The cost of structure learning

Anne G.E. Collins
Department of Psychology, UC Berkeley

Abstract: Human learning is highly efficient and flexible. A key contributor to this learning flexibility is our ability to generalize new information across contexts that we know require the same behavior, and to transfer rules to new contexts we encounter. To do this, we structure the information we learn and represent it hierarchically as abstract, context-dependent rules, that constrain lower level stimulus-action-outcome contingencies. Previous research showed that humans create such structure even when it is not needed, presumably because it usually affords long-term generalization benefits. However, computational models predict that creating structure is costly, with slower learning and slower reaction times. We tested this prediction in a new behavioral experiment. Participants learned to select correct actions for four visual patterns, in a setting that either afforded (but did not promote) structure learning or enforced non-hierarchical learning, while controlling for the difficulty of the learning problem. Results replicated our previous finding that healthy young adults create structure even when unneeded, and that this structure affords later generalization. Furthermore, they supported our prediction that structure learning incurred a major learning cost, and that this cost was specifically tied to the effort in selecting abstract rules, leading to more errors when applying those rules. These findings confirm our theory that humans pay a high short-term cost in learning structure to enable longer-term benefits in learning flexibility.

Introduction:

Human learning is highly efficient and flexible. One key factor in its efficiency is our ability to simplify a problem by distinguishing its relevant aspects from irrelevant ones, and to generalize learned information to different contexts. Consider someone who has always lived in a warm climate vacationing at a ski resort for the first time. This person will learn a number of new ways to behave, such as dressing in the morning (more warmly!), walking in the street (with more careful steps), and driving (slower). If this vacationer later moves to a cold climate, she will be able to transfer these learned skills, rather than needing to relearn them. Creating a set of behavioral tools that are not tied to the context in which they were learned (in the mountains at a ski resort on vacation), but that can be reused as a package in other appropriate contexts (any time there is snow or ice on the ground), greatly speeds up behavioral adaptation. This ability to transfer knowledge is crucial for human intelligence, and mimicking this ability is an active area of research in artificial intelligence(Mesnil et al. 2011; Parisotto et al. 2015; Bengio et al. 2012).

We refer to structure learning as our ability to learn not only stimulus-action-outcome contingencies, but to hierarchically structure them into groups, called task-sets or rules, that correspond to higher level, abstract choices or strategies, and that themselves can be selected in different contexts. One example of structure learning is identifying existing structure in the problem. For example, healthy human adults are able to infer which dimensions or features of a stimulus are relevant (e.g. the snow on the ground, but not the presence or absence of skiers, the time of the day or the color of their jacket), thus greatly simplifying a learning problem from a complex high-dimensional state space to a simpler, lower dimensional one (Wilson & Niv 2011; Niv et al. 2015). They can also do this in a conditional, or hierarchical fashion, such that they identify that some features are relevant in a given context, but not in another, and vice-versa (Badre et al. 2010; Badre & Frank 2011; Frank & Badre 2011). Structure learning is not limited to pruning irrelevant sensory information from the state representations: learners also create structure as a way to share rule information between different contexts that cluster together (Collins & Frank 2016b), and can do so whether the contexts are observable (either snow or ice on the ground means you should dress warmly) or only latent (it's currently winter in a cold

climate, which requires the same rule), inferred as the current temporal context(Collins & Koechlin 2012; Donoso et al. 2014).

Humans create structure when learning, and such structure learning can help identify ways to simplify learning in a given environment and to generalize and transfer information. We recently showed that structure learning occurs a priori, even in environments where there is no benefit to learning structure(Collins & Frank 2013; Collins et al. 2014; Collins & Frank 2016a). This tendency may reflect a prior belief that learning structure is usually beneficial, potentially reflecting statistics of our environment, where contexts are infinitely more changing than our repertoire of behaviors, such that we usually have opportunities to generalize our skills to new situations. This is in line with other research showing that humans tend to see patterns where they do not exist statistically or are not useful(Yu & Cohen 2009). Our adult bias probably reflects a very strongly ingrained tendency to create structure, since we found evidence for such a priori structure learning and generalization even in 8-month old infants(Werchan et al. 2015; Werchan et al. 2016).

Learning structure is thought to rely on cognitive control, and be dependent on prefrontal cortex(Badre & Frank 2011; Donoso et al. 2014; Collins et al. 2014; Werchan et al. 2016; Niv et al. 2015). The use of rules in cognitive control has long been shown to be behaviorally costly, as measured by a behavioral switch-cost(Monsell 2003), and recent research has more generally shown that cognitive control is effortful(Kool & Botvinick 2014; Kool et al. 2013; Westbrook et al. 2013; Westbrook & Braver 2015), so that we trade off the expected benefits of control with the cost of the associated mental effort. We thus predict that structure learning should be costly. We further ask whether this cost may discourage building structure in environments where it is not beneficial.

We explore this topic by asking two questions: is structure learning behaviorally costly, and does this cost make us less likely to learn structure over time if it is not beneficial? To answer these questions, we extend our previous structure learning protocol (Collins & Frank 2013; Collins et al. 2014; Collins & Frank 2016a), where participants learned to select correct actions for four visual patterns, combining two features of two sensory dimensions (e.g. colored shapes; see Fig.

1). We previously showed that rather than learning four associations between each colored shape and each action ("flat" learning, Fig. 1B), participants represented this problem in a more

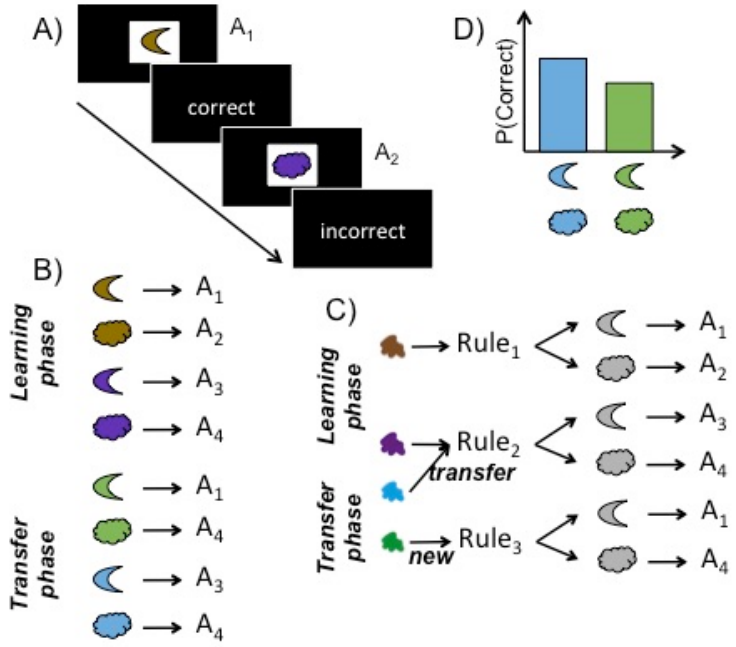

Figure 1: A priori structure learning protocol. This figure describes the protocol used in (Collins & Frank 2013; Collins et al. 2014) to show that participants learn structure a priori, as well as one important behavioral marker of structure learning. **A)** We study structure learning in the context of learning from reinforcement: subjects needed to respond to each colored shape by pressing one of four keys (A_1-A_4), and receive truthful feedback indicating if their choice was correct, allowing them to learn associations. **B)** Participants first learn about four images (e.g. colored shapes) presented in randomly interleaved order (learning phase); then about four new images in the transfer phase. The new images share features along one dimension with the old images (here shape), but incorporate new features along the other dimension (here color). **C)** Our previous research shows that instead of representing the learning problem as in B), participants build structure, whereby one dimension (e.g. here color) cues a rule, which itself cues stimulus action associations. Note that creating structure during the learning phase does not simplify the problem initially (comparing B and C) **D)** One behavioral marker of such structure learning is observing transfer in the transfer phase, where one of the two new colors allows generalization of an old rule, while the other requires learning a new rule. Structure learning predicts that we can generalize an old rule to a new context, leading to higher performance (represented schematically here as more accurate choices for the blue than green shapes.

complex, hierarchically structured way (Collins & Frank 2013; Collins et al. 2014). Specifically, they used one dimension (e.g. color; Fig 1C) as a context that cued one of two rules, thus learning two high level associations between contexts and rules. For each rule, they also learned two associations between the other dimension (e.g. shape) and actions, thus combining four low

level associations. Thus, in addition to the higher complexity of a hierarchical representation and of having to create rules, learning this task with structure required participants to create six associations, rather than four with flat learning. This suggests a high cost for structure learning, in an experimental design where there was no benefit to this structure. Our neural network model of hierarchical structure learning further predicted that this cost should lead to slower learning, and slower reaction times (Collins & Frank 2013).

Here we test this prediction by directly comparing learning in this experimental setup to equivalent learning of four associations in a "flat" way, where there is no opportunity for creating structure. Furthermore, we distinguish this prediction from that of a model of choice conflict: when structure learning is possible, the appropriate action for a given stimulus is different in different contexts, which leads to choice conflict and thus to a behavioral cost. We predict that the cost of structure learning cannot be reduced to conflict cost. Last, we test whether repeatedly encountering similar environments in multiple blocks lessens participants' tendency to build structure as they observe the cost and the lack of benefits. Our results confirm our prediction that structure learning is costly, and that this cost cannot be reduced to conflict cost. Furthermore, we show that despite this cost, participants learn structure a priori even when it is not useful. This bias for structure learning lead participants to select fewer correct actions and to have slower reaction times than when learning without structure; it did not disappear with practice, and it afforded later transfer and generalization of rules.

Methods:

Experimental design. This experiment included thirteen independent learning blocks, with non-overlapping sets of visual patterns for each block. Each block was independent from the others, and required learning the correct action for a new set of patterns. At each trial, subjects were presented with a single pattern consisting of two images presented side-by-side on a black background screen. Subjects could make one of four choices by pressing one of four keys with their right hand fingers. They had 1.5s to answer. For each pattern, a single choice always lead to *correct* feedback, while the other three choices consistently lead to *incorrect* feedback. Feedback was presented as a white word on black background for 0.5s, 0.3s after choice. It was followed by 0.75s fixation cross before the start of the next trial. Input-pattern order presentation across trials was pseudo-randomized to ensure equal presentation, and equal frequency of first-order transitions.

Experimental conditions: In six of the first 12 blocks (called *2-d blocks,* Fig. 2D), the two images forming a pattern came from two distinct "dimensions" of images (e.g., one image from a color dimension and one from a shape dimension), with two images in each dimension (eg. brown and purple colors, moon and cloud shapes). This forms four distinct visual input patterns (Fig. 2C). Data from these six blocks was analyzed in previous work to investigate the specific nature of structure created by participants; this previous work did not investigate the cost of structure learning (Collins & Frank 2016a). In the other six of the first 12 blocks (called *1-d blocks,* Fig. 1B), the two images forming a pattern again came from two distinct dimensions of images, but with four images in each. Each image in each dimension was consistently associated with another image in the other dimension (e.g. triangle always paired with green, square with yellow, etc.) such that these still combined to form four distinct visual input patterns. To ensure that participants paid attention to both dimensions of the stimulus, the left-right position of each part of the image was randomized across trials in all types of blocks (Fig. 2A,C). Participants were instructed that this had no effect on the correct choice to be learned (i.e. the correct action when the triangle was on the left and green on the right was the same as when their position was

reversed). In the first 12 blocks, the correct choice was different for each input pattern (Figure 1E). We matched each 2-d block with a 1-d block, such that their sequence of patterns and correct actions were identical, with only the actual visual pair of images corresponding to each pattern different between the blocks. This allowed us to directly control for sequence effects in the behavior. Blocks 1-12 each included 80 trials. The 1-d and 2-d blocks were randomly interleaved.

Block 13's first half was identical to a previous 2-d block (with new, non-overlapping patterns). In the second half (transfer phase; Fig. 5), two new features were introduced in one of the

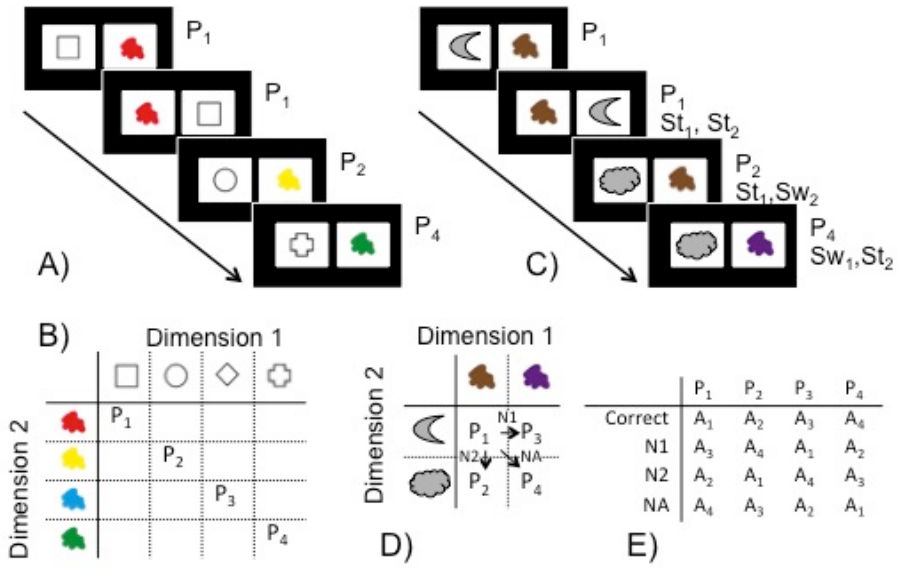

Figure 2: Experimental design. A) Four sample trials in a 1-d block: subjects needed to respond to each pattern Pi by pressing one of four keys (A_1-A_4). Each pattern was characterized as a pair of two images, and participants were instructed that the left-right position of each image did not matter. **B)** Example of stimuli presented in a 1-d block: while the stimuli included two input dimensions, they were effectively 1-dimensional, with no overlap between stimuli. **C)** Example of four trials in a 2-d block, yoked to the four trials in 1-d block in A). Each trial indicates change from the previous trial as whether this is a stay (*St*) or switch (*Sw*) trial relative to each dimension in the pair. **D)** Example of stimuli presented in a 2-d block, where the four patterns are 2-dimensional and present overlap. Arrows give an example of what different errors would be for pattern P_1: neglecting dimension 1 (N1) would correspond to mistakenly selecting the action prescribed for pattern P_3; neglecting dimension 2 (N2) the one for pattern P_2, and neglecting all (NA) the one for pattern P_4. E) For each pattern, the table contains the correct action, and the type of error corresponding to each possible mistake. The mapping from actions A1-A4 to finger presses was randomized in each block.

dimensions (called context), while the other dimension's features remained the same. Thus, subjects learned the correct action for four new patterns, for an added 80 trials. The new associations to be learned were such that in one of the new contexts, one of the previously learned associations between the other dimension and actions could be generalized if it had been abstracted as a sub-rule, while in the other context, this was not the case. This protocol follows our previously published structure learning protocol, testing for generalization of rules.

Reaction time switch-cost. 2-d blocks offer the possibility of learning structure, by using one dimension as a context that cues a lower-order rule, which itself constrains associations between features of the other dimension and actions. Assuming that subjects build structure, we infered which of the two dimensions is used as the "high" dimension, using reaction time (RT) switch-cost (SwC) (Collins & Frank 2013; Collins et al. 2014). We have previously validated this measure, showing that it was predictive of subjects' ability to transfer the structure to novel contexts that involve the same structure, and was also predictive of subjects attention to the high level context as decoded from neural signals (Collins et al. 2014; Collins & Frank 2013). Specifically, for each dimension $i=1:2$, we measured SwC_i as the difference between reaction times in Sw_i trials (where the feature on dimension i changed from trial $t-1$ to trial t) to RTs in St_i trials (where the feature on dimension i stayed the same). We then used the difference SwC_1-SwC_2 as evidence for structure using dimension 1 or 2 as context. Switch-costs were computed during the second half of each block (at asymptotic performance), on correct trials. To control for other sequential factors that may affect reaction times (such as change in side of feature position, change in action, and effect of specific motor sequences), we correct RTs in each 2-d blocks by subtracting RTs from the 1-d block matched to it.

Error analysis: In 2-d blocks, errors can be classified into three types (Fig. 2D,E): selecting the action correct for dimension 1's feature but incorrect for dimension 2's feature (neglecting 2, N2); selecting the action correct for dimension 2's feature but incorrect for dimension 1's feature (neglecting 1, N1), or selecting an action incorrect for both features (neglecting all NA). To compare to the behavior in 1-d blocks, we can map these errors to actions in 1-d blocks: for each 2-d block, we define the sequences of actions that would correspond to N1, N2 and NA errors in this block, and analyze how subjects' choices match with these kinds of error. Because each 1-d

block is yoked to a 2-d block, we can analyze errors in the 1-d blocks using the matching 2-d block action sequences; although these errors don't have the same interpretation in 1-d blocks as they do in 2-d bocks, this provides a baseline for measuring how much each type of errors would be expected based on a flat learning representation, and thus allows us to control for any other potential biases (such as the specific key press used for different errors).

Structure analysis: for 2-d blocks, we identify a high and low dimension such that if *SwC1>SwC2*, *H=1*, *L=2*, and conversely. This allows us to define SwH vs. StH trials; we also re-label error types as NH (neglecting the high dimension) and NL (neglecting the low dimension) based on structure identification.

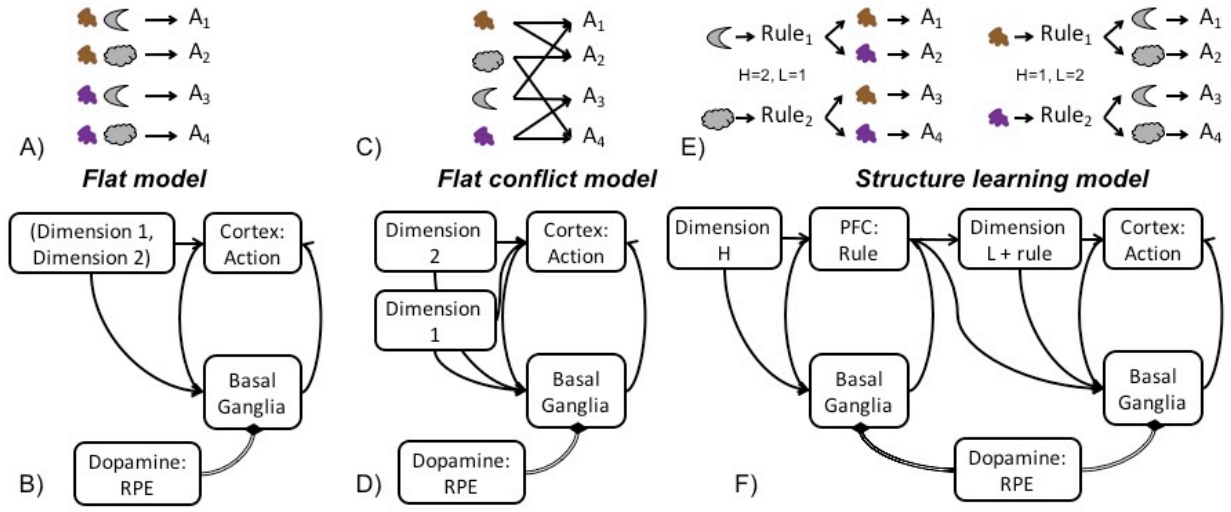

Figure 3: Neural network models capture different assumptions about how information is learned and represented. **A)** The flat model assumes that we learn associations between each pair of images and actions. **B)** The neural network model captures this by assuming that the pairs are represented as distinct, non-overlapping input patterns that serve as input for the cortico-basal-ganglia loop learning mechanism. All neural network models use reward prediction error (RPE)-like signal to learn to select actions in response to an input. **C)** The flat conflict model assumes that each feature learns associations to actions, such that this generates conflict between potential actions associated with a given feature. **D)** The neural network model captures this by entering each dimension's features as a separate input to the same learning loop. **E)** The structure learning model assumes that we learn at two hierarchical levels in parallel: to select rules in response to features of one dimension (the context or high dimension H), and then within this rule, to select actions in response to features of the other dimension (stimulus, or low dimension L). Note that in 2-d blocks, two different structure representations are possible, depending on which dimension serves as H vs. L. **F)** The neural network model captures this by assuming two parallel learning loops, where the lower level loop's input is constrained by the higher level loop's choice of rule.

Computational models: We use neural network simulations to provide predictions from three different models assuming different representations of learned information: a *flat model*, a *flat conflict model*, and a *structure learning model*. The structure learning model and flat conflict model are identical to models published previously; the flat model is nearly identical to the flat conflict model, with the one difference being that inputs vary only on one dimension (Fig. 2B). Briefly, all neural networks use biologically realistic activation and learning rules, and respect the known connectivity of cortico-basal ganglia loops(Hazy et al. 2006; Frank et al. 2007; Collins & Frank 2013). Models learn from reinforcement-like feedback implementing dopamine-dependent plasticity in cortico-striatal connections. Models were implemented and simulated in the emergent neural network software(Aisa et al. 2008), v7. The three models differ by 1) their inputs to the loops 2) the number of loops. The flat and flat conflict model contain only a single set of parallel loops, and differ in that the flat conflict model represents both dimensions of the pattern separately (allowing each dimension to learn association weights to the selected actions; Fig 3C,D), while the flat model represents each pattern separately (Fig 3A,B). The structure-learning model (Fig 3E,F) comprises of two hierarchically organized loops, whereby one dimension leads to selection of a rule in PFC, which then constrains inputs to the second loop, combining the rule context and the second input dimension. Rules initially do not carry any meaning, but come to represent a set of stimulus-action associations by the way they bias lower level action select. We previously showed that this network structure accounts for human structure learning in our structure learning task.

Model simulations: We simulated each model 100 times, and excluded from analyses models that did not reach asymptotic performance within 50 presentations of each pattern (30 in the structure learning model, 0 in both flat models). All simulations presented here used 2-d blocks' inputs (Figure 3), to make predictions about overall performance and reaction times assuming different representations of 2-d problems. Simulations of 1-d blocks inputs (not shown) are identical to simulations of 2-d blocks with the flat model. Indeed, 1) the flat model treats all 4 inputs as independent and thus assumes no difference between 1-d and 2-d; 2) the flat conflict model is identical to the flat model for 1-d, as there is no overlap between the 4 inputs; and 3) the structure learning model collapses to the lower level loop for 1-d inputs, as a single dimension is

sufficient to predict the data. Thus, flat simulations represent 1-d blocks predictions, as well as 2-d blocks predictions under a flat learning assumption; while flat conflict and structure model simulations represent predictions for the 2-d blocks under a conflict or structure learning hypothesis.

Although we plot learning curves for model simulations and participants' behavior, we focus our predictions on the qualitative differences in overall performance (proportion of correct trials and reaction times) across conditions, rather than on the temporal dynamics of the learning curve. This is for two reasons. First, the neural network model is highly parameterized; thus to avoid overfitting, we chose to use the previously published model parameters, instead of attempting to fine-tune model parameters for better quantitative fit of the learning dynamics. Second, we model here only one type of learning (model-free reinforcement learning), and do not include working memory, even though it is likely to contribute in parallel to explicitly remembering discrete associations, thus accelerating learning (Collins & Frank 2012). Thus, we focus on the qualitative predictions of the model that are unaffected by potential contributions of working memory, but not on quantitative dynamics that would be affected by other processes.

Results.

22 participants participated in a learning task where they used feedback to learn to select correct actions for four patterns. In 1-d blocks, the four patterns did not share any features, thus learning for one pattern could be assumed to be fully independent of learning for another pattern (flat learning). In 2-d blocks, the four patterns shared one of two features on one of two dimensions, potentially allowing structure learning (Fig. 2D, Fig. 3E). We first looked at the effect of overlap in pattern features on learning performance and reaction time, by comparing 2-d and 1-d learning blocks. We found a strong effect of block type on proportion of correct trials ($t(21)=7.24$; $p<0.001$) such that 21 out of 22 participants performed better in 1-d blocks (mean 87%; min-max [76-94]) than in 2-d blocks (mean 76% [51-87]). Similarly, we found a strong effect of block type on correct-trial reaction times ($t(21)=17.9$, $p<0.001$) such that all participants responded faster in 1-d blocks (mean 694ms [624 782]) than in 2-d blocks (mean 835ms [743 914]). To check whether this effect persists throughout the experiment, we examined the second half of the twelve blocks: the results remained identical if we only included the last three 2-d and 1-d

blocks. While these results confirm predictions from our structure learning model (Fig. 4, top), it could also be an unsurprising result of conflict between patterns in the 2-d blocks compared to 1-d blocks (Fig. 4, top, Fig. 3C); thus, further analysis is needed to confirm whether slower reaction times and weaker performance are due to structure learning, or to conflict in the 2-d blocks.

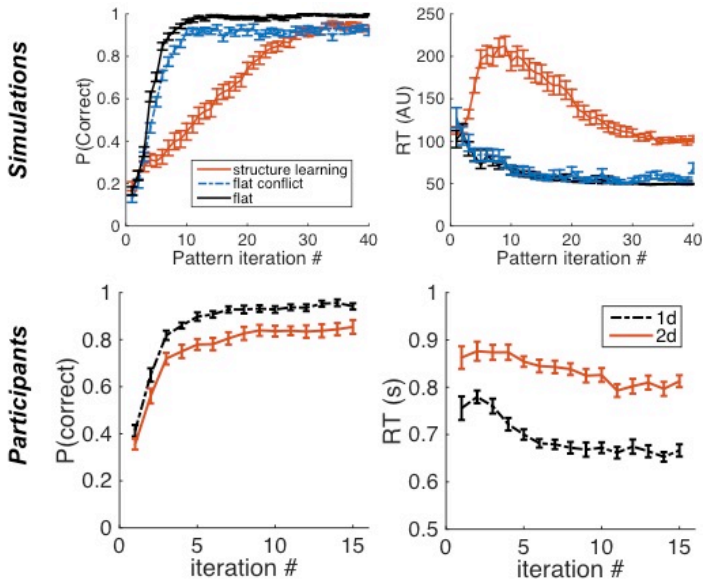

Figure 4: Performance is impaired in 2-d compared to 1-d. Top: network simulations predict worse overall performance (left) and slower reaction times (right) for structure learning over flat conflict, over flat model. **Bottom:** Participants' performance is worse and reaction times are slower for 2-d compared to 1-d. This matches the qualitative prediction that 2-d problems are not represented in a flat manner as 1-d problems, but either in a *flat conflict* or *structure* way. Error bars indicate standard error of the mean.

We next sought to establish that participants were indeed building structure. As a first confirmation, we investigated learning in the last block's transfer phase, where task structure allowed us to test whether participants built rule structure and generalized a previously learned rule in a new context (Fig. 5). Consistent with structure learning model predictions (Fig. 5 top right), we found that participants performed significantly better in the new context corresponding to an old rule (transfer condition) than they did in a new context corresponding to a new rule (new condition; Fig. 5 bottom left). Indeed, the proportion of correct responses over the first 5 trials of each new pattern in the transfer vs. new condition was significantly higher ($t(21)=4.5$, $p<0.001$; this was observed in 17 out of 22 participants – sign test $p=0.003$). This replicated our previous finding that participants build structure in absence of an incentive (at the beginning of block 13), and use this rule structure to later generalize known rules to new contexts. This result was not predicted by either the flat learning model or the flat conflict model (Fig. 5, bottom right; $t(99) = 0.9$, $p=0.38$).

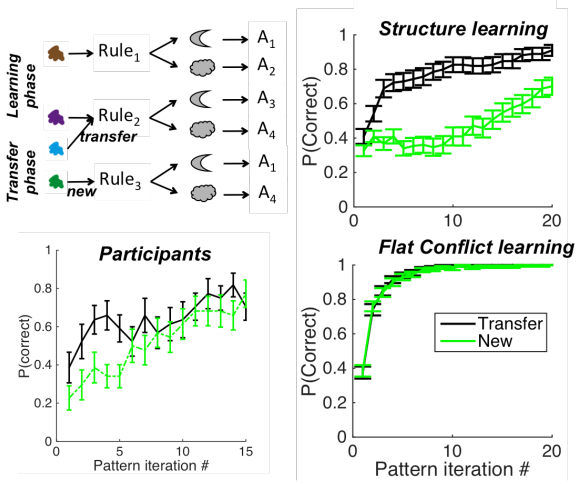

Figure 5: Transfer phase results support structure learning. **Top left:** in the last block, the 2-d learning block transitioned to learning of four new patterns with two shared features on dimension 2, and two new features on dimension 1 (here, color). One new feature afforded transfer of a previously learned rule (here blue), while the other one required learning a new rule (here green). **Bottom left:** participants' learning showed significantly better performance in the transfer than new condition. **Right:** network simulations. This qualitative pattern was predicted by the structure learning model (top), but not by the flat conflict model (bottom) or the flat model (not shown).

Next, we wanted to relate structure learning in 2-d blocks to decreased performance compared to 1-d blocks. To avoid inducing structure learning in participants during 2-d blocks by providing a generalization incentive, we only included a transfer phase in the last block; thus we couldn't use transfer in 2-d blocks as a marker of structure learning. We thus investigated whether structure learning could also be identified in the first 6 2-d blocks by other markers that we previously validated, and that are uniquely predicted by the structure learning model. Specifically, the structure learning model predicts that 1) participants should exhibit task-set (rule) switching reaction time switch cost when successive trials change in the "high level" dimension of their created structure (Fig. 6A) 2) that participants should show more errors neglecting the low dimension of their created structure (NL) than they do the high dimension (NH), and more so than other kinds of errors (NA; Fig 6A). This can be rephrased as saying that the structure learning predicts asymmetry in the role of the two pattern dimensions, both for reaction time switch costs and error types, and that this should be consistent across those measures as markers of the specific structure created. Conversely, the flat model learns for each pattern independently, and thus predicts no asymmetry in errors or switch-costs (Fig. 6C). The flat conflict model also predicts no switch-costs, though it does predicts more errors that correspond to neglecting one of the two input dimensions, but with no asymmetry between those dimensions (Fig. 6B).

Thus, we asked whether asymmetry in reaction time switch-cost *SwC1-SwC2* (with 2-d blocks reaction times corrected by 1-d blocks reaction times to account for other possible sequential effects – see methods) were predictive of asymmetry in the distribution of errors neglecting either input dimension, N2-N1. We measured this separately in each 2-d block of each participant, and found a strong relationship (Spearman $\rho(131)=0.46$, $p<10\text{-}6$; Fig 6D left), showing that the measure of error asymmetry was consistent with the measure of RT switch cost

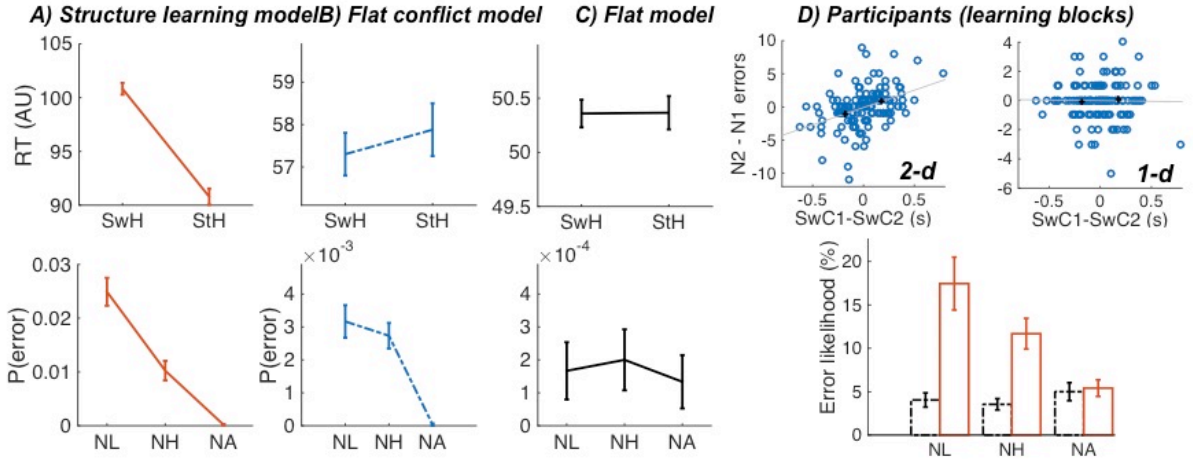

Figure 6: Model simulations show that only the structure learning model **(A, top)** predicts the existence of a reaction-time switch-cost when a pattern's high dimension feature changes from one trial to the next (SwH vs. StH). Simulations show that the flat model **(C, bottom)** predicts no difference in the kinds of error made. In contrast, the flat conflict model **(B, bottom)** predicts more errors neglecting a single dimension (high or low: NH or NL) than errors neglecting both dimensions (NA). The structure learning model predicts significantly more errors neglecting the low dimension than the high dimension, and more so than NA. **D)** Empirical results. **Top:** Asymmetry in individual learning blocks' switch-cost (SwC) between dimension 1 and dimension 2 in 2-d blocks was significantly related to their asymmetry in error distributions. This was not the case in 1-d blocks (right). **Bottom:** Error distribution in 2-d (full red) and 1-d (dotted black) blocks. Participants make as many NA errors both types of blocks, likely reflecting action selection noise. Thus, the increase in errors neglecting the high dimension, and the comparatively greater increase in errors neglecting the low dimension, reflect a decrease in performance in 2-d blocks due to structure learning. Error bars indicate standard error of the mean.

asymmetry, in such a way that participants made more errors neglecting dimension 2 (N2) if dimension 1 showed higher reaction times switch-cost (SwC1>SwC2), and reciprocally. Indeed, if we assigned each block in a binary fashion to a "dimension 1" or "dimension 2" structure, we found that the first group had significantly more N2 than N1 errors ($t(66)=2.6$, $p=0.01$), and the second group significantly more N1 than N2 ($t(64)=2.9$, $p=0.005$). As expected from the correlational result, the two groups differed significantly in their proportion of N2 vs. N1 errors

(t=3.9, p=0.0001). To ensure that the link between RT switch-cost and error distribution was not due other factors than structure, we checked whether the relationship was also present in 1-d blocks. We found that switch cost asymmetry was not related to error distribution in 1-d blocks (Spearman ρ=0.01, Fig 6D right), and that the link between switch cost asymmetry and error distribution was significantly stronger in 2-d than 1-d blocks (Spearman ρ=0.38, $p<10^{-5}$). Using RT switch-cost asymmetry, we assigned each 2-d block to the most likely structure subjects built in this block, and label errors as NH, NL or NA. The results remained identical if we only included the last three 2-d and 1-d blocks.

The results above replicate our previously published finding (Collins & Frank 2013; Collins et al. 2014) showing that participants build structure a priori, even when unneeded, and extend them to showing that participants do so repeatedly, even when exposed multiple times to the evidence that they do not gain from it (by having generalization opportunities). Finally, we investigated whether in addition to being unneeded, structure learning was also costly. Specifically, we asked whether the decreased performance in 2-d blocks was specifically due to structure learning, or if it could only be due to conflict making learning overall more difficult. To do so, we compared the distribution of error types in 1-d and 2-d blocks, by looking at the average observed probability of each type of error per subject (Fig. 6D bottom). An anova with error type and block type as factors confirmed both main effects on error probability (F's>4, p's<0.02), as well as an interaction between the two factors (F=5.6, p=0.004). We confirmed with an anova within 1-d blocks that this was due to the fact that there was no effect of error type within 1-d blocks (F=0.56, ns). Furthermore, post-hoc tests showed that, while NH and NL errors were significantly more likely in 2-d than 1-d blocks (t's(21)>4.1, p's<0.001), there was no difference in the probability of NA errors between 1-d and 2-d blocks (t(21)=0.42, ns, Fig. 6D bottom). This indicates that the decrease in performance from 1-d to 2-d blocks was not due to an overall increase in choice randomness, but specifically in NH and NL errors, and more so in NL errors. This supports the argument that learning errors in 2-d blocks are partly accounted for by using structure. The results remained identical if we only included the last three 2-d and 1-d blocks.

Discussion

The findings in this study replicate and extend our previous findings on structure learning. First, we replicated our finding that participants create hierarchical rule structure when learning from reinforcement, even when the learning problem offers no incentive to do so, and no immediate benefits can be derived from creating structure. We observed three independent, but correlated characteristic behavioral signs that we had previously identified as signatures of structure learning. First, in a test phase at the very end of the experiment, participants were able to generalize a previously learned rule in a new context, allowing them to learn faster in this new context, compared to a context where they needed to learn a new rule. Second, participants exhibited reaction time switch-cost when the trial pattern changed from the previous trial along one input dimension, more than when it changed along the other. We identify this asymmetric use of information for action selection as an indicator of hierarchical rule structure: in a context where both dimensions theoretically play the same role, one dimension serves as high level context that cues rules, while the other serves as a low level stimulus that determines actions to select within a rule. Third, the errors made by participants showed an asymmetry in which part of the visual pattern was neglected. This asymmetry was consistent with the switch-cost asymmetry, with the low dimension neglected more than the high one. This indicates that subjects are more likely to appropriately select a rule and then fail to apply it correctly, than to select the wrong rule and then apply it well.

In addition to replicating our previously published findings, we extend them in two ways. First, we show that structure learning is indeed costly. Comparing performance in blocks where participants could learn structure (2-d) to blocks in which they could not (1-d), we found that participants consistently made more errors and responded slower when learning structure. Looking at the specific nature of the errors confirms that the increase in errors could not be explained by an overall increase in randomness in choice selection. Indeed, there were an equal number of purely "noisy" errors (NA, where the chosen action did not match the action prescribed by any component of the visual pattern) in 2-d and 1-d blocks. Instead, the decrease in performance was partially caused by the increase in specific types of errors predicted by a structure learning model. While the structure learning model captures the qualitative patterns of participants' behavior (including transfer, reaction-time switch-cost, and distribution of error types) in a way that neither the flat nor the flat conflict model can, the quantitative fit of the

model is weak (slower overall learning, etc.). There are several reasons for this; the most salient is that the model simulations used previously published parameters, rather than using parameters optimized for quantitative fit to behavior. This avoids overfitting in a neural network model where a large number of parameters could be tweaked. Another likely reason for observed differences between the model and human behavior is that we do not account for the use of working memory for learning (Collins & Frank 2012), in parallel to the reinforcement learning process modeled here; this could explain why our model learns slower, or sometimes doesn't converge. These limitations are not important for our results, since we use modeling here to show that the qualitative patterns of behavior can only be accounted for by structure learning, not to make precise quantitative predictions.

Another important finding of this paper is that structure learning still occurred after participants experienced six blocks where it was costly and non beneficial. This was shown by observed transfer at the end of the experiment, and by the continued presence of all markers of structure learning in the second half of the experiment—the last three 1-d and 2-d blocks. Thus, participants did not seem to dynamically adjust their prior for structure learning over time based on costs and benefits, as might be predicted by rational resource allocation theory. It is thus remarkable that our tendency to build structure is so strong that it resists evidence for a lack of benefits and the substantial costs in building structure.

It is possible that the strength of this bias simply reflects the statistics of our interactions with our environment: assuming that learning structure is overwhelmingly useful in everyday life, we could expect a prior that structure is helpful to require more contrary evidence to override than provided by an hour long experiment. An alternative though not necessarily conflicting explanation is that we may have evolved a hardwired network to learn structure by default, which would be more efficient than arbitrating between structure and flat learning if indeed structure learning is overwhelmingly useful. There is some support for the latter hypothesis. First, we observe structure learning very early on in life, prior to having much exposure to the world(Werchan et al. 2015; Werchan et al. 2016). Second, the cortico-basal-ganglia loops that enable simple reinforcement learning are also used, with different inputs from the hierarchically organized subregions of lateral PFC(Koechlin et al. 2003; Koechlin & Summerfield 2007; Badre

2008), to afford hierarchical structure learning(Alexander & DeLong 1986; Haber & Behrens 2014; Collins & Frank 2013; Collins & Frank 2016a). Last, rostral regions of PFC, linked to higher level, more abstract processing, are engaged by default in early learning(Badre & Frank 2011; Frank & Badre 2011). Jointly, these findings may support the hypothesis that PFC-basal-ganglia loop networks perform a priori structure learning.

Our experimental design enabled us to study whether the behavioral cost for 2-d blocks may be due to choice conflict rather than structure learning. We tested the cost of structure learning by comparing learning in 2-d blocks to learning in 1-d blocks — 1-d blocks cannot afford any hierarchical structure and thus form a good baseline for evaluating flat learning. Furthermore, structure learning requires inputs to be at least two dimensional, to allow participants to use one dimension as a context cuing rules. However, this choice implies that 2-d blocks differ from 1-d blocks not only in that they provide an opportunity for structure learning, but also in two other ways. First, 2-d blocks require participants to integrate two dimensions, while they could perform 1-d blocks by focusing on only one dimension. In an attempt to mitigate this difference, the positions of the two dimensions on the screen were unpredictable. Thus, if participants focused on only one dimension, a prediction would be that a switch in the side of this dimension (compared to a stay) from the previous trial would result in a greater change in reaction times than in 2-d trials, where participants always needed to consider both dimensions. Instead, we observed no difference ($t=0.11$, $p=0.9$) between the two conditions in participants, supporting the hypothesis that participants do pay attention to both dimensions in 1-d blocks. Second, in 2-d blocks, different patterns have overlapping features, which generates choice conflict (Zhang & Kornblum 1998; Monsell 2003). As conflict is behaviorally costly, leading to errors and slower reaction times, it is important to ensure that conflict is not the only source of the observed performance cost in our results. Our computational simulation of a model that accounts for conflict does predict a drop in performance. However, it does not account for the observed asymmetrical pattern of errors, for the reaction-time switch-cost, or for the link between these two independent measures; nor would a model that assumed that participants did not integrate dimensions in 1-d blocks. Thus, our results support our argument that, while conflict may cause some of the observed performance difference between 1-d and 2-d blocks, part of it must be attributed to structure learning.

Another question one might ask is whether the observed asymmetry in the role of the two pattern dimensions in 2-d blocks, which is consistent between RT switch-cost and error distribution, might arise from differential attention to the two dimensions, for example because one was preferred or more salient. Higher attention to dimension one should lead to faster switching in that dimension and to fewer errors neglecting it; thus it would predict a negative correlation between SwC1-SwC2 and N2-N1. Instead, we observe the opposite relationship, as predicted by our structure learning model. Thus, while attention may play a role in which dimension is selected as "context" in structure building, structure learning is a better explanation of the observed error and switch-cost patterns.

We have shown that creating structure when learning is costly, but that this cost does not prevent us from learning structure a priori, even when repeatedly experiencing its lack of benefits. While this may seem irrational, it may actually be an effective long-term strategy in an environment where structure usually comes in useful for generalizing knowledge. However, this highlights a remarkable ability to pay short term costs for long-term optimality. We have focused on a specific form of prefrontal-cortex-dependent structure learning, the learning of simple sets of stimulus-action-outcome associations into rules that can themselves be selected as an abstract action in different contexts. Other related forms of hierarchical reinforcement learning focus on other ways to learn to select strategies that are more abstract than simple actions, such as options (which represent policies to select sequences of actions in environments where actions condition next states)(Botvinick et al. 2009; Botvinick 2008; Ribas-Fernandes et al. 2011). Theoretical work has shown that adding options to a reinforcement learning problem may also incur a cost, as they increase the possible number of decisions at a given point, but lead to more efficient long term learning(Botvinick et al. 2009). It would be interesting to see if the pattern we observe here holds in this type of hierarchical structure learning: do humans create options a priori, even when they're not useful? Does this incur a cost? Does this cost have an effect on the tendency to create options? These questions are important for future research, to further our understanding of how organizing information during learning, even at short-term cost, helps humans learn more quickly and flexibly.

Bibliography

Aisa, B., Mingus, B. & O'Reilly, R., 2008. The emergent neural modeling system. *Neural networks : the official journal of the International Neural Network Society*, 21(8), pp.1146–52.

Alexander, G. & DeLong, M., 1986. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. *Annual review of neuroscience*.

Badre, D., 2008. Cognitive control, hierarchy, and the rostro-caudal organization of the frontal lobes. *Trends in cognitive sciences*, 12(5), pp.193–200.

Badre, D. & Frank, M.J., 2011. Mechanisms of Hierarchical Reinforcement Learning in Cortico-Striatal Circuits 2: Evidence from fMRI. *Cerebral cortex (New York, N.Y. : 1991)*, pp.1–10.

Badre, D., Kayser, A.S. & Esposito, M.D., 2010. Article Frontal Cortex and the Discovery of Abstract Action Rules. *Neuron*, 66(2), pp.315–326.

Bengio, Y., Courville, A. & Vincent, P., 2012. Representation Learning: A Review and New Perspectives.

Botvinick, M.M., 2008. Hierarchical models of behavior and prefrontal function. *Trends in cognitive sciences*, 12(5), pp.201–8.

Botvinick, M.M., Niv, Y. & Barto, A.C., 2009. Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition*, 113(3), pp.262–80.

Collins, A. & Koechlin, E., 2012. Reasoning, Learning, and Creativity: Frontal Lobe Function and Human Decision-Making J. P. O'Doherty, ed. *PLoS Biology*, 10(3), p.e1001293.

Collins, A.G.E. & Frank, M.J., 2013. Cognitive control over learning: Creating, clustering, and generalizing task-set structure. *Psychological Review*, 120(1), pp.190–229.

Collins, A.G.E. & Frank, M.J., 2012. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *The European journal of neuroscience*, 35(7), pp.1024–35.

Collins, A.G.E. & Frank, M.J., 2016a. Motor Demands Constrain Cognitive Rule Structures J. Daunizeau, ed. *PLOS Computational Biology*, 12(3), p.e1004785.

Collins, A.G.E. & Frank, M.J., 2016b. Neural signature of hierarchically structured expectations predicts clustering and transfer of rule sets in reinforcement learning. *Cognition*, 152, pp.160–169.

Collins, a. G.E., Cavanagh, J.F. & Frank, M.J., 2014. Human EEG Uncovers Latent Generalizable Rule Structure during Learning. *Journal of Neuroscience*, 34(13), pp.4677–4685.

Donoso, M., Collins, A.G.E. & Koechlin, E., 2014. Foundations of human reasoning in the prefrontal cortex. *Science*, p.science.1252254-.

Frank, M.J. et al., 2007. Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. *Science (New York, N.Y.)*, 318(5854), pp.1309–12.

Frank, M.J. & Badre, D., 2011. Mechanisms of Hierarchical Reinforcement Learning in Corticostriatal Circuits 1: Computational Analysis. *Cerebral cortex (New York, N.Y. : 1991)*, (2010), pp.1–18.

Haber, S.N. & Behrens, T.E.J., 2014. The Neural Network Underlying Incentive-Based Learning: Implications for Interpreting Circuit Disruptions in Psychiatric Disorders. *Neuron*, 83(5), pp.1019–1039.

Hazy, T.E., Frank, M.J. & O'Reilly, R.C., 2006. Banishing the homunculus: making working memory work. *Neuroscience*, 139(1), pp.105–18.

Koechlin, E., Ody, C. & Kouneiher, F., 2003. The architecture of cognitive control in the human prefrontal cortex. *Science (New York, N.Y.)*, 302(5648), pp.1181–5.
Koechlin, E. & Summerfield, C., 2007. An information theoretical approach to prefrontal executive function. *Trends in cognitive sciences*, 11(6), pp.229–35.
Kool, W. et al., 2013. Neural and Behavioral Evidence for an Intrinsic Cost of Self-Control S. F. Brosnan, ed. *PLoS ONE*, 8(8), p.e72626.
Kool, W. & Botvinick, M., 2014. A labor/leisure tradeoff in cognitive control. *Journal of Experimental Psychology: General*, 143(1), pp.131–141.
Mesnil, G. et al., 2011. Unsupervised and Transfer Learning Challenge: a Deep Learning Approach. , 7, pp.1–15.
Monsell, S., 2003. Task switching. *Trends in Cognitive Sciences*, 7(3), pp.134–140.
Niv, Y. et al., 2015. Reinforcement learning in multidimensional environments relies on attention mechanisms. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 35(21), pp.8145–57.
Parisotto, E., Ba, J.L. & Salakhutdinov, R., 2015. Actor-Mimic: Deep Multitask and Transfer Reinforcement Learning.
Ribas-Fernandes, J.J.F. et al., 2011. A neural signature of hierarchical reinforcement learning. *Neuron*, 71(2), pp.370–9.
Werchan, D.M. et al., 2015. 8-Month-Old Infants Spontaneously Learn and Generalize Hierarchical Rules. *Psychological science*.
Werchan, D.M. et al., 2016. Role of Prefrontal Cortex in Learning and Generalizing Hierarchical Rules in 8-Month-Old Infants. *Journal of Neuroscience*, 36(40), pp.10314–10322.
Westbrook, A. et al., 2013. What Is the Subjective Cost of Cognitive Effort? Load, Trait, and Aging Effects Revealed by Economic Preference M. Pessiglione, ed. *PLoS ONE*, 8(7), p.e68210.
Westbrook, A. & Braver, T.S., 2015. Cognitive effort: A neuroeconomic approach. *Cognitive, Affective, & Behavioral Neuroscience*, 15(2), pp.395–415.
Wilson, R.C. & Niv, Y., 2011. Inferring relevance in a changing world. *Frontiers in human neuroscience*, 5(January), p.189.
Yu, A.J. & Cohen, J.D., 2009. Sequential effects: Superstition or rational behavior. *Advances in neural information processing systems*, 21, pp.1873–1880.
Zhang, H. & Kornblum, S., 1998. The effects of stimulus–response mapping and irrelevant stimulus–response and stimulus–stimulus overlap in four-choice Stroop tasks with single-carrier stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, 24(1), pp.3–19.