Higher perceptual attention cost slows contingency learning after a modality shift

Natalia Postnova (natalia.postnova@helsinki.fi)
Faculty of Educational Sciences, University of Helsinki, Helsinki, 00014, Finland

Nelson Totah (nelson.totah@helsinki.fi)
Helsinki Institute of Life Science HiLIFE, University of Helsinki
Faculty of Pharmacy, University of Helsinki

Benjamin Ultan Cowley (ben.cowley@helsinki.fi)
Cognitive Science, Faculty of Arts, University of Helsinki
Faculty of Educational Sciences, University of Helsinki

Abstract

Learning perception-action contingencies from the environment depends on attention, to efficiently control cognition and selectively process sensory information. Shifting attention between sensory modalities incurs a cost in humans, other primates, and rodents, resulting in slower learning in the new modality. Previous set-shifting work in rats showed that increased difficulty of perceptual discrimination in the preceding modality increases the following shift cost. We studied this in humans by manipulating perceptual attention in one sensory modality, titrating task demand by staircase design, to test the effect on perceptual contingency learning in another modality. To accommodate the complexity of human learning, we introduce a Bayesian method to decompose and estimate learning characteristics from performance data. This method identifies the completion of rule acquisition and consolidation, accounting for individual variation in learning. Results show the expected modality shift cost, and offer new evidence in humans that shift cost is exacerbated by prior demands on attention.

Keywords: perceptual learning, computational cognition, attention, set-shifting task

Introduction

Human learning is a complex process that depends on cognitive faculties including attention, perception, and memory, all interacting as we acquire and consolidate new information. Humans learn *perceptual contingencies*, such as how to safely navigate traffic lights, using these faculties to do associative and then inference-based learning.

Efficient learning, that is, learning under time or resource constraints, depends critically on paying attention. This view conceives of attention as the flexible control of limited computational resources (Lindsay, 2020), with the emphasis on control. A slightly more general view from predictive processing literature, however, sees attention as “context or state-dependent optimization of the precision of prediction errors” (H. Brown et al., 2011). This allows that attention operate not only at the level of cognitive control, but also as a perceptual filter, enhancing discrimination of intra-modal stimuli near the perceptual threshold (Dermody et al., 2024).

The juxtaposition of both modes of attention when learning contingencies highlights their interaction. As we navigate our sensory environment, attention acts as a filter, allowing us to focus on specific aspects of our environment while inhibiting others by altering sensory neuron representations (Desimone & Duncan, 1995; Luck et al., 1997). This filtering function is dynamic and must respond to varying demand levels (e.g. distraction, occlusion, low discriminability) as well as manage competition and shift between targets or sensory modalities. This raises the research question we study here:

How does attentional resource allocation influence contingency learning in humans under conditions of high vs low perceptual discriminability of task stimuli?

We used an attentional set-shifting task to test the hypothesis that *increasing perceptual focus on a given modality (visual/auditory) will tax attention and delay perceptual contingency learning after a modality switch*. The attentional set-shifting paradigm is widely used to investigate how attention relates to learning of stimulus-response contingencies, particularly when shifting attention between different stimulus dimensions (Dias et al., 1996; Leong et al., 2017; Owen et al., 1991). The key feature of this paradigm is the introduction of novel stimuli and stimulus-response rules, requiring participants to shift their attention from one stimulus dimension to another. Such shifts of attention demand additional processing and manifest as a performance decrement, commonly termed the *shift cost*.

While the phenomena of modality- or task-driven shift cost is well established, here we leverage the expected shift cost to test the role of perceptual attention demand on *later* learning of a new stimulus-response rule. That is, the study tests whether being more heavily taxed in attention during one session will cause participants to be slower to acquire and consolidate the modality-shifted rule in a subsequent session.

The current study builds on Vasilev et al. (2022), who demonstrated the effect of attention manipulation on subsequent shift costs in rats. We adapted their GO/NOGO task for human participants, with functionally identical study designs for cross-species comparison. Compared to rats, humans require significantly fewer trials to learn new stimuli, and may exhibit sudden substantial behavior shifts due to inference. Thus, we also present a novel learning-point detection approach, to handle our data where the number of trials is limited and behavioral changes can occur rapidly.

In order to better operationalise the potential complexity of human learning, we do not treat it as linear but apply the classic three stage model of acquisition, consolidation, and maintenance (Fitts & Posner, 1967; Karni & Sagi, 1993). Acquisition occurs when individuals first encounter and acquire new contingencies, as exposure, response, and feedback cycles lead to initial neuroplastic connections and memory for-

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

mation. During this stage, performance fluctuates and the success rate at this stage reflects the process of trial and error. Consolidation is an intermediate stage, during which the newly acquired information becomes more stable and is integrated into existing knowledge structures. Performance becomes more consistent and probability of success increases. Maintenance involves longer-term retention and reinforcement of prior learning.

The results validate the experimental translation: we observed modality shift costs and, consistent with the rat study, found increased costs following tasks with high attentional demand. Thus, shift cost is not simply due to reward-guided attention. While perceptual attention is rarely central in human set-shifting research, our findings highlight the strong role it plays, that goes beyond dimensional salience and implicates the active information sampling (Esber & Haselgrove, 2011).

Method

Participants

A total of 55 participants were recruited to the study. Six were excluded due to data loss caused by technical errors. Our two core inclusion criteria were that participants should maintain adequate performance throughout the experiment and that the attention manipulation should sufficiently impact performance. These resulted in exclusion of 10 participants. In addition, 13 participants with musical education were recorded and are excluded from this report, as the effect of music training is not in scope here. The reported sample thus consists of 26 participants (15 female, 10 male, 1 other; age M=27.9, SD=4.5).

The study was conducted according to the guidelines of the Declaration of Helsinki, and approved by The University of Helsinki Ethical Review Board in Humanities and Social and Behavioural Sciences, 27/2022.

Apparatus

The experiment was conducted in an experimental room with controlled lighting and sound conditions. Two lights (LEDGO LG-600CSCII) were used to provide sufficient and consistent lighting throughout data collection. The Tobii Pro Spectrum 600 eye-tracking system and Tobii Pro Lab software were employed to monitor gaze location during the task; that data is not analysed herein. The stimuli were presented to the participants using Tobii monitor (EIZO FlexScan EV2451, diagonal=60.5 cm) and Sony MDR-7508 headphones. Tobii Pro Lab, and PsychoPy v2021.2.3 for task presentation, both ran on a Lenovo ThinkStation P350 i9 32GB with Nvidia T1000 8GB graphics, Windows 10 Pro.

Procedure

Participants were provided with consent forms and background questionnaires. After completing the questionnaires, participants were introduced to the experimental task, and eye-tracking calibration procedures were performed.

The task required participants to identify the GO stimulus through trial and error. For each correct response, indicated by green feedback, they earned one point, which was later converted into a reward, as explained before the experiment. Participants could track their cumulative points on the screen between experimental blocks. The point system was implemented to sustain motivation and encourage continued effort, even after the GO stimulus was identified.

The experiment, illustrated in Figure 1, consisted of nine sessions of at least 60 trials each: one training and eight test sessions. After training, we orally confirmed with participants that they understood the task and could identify the stimulus contingencies. The experimenter described any necessary clarifications, and if needed, training was reiterated. All participants completed the training successfully. Before both staircase sessions, compulsory breaks were implemented, to prevent fatigue and ensure best possible performance during the staircase.

The experimenter followed the progression of the experiment on a second monitor, where Tobii Pro Lab overlaid gaze location on the task, in order to verify that participants fixated on stimuli. This was particularly crucial during auditory modality sessions, as some participants naturally tended to let their gaze drift away from irrelevant visual stimuli.

The task took approximately 40 minutes to complete, with another 20 minutes for paperwork.

Stimuli

Within each session of the experiment (described below), stimuli comprised 2 auditory stimuli (pure tones) and 2 visual stimuli (sinusoidal grating), see Figure 1, panel B and C. In each trial, one auditory and one visual stimulus were presented simultaneously in one of four possible random combinations. At each experiment session a specific modality was designated as relevant. In this modality, one stimulus served as the 'GO' stimulus, requiring a key press. The other stimulus in the relevant modality was the 'NOGO' stimulus, requiring no action. The presentation of the stimuli lasted for 1 second. For the response to be registered, the key should be pressed within the second the stimuli are on. Positive feedback (green light) was given for the key press to GO stimuli, and negative feedback (red light and a longer inter-trial interval) was provided if a key was pressed for NOGO stimuli. No feedback was given when the key was not pressed. The second modality was considered irrelevant and had no effect on participants' performance score.

The stimuli in each modality were easily discriminable. Visual stimuli varied in grating orientation by 30 degrees, while auditory stimuli differed in frequency by an amount roughly equivalent to one tone (30/32 of a tone).

Throughout the experiment, five distinct stimulus sets were used. Each set contained stimuli that were sufficiently different from those in other sets to prevent interference from previously learned stimuli. Despite the limited number of stimulus combinations per session (only four, as in the original rat study), the task was challenging because it was speeded and the abstract stimuli resist memorisation by naming.

Experiment structure

The experiment started with a training session to ensure that the participants understood the task rules. Thereafter, the experiment was divided into sessions of four types, repeating (see Figure 1 panel A).

Inter-Dimentional Shift (IDS): Following training, the first IDS occurred, replacing stimuli with a different set of 2 auditory and 2 visual stimuli. However, the relevant modality for the GO and NOGO stimuli remained consistent with the previous session modality (training for first IDS).

Staircase Method: After the IDS session, the staircase procedure was conducted. This phase began with the same set of stimuli used in the IDS session, but the NOGO stimuli were adjusted using a 2-down/1-up staircase method. As the NOGO stimulus became more similar to the GO stimulus, this adjustment aimed to determine the appropriate level of difficulty for each participant, corresponding to approximately 70% performance accuracy. The stimuli in irrelevant modality stayed the same.

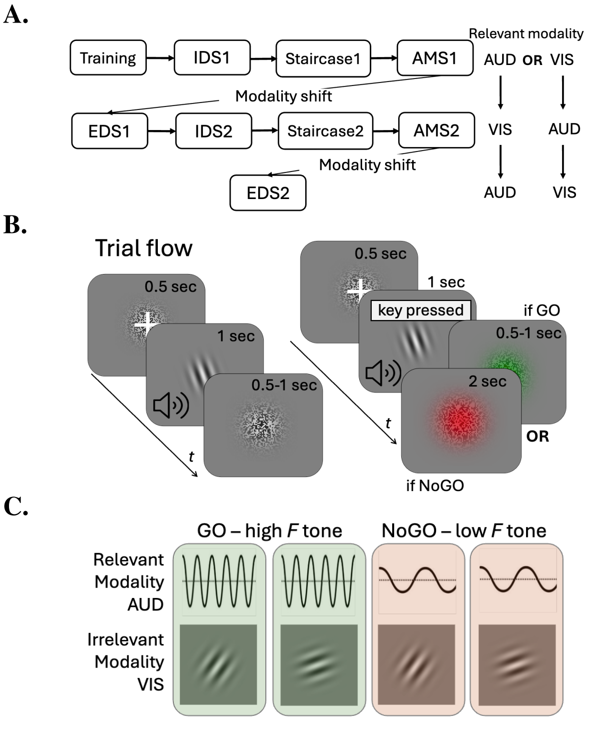

Figure 1: Experiment design at structure, trial, and stimulus levels. **Panel A**: Experimental structure: AUD (auditory) and VIS (visual) indicate the relevant modality for corresponding sessions. Participants were randomly assigned to begin with either the AUD or VIS modality. **Panel B**: trial schematic (visual elements not to scale), with structure: fixation cross, target stimulus, feedback. Three feedback possibilities are shown: no key press, no feedback (left); key press, green light if GO, red light of longer duration if NOGO. **Panel C**: example stimuli under the rule *GO to auditory high 'F' tone*.

Attention Manipulation Sessions (AMS): The AMS session consisted of two blocks (120 trials each) using the same discrimination task. The task was either easy, using the same stimuli as in the IDS, or difficult, with the GO stimulus unchanged and the NOGO stimulus adjusted based on the staircase results. The stimuli in the irrelevant modality remained the same as in the IDS.

Extra-Dimensional Shift (EDS): After AMS, an EDS occurred, introducing a new set of stimuli along with a change of the relevant modality.

Session order: Following EDS, the entire cycle of IDS - staircase - AMS - EDS was repeated one more time, ensuring that all participants completed both easy and challenging attention manipulation tasks.

The experiment sessions initially consisted of 60 trials each. Progression to the next session would occur if the windowed average performance (window=20 trials) reached 80% accuracy. If not, another block of 60 trials, with the same parameters, was added to the session. After completing each block, participants had to press space to continue, giving them a self-paced but brief pause for comfort.

Performance control: The efficacy of the AMS depended on the staircase converging to suitably hard-to-discriminate stimuli. To ensure this, the procedure would be repeated if after staircase procedure, the achieved difficulty of the stimuli resulted in too low performance ($<55\%$) suggesting guessing, or too high ($>75\%$) indicating insufficient discrimination difficulty. Using this method to control performance for only a subset of participants, we could establish the minimum value of performance impact due to the attention manipulation, and thus exclude (as described above) any of the other participants for whom the manipulation failed.

Impact of performance was measured by the Euclidean distance between the 2D points defined by normalised error rate and response time for sessions following the AMS (see Figure 3 panel B). In this way we accommodate performance impact expressed as speed and/or accuracy (rather than accuracy alone). The minimum value found among the controlled participants was about 30% of the normalised maximum of all participants' performance.

Estimation of learning point & shift cost

All modelling and analyses used Python 3 and common analysis libraries, as well as pymc, statsmodels, and pingouin.

To test our hypothesis that inducing an attention 'cost' (i.e. taxing perceptual attention in one modality) would influence a subsequent attention 'shift' (i.e. contingency learning in the other modality), we computed shift costs as follows. We define and estimate a 'learning point' (see below and Figure 2), as the trial number at which a specific degree of task proficiency is attained. Estimating a learning point for each session, we calculate shift costs by subtracting the learning *baseline* (learning point at the first IDS session) from the learning point achieved in sessions following a modality shift.

Learning point detection

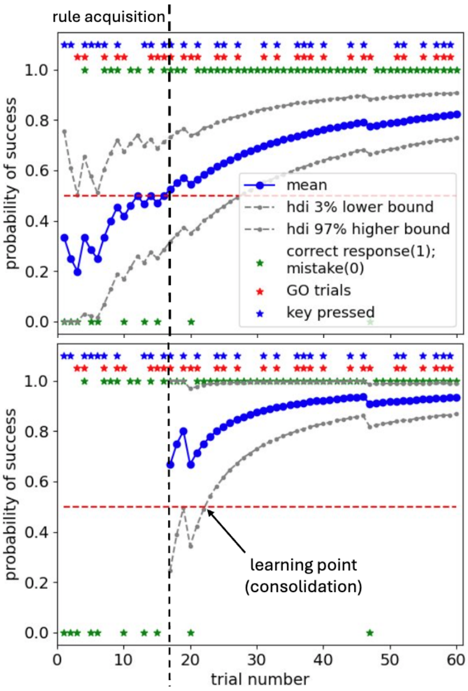

Figure 2: Example data showing two-stage fitted Bayesian framework model for (top) the acquisition stage and (bottom) the consolidation stage. Probability of a successful response is on the y-axis; GO trials are red stars and responses are blue stars; response outcomes are shown as green stars, correct=1 and error=0. Trials are on the x-axis; dashed line shows the estimated trial for rule acquisition, when model mean probability ≥ 0.5 in the top plot. The bottom plot shows the estimated learning point when HDI 3% lower bound ≥ 0.5

Trial and error learning of implicit information, such as a sorting rule, has been modelled in several ways, including sigmoid curve fitting and Bayesian hypothesis testing (Tenenbaum & Griffiths, 2001). While the original rat study used a sigmoid approach (Vasilev et al., 2022), the method proved unsuited to our data. The Sigmoid function, or S-Curve, is sensitive to individual errors when the number of trials is small, and cannot capture abrupt shifts in behavior.

Also, in contrast to the related card sorting tasks (Lange et al., 2016), in our task learning is partly probabilistic rather than purely rational. This is because the GO and NOGO stimuli may be close to perceptual discrimination threshold and the task is speeded rather than self-paced, requiring an automatic rather than deliberative response.

Thus, our approach combined Bayesian evidence updating with the conceptual three-stage model of learning mentioned in the Introduction: rule acquisition, consolidation, and maintenance. This allowed us to model learning from fewer trials and also account for progressive changes in humans' information updating.

During the acquisition stage, when the target remains ambiguous, the learning process is characterized by trial and error, wherein various options are explored randomly. Until a sufficient amount of feedback is gathered, these trials can be regarded as independent events (every trial a guess), and their outcomes modeled using a binomial distribution. The Bayesian framework dynamically calculates the evolution of the distribution of outcomes following each trial.

The calculation begins with the first trial that includes feedback. Since feedback was provided only for trials where an action (key press) was taken, any initial trials following a stimulus change in which the participant did not act lacked feedback and, therefore, did not provide information about which stimulus should be considered the GO stimulus.

The search for the correct GO stimulus can occur either within the correct or incorrect modality. If the search starts in the incorrect modality, the probability of a correct response may drop below 0.5. However, as the search progresses and the correct modality is identified, the probability of a correct response increases, eventually crossing the 0.5 threshold. This point, where the mean probability surpasses 0.5, is considered the moment of 'rule acquisition'.

Humans also use inference to grasp meaning beyond the information given, driving abrupt shifts in mindset and decoupling future performance from prior success or failure. To account for this, a model update cutoff point is introduced at the moment of rule acquisition (when the probability of success reaches or exceeds 0.5), and the consolidation stage prior is updated using only trials from this point onward.

The consolidation stage is marked by performance confidently exceeding chance. This stage is defined as the point where the 3% highest density interval (HDI) surpasses 0.5, signifying that performance is reliably above chance and the participant has identified the correct GO stimulus. In cases when several crossing points are detected, the point is defined as the average of the first and last crossing point.

We define the learning point as the number of trials from first feedback to the consolidation stage, i.e. the trial when HDI>0.5.

Results

We first assessed the effect of modality shifts on learning to validate the experimental design. Even with a relatively simple task, the results demonstrated a clear impact of modality shifts on learning duration. As expected, modality shifts (EDS sessions) resulted in longer learning times compared to IDS sessions where the modality of the GO stimulus was the same as the previous session (though the stimuli themselves changed). Participants required, on average, 13 more trials in the EDS compared to the IDS ($p < 0.001$, 95% CI: -19.575, -6.810; Figure 3 panel A).

An additional *order effect* was seen, i.e. the first round of IDS and EDS sessions required, on average, 7 more trials before participants learned the rule for the GO stimulus, compared to the second round ($p = 0.038$, 95% CI: -13.152, -0.387). This suggests that participants not only learned within individual sessions but also across sessions, becoming more efficient at handling the task and likely refining their under-

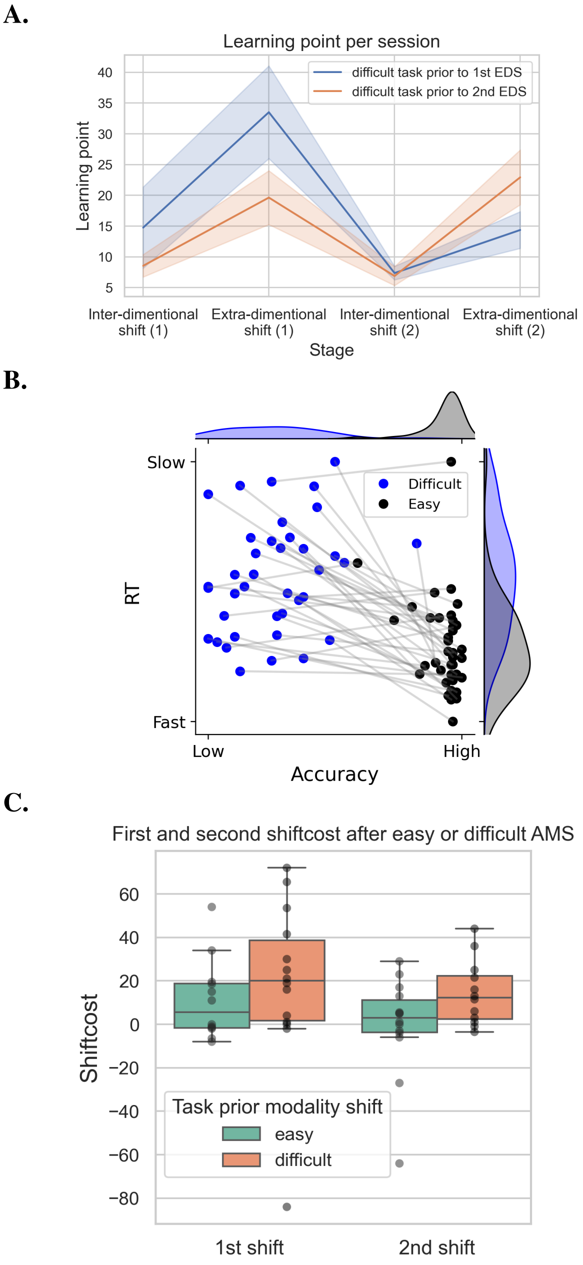

Figure 3: Results. **Panel A**: the learning point (first feedback to consolidation trial) of shift sessions, in order: IDS1, EDS1, IDS2, EDS2. The blue line indicates the average performance of participants who completed a difficult discrimination task prior to the first EDS, while the orange line represents those who completed it prior to the second EDS. Shaded areas indicate ±1 standard error of the mean. **Panel B**: participant-wise behavioural performance in accuracy (x-axis) and response time (y-axis), for easy (black) vs. difficult (blue) AMS. Axes range from minimum to maximum observed values on scales normalised from 0 to 1. **Panel C**: our *primary result* shows the shift costs of the first and second rounds of IDS and EDS depend on the attention manipulation (coded by box colours).

standing of task demands.

As described above, we established that attention manipulation worked using combined RT and accuracy results (see Figure 3 panel B). Given this, further analysis of the EDS sessions showed that learning following the difficult AMS required, on average, 11 more trials (p = 0.031, 95% CI: -21.429, -1.006), indicating that a more challenging prior task (with high discrimination difficulty set individually by staircase) increased learning time for subsequent sessions. This analysis examines the absolute number of trials it takes for the participant to learn the rule for the new GO stimulus.

We have also analysed the shift cost using the IDS1 as an individual baseline, and found similar results. Bayesian one-sided t-test of the shift costs yielded a Bayes factor of 2.83, providing some evidence of a difference between the two groups, with a significant difference between costs observed for the modality shifts following the difficult and easy discrimination tasks ($p = 0.022$, 95% CI: [-inf, -2.34]). These results show that shift costs were greater for participants who had completed the difficult discrimination task, supporting the hypothesis that modality shifts lead to higher learning costs when attention is more taxed by task difficulty.

In summary, the task design successfully evoked shift costs with modality changes, and manipulation of attention via task difficulty was shown to significantly affect subsequent learning time. Both the analysis of absolute trials and the shift cost indicated that participants required more trials to learn the GO stimulus following the difficult AMS.

Discussion

Our main result supports the hypothesis that taxing attention with task difficulty will increase subsequent shift costs during a modality shift. This builds on two intermediate outcomes.

Shift costs were calculated based on session-wise learning point estimates that account for the non-linearity and noisiness in how human performance data captures the learning process. This novel method, using a Bayesian framework, compared well to the sigmoid fitting approach used by Vasilev et al. (2022), when we tested both on human data with far fewer trials than the rats. Specifically, the Bayesian approach was robust to isolated incorrect trials, but better captured sudden ‘belief updating’ (even if in the incorrect direction). The method also provides a principled account of what is learned, based on the outside-of-chance (HDI$\geq$ 0.5) posterior estimate of the binomial distribution.

AMS sessions support the claim that attention was indeed taxed; the manipulation used forces attention allocation to achieve acceptable performance and thus progression. Results in Figure 3 panel B, show the manipulation successfully modulated participant performance in both accuracy and RTs.

Based on the successful induction of a modality shift cost and attention manipulation, our final result Figure 3, panel C illustrates the support for our main hypothesis. Vasilev et al. (2022) previously studied the role of perceptual attention and concurrent changes in neuronal representations on subsequent learning. They trained head-fixed rats in a GO/NOGO auditory-visual attentional set-shifting task, which our task aimed to directly translate to humans. Thus, Vasilev et al. (2022) task used similar stimuli, session structure, and manipulation; differing (aside from apparatus) mainly in that rats had many more trials per session. It was observed that learning to respond to novel and easily discriminable stimuli in one modality was slower after difficult discriminations in another previously-rewarded modality. That is, taxing attention at one time had an adverse effect on later contingency learning (with outcomes based on perceptual attention). With these two complementary results, evidence is strengthened that the fundamental processes of mammalian perception and attention are interdependent in support of contingency learning.

Implications and further work

Attention set-shifting tasks support the view of attention as a flexible control of limited resources (Barceló, 2020; Cowley & Lukander, 2016): learning after a shift illustrates flexibility, but modality shift costs illustrate that flexibility is limited. Our findings extend this, showing the sensitivity to demand on attention. An important question is why? Learning a new rule in an EDS has moderate perceptual demand and deterministic (not probabilistic) contingency, so should be a straightforward associative learning task. What mechanism 'carries over' the taxing of attention from the prior AMS?

In a difficult perceptual task, attention demand requires dynamic modulation of sensory neuron representations (Desimone & Duncan, 1995; Luck et al., 1997) to enhance sensory processing, which can create a temporary bottleneck. This matches the computational model of higher precision 'prior' prediction needed to discriminate perceptually 'close' stimuli. A contrasting easy perceptual task would require lower precision 'broad' priors. A consequence, illustrated by training any Bayesian update model, is that more precise priors require more updates to shift.

Additionally, the more finely tuned attention is to the previous task, the harder it is to switch to a new one, leading to more mistakes. This underscores the role of attentional focus and neural adaptation in task-switching performance.

Interestingly, research in psychophysics of perceptual learning transfer has shown that task difficulty influences how well learning generalizes to untrained conditions: easier training tasks lead to greater transfer (Ahissar & Hochstein, 2004; Wang et al., 2013). Easier tasks promote broader generalization across motion directions, visual fields, and perceptual dimensions — suggested by Reverse Hierarchy Theory to involve higher cortical areas. In contrast, more challenging tasks typically result in learning that is more specific and less transferable (Pavlovskaya & Hochstein, 2011). These earlier studies raise important questions about how task difficulty may shape subsequent performance. Our findings complement these studies by addressing these questions in terms of perceptual learning across modalities.

Our findings also highlighted a strong influence of *order* of the modality shift, as the first shift was more challenging regardless of difficulty of the preceding AMS. Thus, task learning carried through from EDS1 to EDS2.

Humans possess the powerful ability to grasp meaning beyond the information given, by forming concepts and making inferences. This is distinct from associative learning (conditioning), because the lessons learned via inference are not required to be reinforced by direct experience (Jensen et al., 2019). For example, one may use observed relations, e.g. A>B and B>C, to infer never-before-seen relations, A>C, a case of *transitive inference*. In our task, the introduction of the first modality shift poses a significant challenge to the existing task profile, necessitating updates in the task-relevant information and task-redundant information categories. Once these updates are completed, this updated task profile facilitates further modality shifts, as it becomes encoded within the learned task rule. This should be investigated in future with a study design that controls the task profile update. That could be accomplished by introducing more than two dimensions (modalities) to create equally 'new' modality shifts, or an experiment where the full verbal explanation is given, explicitly indicating the relevance of both modalities.

In Vasilev et al. (2022), motivation stemmed from receiving rewards for correct responses throughout the experiment. Most human participants were similarly motivated by receiving 'points' for correct responses, traded for a book voucher. However, some of the first participants were recorded without this reward due to an oversight, and a subset of these showed diminished performance in the maintenance stage of the task (perhaps due to boredom). This subset was excluded from analysis as per our first inclusion criterion. Yet their performance profile – comparable to extrinsically rewarded participants in the acquisition and consolidation stages (as required to progress in the task), but dropping off during maintenance – illustrates how humans' behaviour is dependent on their (potentially idiosyncratic) interpretation of the task demands.

Maintenance stage performance also highlights the role of cognitive depletion. While this may contribute to shift costs, it does not truly explain our findings. If it did, we would expect increasing shift costs with successive task switches. Instead, our data show that the first shift resulted in longer learning times, indicating that initial adaptation difficulties are not due to cognitive depletion.

Conclusions

In conclusion, we studied the impact of task difficulty on modality shift cost in humans. Our findings confirm that the reward-guided theory of shift cost is incomplete, since inducing a state of high-demand, focused, goal-directed perceptual attention negatively affects subsequent learning following a modality shift.

Also, our translation of a matching attention set-shifting task from rodent to human is novel, as most prior rodent studies have used olfactory cues unsuited to translation (V. J. Brown & Tait, 2016). Critically, we thus demonstrate cross-species similarity for the role of perceptual attention filtering in the shift cost, in two evolutionarily distant species.

References

Ahissar, M., & Hochstein, S. (2004). The reverse hierarchy theory of visual perceptual learning. *Trends in Cognitive Sciences*, *8*(10), 457–464.

Barceló, F. (2020). A Predictive Processing Account of Card Sorting: Fast Proactive and Reactive Frontoparietal Cortical Dynamics during Inference and Learning of Perceptual Categories. *Journal of Cognitive Neuroscience*, 1–21.

Brown, H., Friston, K., & Bestmann, S. (2011). Active Inference, Attention, and Motor Preparation. *Frontiers in Psychology*, *2*, 218.

Brown, V. J., & Tait, D. S. (2016). Attentional Set-Shifting Across Species. In T. W. Robbins & B. J. Sahakian (Eds.), *Translational Neuropsychopharmacology* (pp. 363–395). Springer International Publishing.

Cowley, B., & Lukander, K. (2016). Forest, Trees, Dynamics: Results from a Novel Wisconsin Card Sorting Test Variant Protocol for Studying Global-Local Attention and Complex Cognitive Processes. *Frontiers in Psychology*, *7*.

Dermody, N., Lorenz, R., Goddard, E., Villringer, A., & Woolgar, A. (2024, December 6). *Spatial and feature-selective attention interact to drive selective coding in frontoparietal cortex*.

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annual Review of Neuroscience*, *18*(1), 193–222.

Dias, R., Robbins, T. W., & Roberts, A. C. (1996). Dissociation in prefrontal cortex of affective and attentional shifts. *Nature*, *380*(6569), 69–72.

Esber, G. R., & Haselgrove, M. (2011). Reconciling the influence of predictiveness and uncertainty on stimulus salience: A model of attention in associative learning. *Proceedings of the Royal Society B: Biological Sciences*, *278*(1718), 2553–2561.

Fitts, P., & Posner, M. (1967). *Human performance*. Brooks/Cole.

Jensen, G., Alkan, Y., Ferrera, V. P., & Terrace, H. S. (2019). Reward associations do not explain transitive inference performance in monkeys. *Science Advances*.

Karni, A., & Sagi, D. (1993). The time course of learning a visual skill. *Nature*, *365*(6443), 250–252.

Lange, F., Kröger, B., Steinke, A., Seer, C., Dengler, R., & Kopp, B. (2016). Decomposing card-sorting performance: Effects of working memory load and age-related changes. *Neuropsychology*, *30*, 579–590.

Leong, Y. C., Radulescu, A., Daniel, R., DeWoskin, V., & Niv, Y. (2017). Dynamic interaction between reinforcement learning and attention in multidimensional environments. *Neuron*, *93*(2), 451–463.

Lindsay, G. W. (2020). Attention in Psychology, Neuroscience, and Machine Learning. *Frontiers in Computational Neuroscience*, *14*, 29.

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural mechanisms of spatial selective attention in areas v1, v2, and v4 of macaque visual cortex. *Journal of Neurophysiology*, *77*(1), 24–42.

Owen, A. M., Roberts, A. C., Polkey, C. E., Sahakian, B. J., & Robbins, T. W. (1991). Extra-dimensional versus intra-dimensional set shifting performance following frontal lobe excisions, temporal lobe excisions or amygdalo-hippocampectomy in man. *Neuropsychologia*, *29*(10), 993–1006.

Pavlovskaya, M., & Hochstein, S. (2011). Perceptual learning transfer between hemispheres and tasks for easy and hard feature search conditions. *Journal of Vision*, *11*(1), 8–8.

Tenenbaum, J. B., & Griffiths, T. L. (2001). Generalization, similarity, and Bayesian inference. *The Behavioral and Brain Sciences*, *24*(4), 629–640, discussion 652–791.

Vasilev, D., Safaei, N., Iwai, R., Bahmani, H., Zouridis, I. S., Watanabe, M., Logothetis, N. K., & Totah, N. K. (2022). Focusing perceptual attention in the past constrains outcome-based learning in the future by adjusting cortico-cortical interactions. *bioRxiv*.

Wang, X., Zhou, Y., & Liu, Z. (2013). Transfer in motion perceptual learning depends on the difficulty of the training task. *Journal of Vision*, *13*(7), 5–5.