Predicting human prospective beliefs and decisions to engage using multivariate classification analyses of behavioural data

David Soto (d.soto@bcbl.eu)

Ning Mei (n.mei@bcbl.eu)

Sean Rankine

Einar Olafsson

Basque Center of Cognition, Brain, and Language, San Sebastian (Spain); Ikerbasque, Basque Foundation for Science

Abstract

Metacognition can be deployed retrospectively (i.e. to reflect on the correctness of our recent behaviour) or prospectively (i.e. to make predictions of success in one's future behaviour or make decisions about strategies to solve future problems). We sought to investigate the factors that determine this sort of prospective decision making. Human participants performed a visual discrimination task followed by ratings of stimulus visibility and response confidence. Prior to each discrimination trial participants made prospective judgments concerning the upcoming task. In Experiment 1, they rated their belief of future success. In Experiment 2, they rated their decision to adopt a focussed attentional state. Both types of prospective decisions were related to behavioural performance in different ways. Prospective beliefs of success were associated with no performance changes while prospective decisions to engage attention were followed by better self-evaluation of the correctness of behavioural responses. Using standard machine learning classifiers we found that the current prospective decision could be predicted from information concerning task-correctness, stimulus visibility and response confidence from previous trials. In both Experiments, awareness and confidence were more diagnostic of the prospective decision than task correctness. Notably, classifiers trained with prospective beliefs of success in Experiment 1 predicted decisions to engage in Experiment 2 and vice-versa. These results indicate that the formation of these seemingly different prospective decisions share a common, dynamic representational structure.

`https://tinyurl.com/yxjsosnf`.

Keywords: metacognition, prospective decisions, confidence, visual awareness, multivariate classification

Introduction

Prior research in the memory domain addressed how people make prospective judgments of learning during study (Nelson & Dunlosky, 1991; Koriat, 1997), and revealed, for instance, how decisions to study rely on the evaluation of one's own learning (Nelson & Leonesio, 1988; Metcalfe & Finn, 2008) and how this self-evaluation during study relates to subsequent memory accuracy. However, little is know about the factors that influence prospective metacognition during perceptual decision making. In addition to monitoring one's own behavioral performance and forming prospective beliefs about future success, people also engage in self-regulation. For instance, people may also decide to put more attention when they lack confidence in their knowledge or stop a further study when they are confident.

We here sought to investigate whether or not the formation of seemingly different types of prospective decisions (i.e. beliefs of success and decisions to engage with the environment) make use of similar information and recruit similar processes. After all, one's certainty of the adequacy of one's behavioural responses may be dependent on a host of different factors, including stimulus visibility, interference from distracting information and additional biases and heuristics (Koriat, 2007). For instance, given a challenging perceptual task, a state of the low visibility of the critical target may encourage the observer to decide to invest more effort in the next trial but may also lead to a reduction in confidence about his prospective accuracy. It is not known whether or not different prospective beliefs (i.e. involving a decision to adopt a more focused attention state on the next trial vs. prospective beliefs of success) are dependent on the same factors.

Method

Experimental task and procedure

Using a paradigm involving visual perceptual decisions we investigated the factors that predict future prospective beliefs of performance success vs. prospective decisions to engage with the environment (i.e. decisions to adopt a focused attention state). Specifically, we modified an existing paradigm (Jachs, Blanco, Grantham-Hill, & Soto, 2015) to quantify the contribution of the state of visual awareness, task-confidence and task-correctness from previous trials to the prospective decision making in the upcoming trial. Figure 1 illustrates the experimental task. Participants were presented with an oriented Gabor patch near the threshold of visual awareness. Prior to the presentation of the Gabor, on each trial, participants indicated their belief of success (i.e. low or high) in the upcoming orientation discrimination task (Experiment 1) or indicated their decision to engage a focused attention state (low or high; Experiment 2). Following the presentation of the

Gabor, participants rated their visual awareness, responded to the Gabor orientation and rated their confidence in the orientation response (Jachs et al., 2015; Charles, Van Opstal, Marti, & Dehaene, 2013).

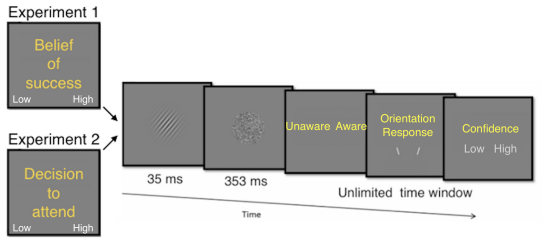

Figure 1: Illustration of the experimental task

Machine learning protocols

We used standard machine learning algorithms to investigate whether prospective beliefs of success on the current trial (Experiment 1) or the current prospective decision to engage attention (Experiment 2) can be predicted based on the participants' experiences during task performance on previous trials. Hence we aimed to predict whether the belief of success was low or high (Experiment 1) or whether participants decide to engage in high or low attention state (Experiment 2) given a vector of features including correctness, visual awareness, and confidence from the previous trials, critically, considering 1-back, 2-back, 3-back, and 4-back trials separately for training and testing the classifier. We employed a regularized random forest and a logistic regression classifier to perform the same classification problem in order to estimate the stability of the decoding performance and to provide more information about the pattern of results based on the feature importance and coefficients of the different features used for classification. In order to estimate the variance of the decoding performance, we conducted a 100-fold shuffle splitting cross-validation for each subject, each N-back, and each classifier. Each fold was constructed by shuffling the order of the instances including both the features and targets, and then 80% of the instances were selected to form a training set while the rest 20% became the testing set. After fitting the classifier with the training set, probabilistic predictions were made for the testing set. The performance was measured by the area under the receiver operating curve (ROC AUC). The statistical significance of the classification scores in each condition (i.e. N-back) were determined by a permutation t-test (Ludbrook & Dudley, 1998). In it, decoding scores were assessed relative to their corresponding chance level estimates pairwise.

Results

Across the two experiments, we found that a logistic regression classifier significantly predicted the upcoming prospective belief of success (Experiment 1) and the prospective decision to engage attention (Experiment 2) based on the pattern of awareness, confidence, and correctness exhibited in the previous trials. This result was replicated with a regularized random forest classifier. Information from the previous trial led to the highest accuracy in the classification of the prospective belief of success/decision to engage, with classification accuracy increasingly dropping with up to 4 trials back (see Figure 2).

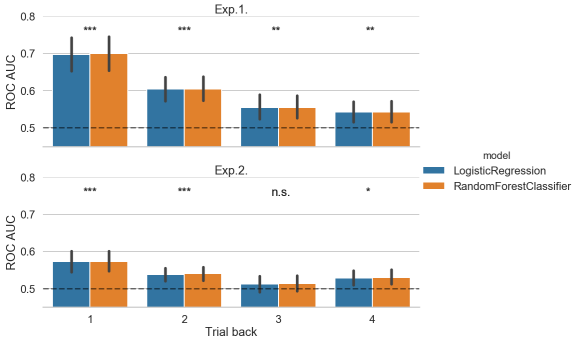

Figure 2: Results from the logistic regression and random forest models tested separately for each of trials back. ***: $p < 0.001$, **: $p < 0.01$, *: $p < 0.05$, Bonferroni corrected; n.s.: not significant. Error bars represent bootstrapped 95% confidence intervals.

Next, we assessed the relevance of each of the different attributes (awareness, confidence, and correctness) for the classification. As we were interested in understanding the factors that contribute to future beliefs of success, we analyzed the weight coefficients (odd ratios) from the logistic regression, and the feature importance of the random forest classifier by means of an ANOVA with time window (1, 2, 3 and 4 trials back) and feature attribute as factors. Since the above classification results already showed that classification accuracy decreases with the number of trials back, additional analyses based on significant main effects of time window are not considered further. As illustrated in Figure 3 (top-left panel), following a rating of high awareness/confidence in the previous trial, observers were four times more likely to report a high belief of success; following a rating of high confidence in the previous trial, observers were 1.4 times more likely to report a high belief of success. Similar patterns were observed for the feature importance of the random forest model (bottom-left panel). Figure 3, panels (top-right and bottom-right) illustrate similar findings for the prospective decision to engage attention.

Then, we analyzed the data from both experiments together in order to estimate how much information learned from one experiment can be transferred to the other experiment. We found that the logistic regression model was able to decode the decision to attend (Experiment 2) based on the pattern of awareness, confidence, and correctness from the previous trial of Experiment 1 in which participants rated their belief of success ($p = 0.0106$), but not with the attributes in 2, 3, or 4 trial back ($p > 0.9$). Then we trained the classifier using

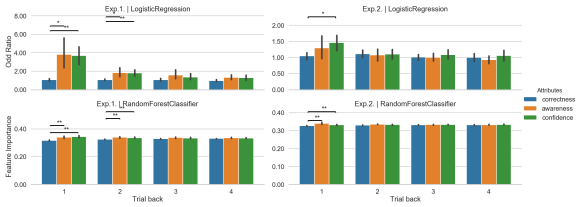

Figure 3: Odd ratios from the logistic regression and feature importance from the random forest models tested separately for each of the trials back. **: $p < 0.01$, *: $p < 0.05$, Bonferroni corrected, n.s.: not significant. Error bars represent bootstrapped 95% confidence intervals.

the data from Experiment 2 and tested the classifier using the data from Experiment 1. The classifier was able to decode the prospective belief of success based on the pattern of awareness, confidence, and correctness from the previous trial of Experiment 2 in which observers made decisions to engage ($p = 0.0004$); however this was not the case for N = 2 or 3 or 4 trials back (lowest p-value = 0.0739). These results are depicted in Figure 4. This pattern of results was fully replicated using analyses based on a random forest classifier.

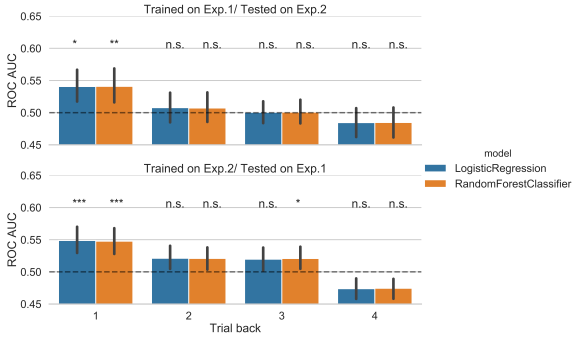

Figure 4: Across-experiment generalization results. Classifiers were trained on data from either experiment and tested on the other experiment. This was done separately for each of trials back. **: $p < 0.001$, *: $p < 0.05$, Bonferroni corrected, n.s.: not significant. Error bars represent bootstrapped 95% confidence intervals.

Finally, we asked whether these seemingly different prospective decisions play a functional role in shaping subsequent perceptual discrimination or metacognitive performance. We then tested whether the performance in the Gabor discrimination task was affected by the type of prospective belief or decision to engage attention. One possibility is that estimations of the high probability of success may encourage observers to invest more cognitive resources in the upcoming trial and hence facilitate performance in a similar way to decisions to engage focused attention. We found that prospective beliefs of success were associated with no performance changes in the Gabor orientation discrimination performance (d') while prospective decisions to engage attention were followed by better self-evaluation of the correctness of behavioural response (i.e. the relationship between confidence ratings and task accuracy - meta-d')(Maniscalco & Lau, 2012). Figure 5 illustrates these results.

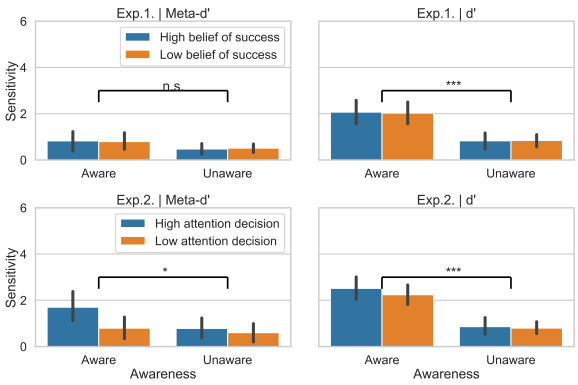

Figure 5: Perceptual sensitivity (d') and metacognitive sensitivity (meta-d') scores as a function of the state of awareness and the decision to engage attention. Error bars represent standard error of the mean.

Discussion

Across two experiments, we found that standard machine learning classifiers significantly predicted the upcoming prospective belief of success based on the pattern of awareness, confidence, and correctness exhibited in previous trials, each tested separately. Information from the previous trial led to the highest accuracy in the prospective belief of success, with classification accuracy increasingly dropping with up to four trials back. This finding is consistent with prior work (Fleming, Massoni, Gajdos, & Vergnaud, 2016). The present study goes beyond to show that this pattern of results generalises to different types of prospective decisions, namely, individual decisions to engage with the task. This conclusion is further supported by the finding that a classifier trained on data from one experimental context (e.g. involving prospective beliefs) could be used to predict a different prospective decision (i.e., to engage attention) and vice-versa. However, this generalization only occurred when data from just the prior trial was considered. We suggest that a common representational structure supports the dynamic formation of seemingly different types of prospective judgements.

However, despite the common information pattern based on past confidence and awareness that underlies the formation of prospective judgments, only the prospective decisions to attend appeared to influence the observers retrospective evaluation of the correctness of perceptual decisions (meta-d'), but this was not the case following a prospective belief of high success. From the perspective of the 'self-fulfilling prophecy', prospective beliefs of success may set an expectation concerning upcoming behavioural performance that the participant is motivated to meet (Weinberg, Gould, & Jack-

son, 1979; Zacharopoulos, Binetti, Walsh, & Kanai, 2014) and accordingly, a belief of performance success might in principle encourage observers to invest more cognitive resources in the upcoming trial. Our results suggest that this is not the case.

Prospective beliefs of success concerned here a low-level perceptual discrimination task. It is possible that prospective beliefs are more diagnostic of the upcoming behavioural performance in different task domains, namely, memory (Nelson & Dunlosky, 1991; Kao, Davis, & Gabrieli, 2005). Another possibility is that decisions to engage attention are more likely to be embodied by comparison to beliefs of success. Accordingly, recent theoretical frameworks borrowing from ecological psychology (Gibson, 1979) propose that perceptual biases and decisions are not independent of action. In this framework, perception drives decisions and action, but actions also drive subsequent experiences in a dynamic perception-action loop (Lepora & Pezzulo, 2015). We propose that decisions to engage with the environment (i.e. to deploy a focussed attention state) are more likely to be embodied in the action system and hence are very likely to trigger commitment towards that action, while prospective beliefs may not. It is possible that decisions to engage attention trigger preparatory control which in turn can influence subsequent cognitive processing.

In summary, this study indicates that a common representational structure supports the dynamic formation of seemingly different types of prospective judgements. Additional research is however needed to test whether these observations generalize to different task contexts and cognitive domains beyond perceptual decision making, including those in which the precision of retrospective metacognition is not correlated across tasks (e.g. (Kelemen, Frost, & Weaver, 2000)).

Acknowledgments

D.S. acknowledges support from the Spanish Ministry of Economy and Competitiveness, through the 'Severo Ochoa' Programme for Centres/Units of Excellence in R & D (SEV-2015-490) and project grants PSI2016-76443-P from MINECO and PI-2017-25 from the Basque Government.

References

Charles, L., Van Opstal, F., Marti, S., & Dehaene, S. (2013). Distinct brain mechanisms for conscious versus subliminal error detection. *Neuroimage*, *73*, 80–94.

Fleming, S. M., Massoni, S., Gajdos, T., & Vergnaud, J.-C. (2016). Metacognition about the past and future: quantifying common and distinct influences on prospective and retrospective judgments of self-performance. *Neuroscience of Consciousness*, *2016*(1).

Gibson, J. J. (1979). *The ecological approach to visual perception: classic edition*. Psychology Press.

Jachs, B., Blanco, M. J., Grantham-Hill, S., & Soto, D. (2015). On the independence of visual awareness and metacognition: A signal detection theoretic analysis. *Journal of experimental psychology: human perception and performance*, *41*(2), 269.

Kao, Y.-C., Davis, E. S., & Gabrieli, J. D. (2005). Neural correlates of actual and predicted memory formation. *Nature neuroscience*, *8*(12), 1776.

Kelemen, W. L., Frost, P. J., & Weaver, C. A. (2000, jan). Individual differences in metacognition: Evidence against a general metacognitive ability. *Memory & Cognition*, *28*(1), 92–107. Retrieved from https://doi.org/10.3758/bf03211579 doi: 10.3758/bf03211579

Koriat, A. (1997). Monitoring one s own knowledge during study: A cue-utilization approach to judgments of learning. *Journal of Experimental Psychology: General*, *126*(4), 349–370. Retrieved from https://doi.org/10.1037/0096-3445.126.4.349 doi: 10.1037/0096-3445.126.4.349

Koriat, A. (2007). Metacognition and consciousness. In P. D. Zelazo, M. Moscovitch, & E. Thompson (Eds.), *The cambridge handbook of consciousness* (pp. 289–326). Cambridge University Press. Retrieved from https://doi.org/10.1017/cbo9780511816789.012 doi: 10.1017/cbo9780511816789.012

Lepora, N. F., & Pezzulo, G. (2015, apr). Embodied choice: How action influences perceptual decision making. *PLOS Computational Biology*, *11*(4), e1004110. Retrieved from https://doi.org/10.1371/journal.pcbi.1004110 doi: 10.1371/journal.pcbi.1004110

Ludbrook, J., & Dudley, H. (1998). Why permutation tests are superior to t and f tests in biomedical research. *The American Statistician*, *52*(2), 127–132.

Maniscalco, B., & Lau, H. (2012). A signal detection theoretic approach for estimating metacognitive sensitivity from confidence ratings. *Consciousness and cognition*, *21*(1), 422–430.

Metcalfe, J., & Finn, B. (2008, feb). Evidence that judgments of learning are causally related to study choice. *Psychonomic Bulletin & Review*, *15*(1), 174–179. Retrieved from https://doi.org/10.3758/pbr.15.1.174 doi: 10.3758/pbr.15.1.174

Nelson, T. O., & Dunlosky, J. (1991). When people's judgments of learning (jols) are extremely accurate at predicting subsequent recall: The 'delayed-jol effect'. *Psychological Science*, *2*(4), 267–271.

Nelson, T. O., & Leonesio, R. J. (1988). Allocation of self-paced study time and the "labor-in-vain effect.". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *14*(4), 676–686. Retrieved from https://doi.org/10.1037/0278-7393.14.4.676 doi: 10.1037/0278-7393.14.4.676

Weinberg, R., Gould, D., & Jackson, A. (1979). Expectations and performance: An empirical test of bandura's self-efficacy theory. *Journal of sport psychology*, *1*(4), 320–331.

Zacharopoulos, G., Binetti, N., Walsh, V., & Kanai, R. (2014). The effect of self-efficacy on visual discrimination sensitivity. *PloS one*, *9*(10), e109392.