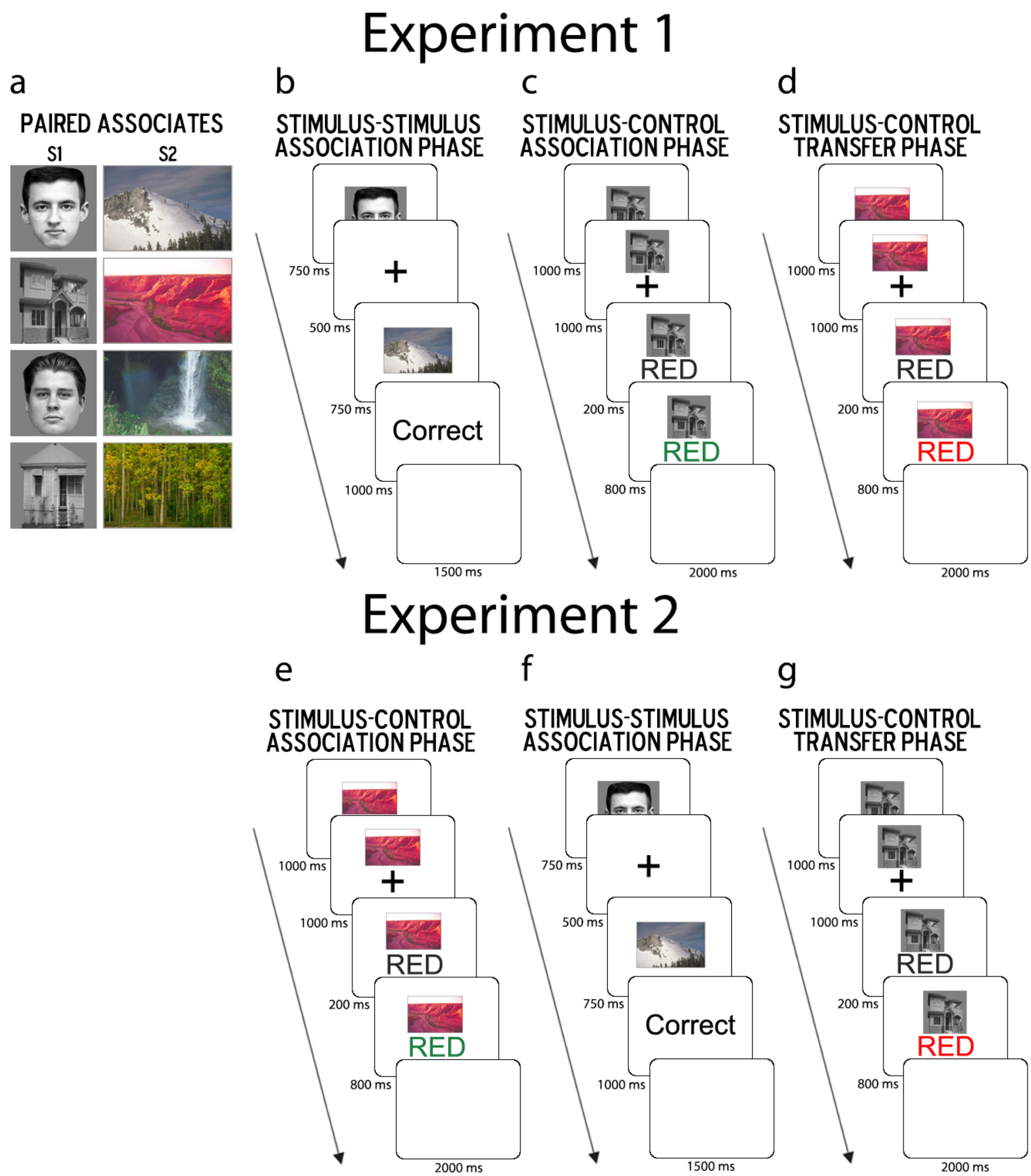

Figure 1: Summary of task procedure. (**A**) Stimuli consisted of two grayscale male face images from Kanade, Cohn, and Tian (2000), two grayscale house images from online real estate websites, and four scene images (mountain, canyon, waterfall, forest). Face and house images were used for their known neural selectivity, in anticipation of running an fMRI study at a later date. **(B** and **F)** In the Stimulus-Stimulus (S-S) association phase, a face or house (S1) repeatedly predicted a particular scene (S2) image to form paired associates in memory. **(C** and **E**) In the Stimulus-Control (S-C) association phase, S1 (Experiment 1) or S2 (Experiment 2) images predicted stimulus congruency in a Stroop task to create implicit control-demand cues. (**D** and

G) In the Stimulus-Control Transfer (S-CT) phase, S2 (Experiment 1) or S1 (Experiment 2) transfer probes likewise preceded the onset of Stroop trials but were not predictive of congruency. Experiment 2 differed from Experiment 1 in that the S-C phase preceded the S-S phase, and Experiment 1 differed from Bejjani, Zhang, and Egner (2018) in that S1 images acted as predictive cues in both the S-S and S-C phase.

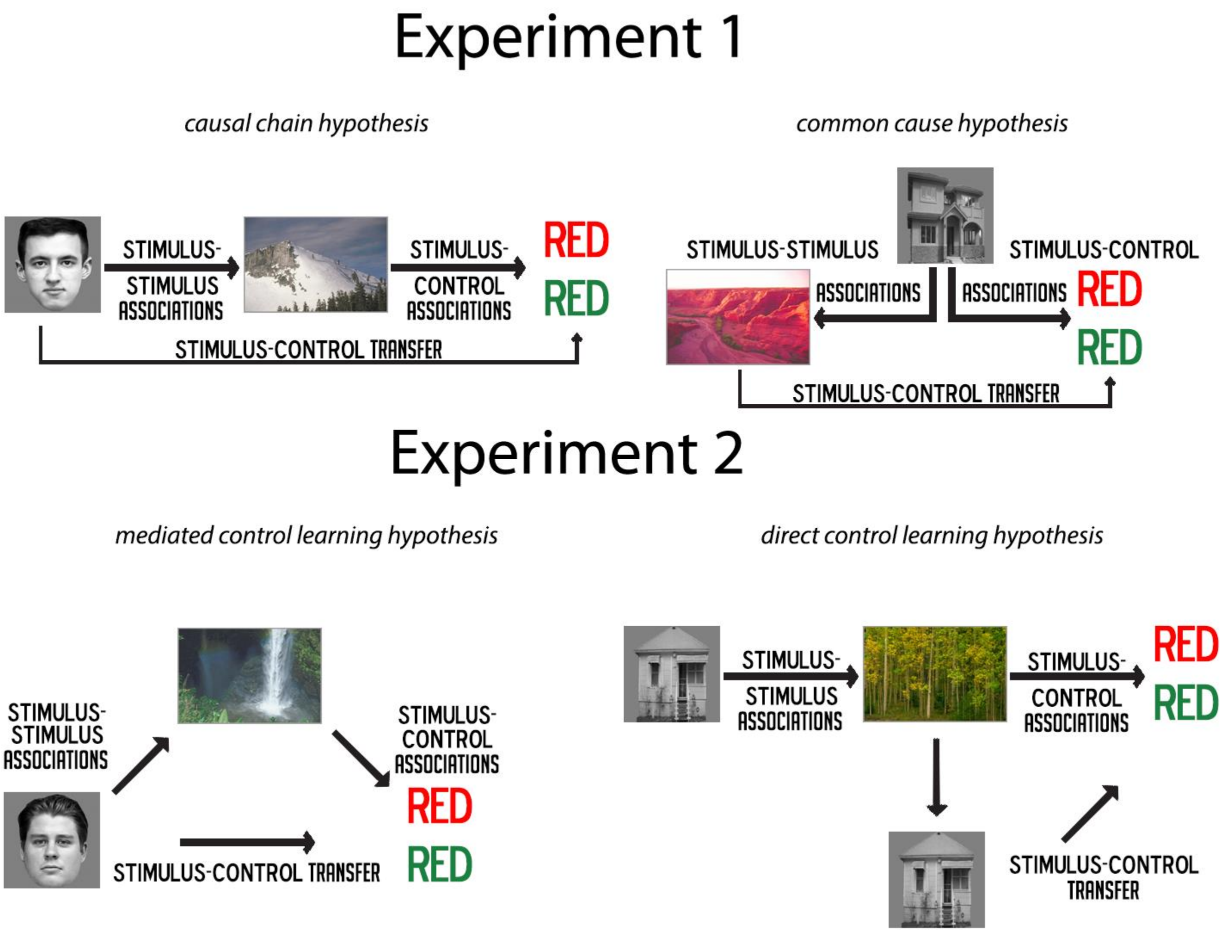

Figure 2: Summary of hypotheses tested. Experiment 1 adjudicated between the causal chain and common cause hypotheses. Previously, we found that control-state transfer can occur through a causal chain (Bejjani, Zhang, and Egner, 2018), whereby face/house images cued scene images and scene images cued specific attentional control-states, causing the transfer of learned control associations through the face/house-scene paired-associate chain. However, learned stimulus-control associations may also transfer across linked cues through knowledge of the covariational information provided by a common-cause (face/house → scene, face/house → control-state) structure. Experiment 2 adjudicated between the mediated and direct control-learning hypotheses. Because transfer can occur through a causal chain (Bejjani, Zhang, Egner, 2018), it is possible that control is learned through mediated hippocampal interactions with the striatum ("mediated learning"; cf. Wimmer and Shohamy, 2012), whereby face/house images accrue model-based value downstream through the chain of associations. However, it is also possible that face/house images accrue immediate model-free value during the S-C phase (cf. Sharpe et al., 2017): here, scene images evoke a representation of face/house images during conditioning so that face/house images become directly associated with learned control-states

and show the desired transfer effect in the S-CT phase.

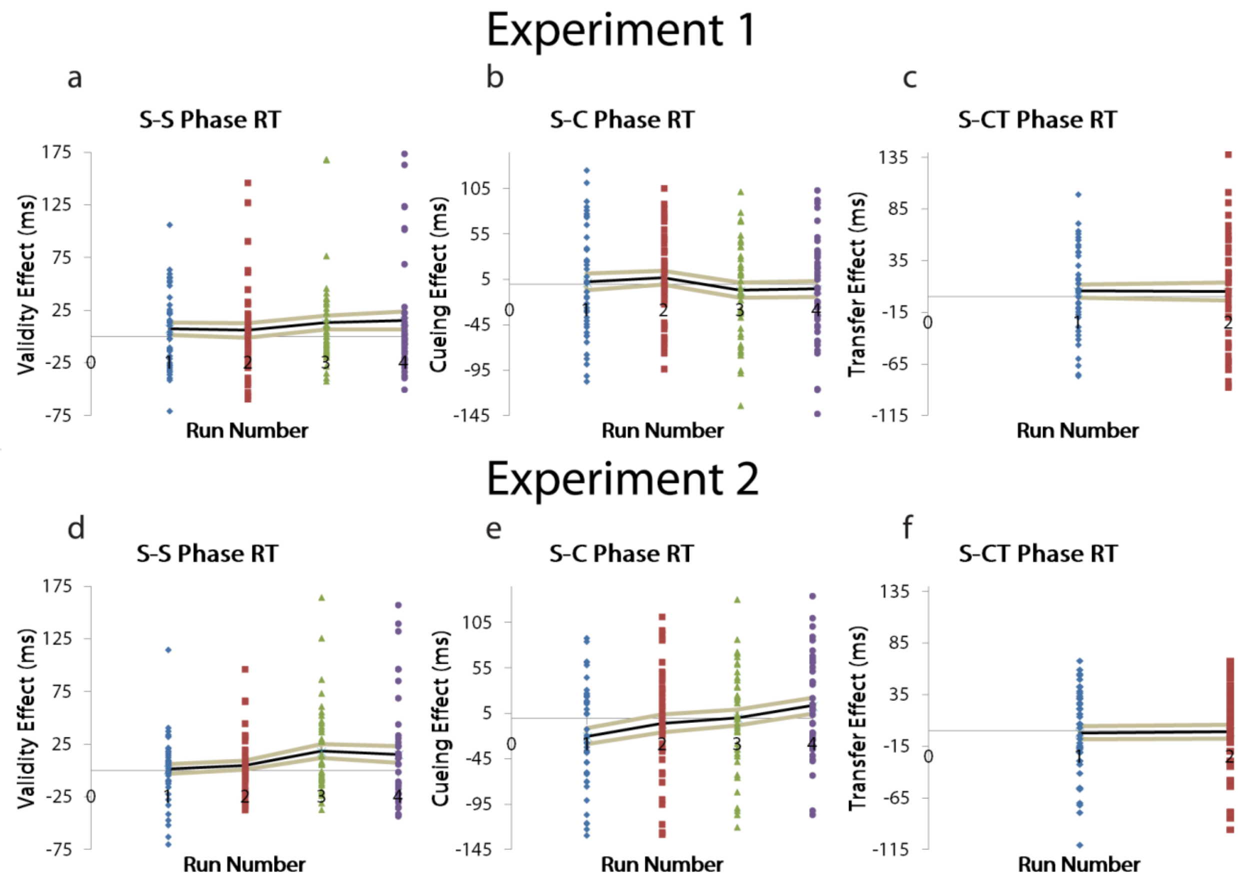

Figure 3: Experiments 1 and 2 results. (**a** and **d**) S-S phase mean RT validity effects (invalid - valid), (**b** and **e**) S-C phase mean RT cueing effects (low demand cue congruency effect - high demand cue congruency effect), and (**c** and **f**) S-CT phase mean RT transfer effects (low demand transfer probe congruency effect - high demand transfer probe congruency effect) are plotted in black, ± SEM in grey, as a function of run number.

References:

1. Bugg, J. M. (2017). Context, Conflict, and Control. In: T. Egner (Ed.), *The Wiley Handbook of Cognitive Control*, pp. 79-96. Oxford: Wiley-Blackwell.
2. Bugg, J. M., & Crump, M. J. (2012). In support of a distinction between voluntary and stimulus-driven control: a review of the literature on proportion congruent effects. *Frontiers in Psychology, 3*, 367.
3. Egner, T. (2014). Creatures of habit (and control): a multi-level learning perspective on the modulation of congruency effects. *Frontiers in Psychology,* 5, 1247.
4. Wimmer, G. E., & Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science, 338*(6104), 270-273
5. Bejjani, C., Zhang, Z., and Egner, T. (2018). Control by association: Transfer of implicitly primed attentional states across linked stimuli. *Psychonomic Bulletin and Review*, 25(2): 617-626.
6. Sharpe, M. J., Batchelor, H. M., & Schoenbaum, G. (2017). Preconditioned cues have no value. *ELife*, *6*.
7. Abrahamse, E., Braem, S., Notebaert, W., & Verguts, T. (2016). Grounding cognitive control in associative learning. *Psychological Bulletin,* 142(7), 693-728.