Resolving the reversal paradox of memory confidence

Paul Masset (pmasset@cshl.edu)
Watson School of Biological Sciences
Cold Spring Harbor Laboratory
Cold Spring Harbor NY 11724 USA

Adam Kepecs (kepecs@cshl.edu)
Cold Spring Harbor Laboratory
Cold Spring Harbor NY 11724 USA

Abstract

The ability to remember previously encountered situations and the ability to give a subjective judgment on the accuracy of a decision are two fundamental aspects of adaptive behavior. Memory allows an individual to decouple actions from momentary percepts, while confidence provides graded degrees of belief that enable choosing between competing alternatives. Humans possess the ability to know when to trust recalled memories (high confidence) and when to doubt them (low confidence). However, despite the evolutionary advantage of being able to assign appropriate levels of confidence to memory recall, the accuracy of such confidence reports have been questioned. In fact, previous studies have found positive, null or even negative correlations between confidence and recall accuracy. Here we present a framework that resolves the apparent paradox of negative correlations. We define a decision variable that assigns a distinct difficulty level for each memory recall decision. We show that within our framework the confidence reports in memory recall follow the signatures of a statistical definition of decision confidence. The apparent paradox can be explained as a difference between the objective and subjective category of the stimuli.

Keywords: memory ; confidence ; decision variable ; false memory ; paradox ; model

The reversal paradox of confidence in memories

Are confidence reports in memories reliable? The answer to this question requires an operational definition of confidence. Confidence can be defined mathematically as the probability of an event occurring given available evidence (Hangya, Sanders, & Kepecs, 2016). Confidence can also be described psychologically as a subjective feeling. A number of studies have shown that confidence in memories can be correlated with accuracy (Wixted & Wells, 2017). However, a number of studies have also shown that in some situations confidence and accuracy can be negatively correlated (Roediger & DeSoto, 2014).

Here we reanalyze data from DeSoto and Roediger who examined the relationship between confidence reports and accuracy in the DeeseRoedigerMcDermott (DRM) paradigm (Roediger & DeSoto, 2014). Briefly, in their study subjects listened to 150 words taken from 10 lists that they had to remember. They then performed an unrelated general knowledge task to ensure the absence of effects by short term memory or working memory (Figure 1a). In the test phase, subjects listened to 300 words and were asked to classify the words as studied or unstudied . In addition to the 150 words previously studies, the 150 new words were divided equally into 3 categories: (i) strongly related lures, words with strong semantic links to the studied words, (ii) weakly related lures, weakly semantically related to studied words and (iii) unrelated lures that are not related to previously studied words. After hearing each word, subjects were asked to report their confidence in their choice. In the first experiment, subjects were asked to enter a confidence rating in the range 0-100 with a keyboard and in the second experiment subjects entered their confidence rating using a sliding scale.

The authors showed that for both confidence reporting techniques, confidence was positively correlated with accuracy for correctly identified targets and negatively correlated with accuracy for incorrectly identified lures. These observations lead to a paradox in which confidence and accuracy are positively correlated for correct choices and negatively correlated for incorrect choices. This could imply that different cognitive processes are involved in generating confidence estimates in both situations. In the rest of this paper we resolve this paradox by showing that it arises from the lack of an explicit model linking decision to confidence reports.

Identifying a decision variable for memory recall

In the traditional signal detection theory analysis of memory recall, targets and lures are grouped in one position for each category. On a given trial, the memory of the stimulus will be a sample taken from the corresponding gaussian distribution. It will classified by the subject as target or lure depending on where the sample falls with respect to the decision boundary.Using the hits, correct rejections, false alarms and misses, one can estimate the parameters of the gaussians as shown in figure 1A. There is no distinction in the difficulty of the recall within each category. In sensory psychophysics the difficulty to discriminate a sensory stimulus is a parameter that can be explicitly manipulated and often so along an intuitive scale. For example, in the classic random dot task, the motion coherence towards one or the other side provides a simple measure of stimulus discriminability. We define the position of a given stimulus along this metric as the decision variable associated with this stimulus. Here we propose to define a decision variable for each memory recall decision. The memory noise is still captured by a gaussian variable but the mean of the gaussian depends on the value of the decision variable for the given word as shown in figure 1B.

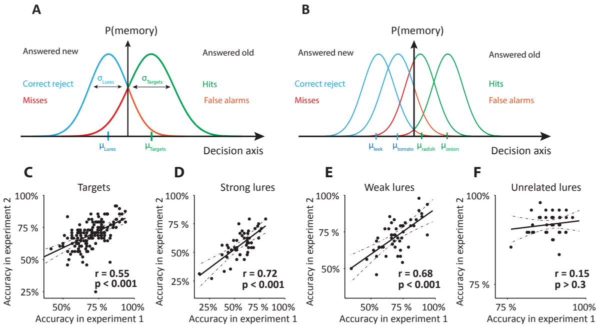

Figure 1: Identifying a decision variable

The difficulty of the DRM task arises from the semantic relatedness of the words presented. Here we show that the decision variable for a given word can be defined as the average accuracy for this word. For semantically related words, the targets and the weak and strongly related lures, the accuracy for a given word across the two experimental populations is strongly correlated, figure 1C-E. For semantically unrelated lures, there is no significant correlation between the accuracy across the two experiments, Figure1F. These results suggest that the difficulty induced by the semantic relatedness is shared across the population and therefore that the average accuracy captures a measure of recall difficulty that is shared across the population. Using this decision variable we can assign a difficulty level to each memory decision in the DRM task and apply decision models developed in sensory psychophysics.

Confidence reports about memory recall have signatures of statistical confidence

The confidence in a decision can be mathematically defined as the probability of being correct given the available evidence. Using assumptions similar to those presented above, one can derive a statistical model of decision confidence which shows how confidence should vary with the decision variable and choice (Hangya et al., 2016). The statistical model of decision confidence predicts that confidence follows 3 signatures: (i) Confidence predicts accuracy, (ii) confidence increases with discriminability for correct choices and decreases with accuracy for error choices and (iii) confidence predicts accuracy behind stimulus discriminability. Here we show that confidence reports in memory recall in the DRM task follow the 3 signatures of the statistical model of decision confidence.

In Figure 2A, we plot the calibration curve. We show that confidence predicts accuracy for trials where the subjects answered that they had seen the stimulus previously. In figure 2B, we plot the vevaiometric curve (from the Greek: βεβαιοζ , certain). We show that confidence increases with discriminability for correctly identified targets and decreases with discriminability for incorrectly identified lures. In figure 2C, we plot the conditioned psychometric curve. We show that, for a given level of the decision variable, if we condition the accuracy on the confidence report, subjects are more accurate when reporting a higher confidence. The analysis presented here is performed across the population but having defined a decision variable for each stimulus we can perform a similar analysis at the single subject level. We extract some metrics that allow us to compare and contrast the variability in the ability of subjects to report their confidence.

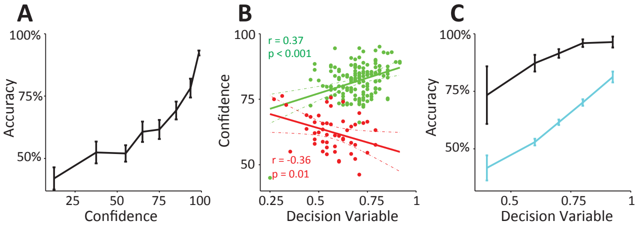

Figure 2: Signatures of decision confidence in memory recall

Resolving the confidence reversal paradox

In this work, we explain the apparent paradox of negative correlations in confidence report as originating from a lack of explicit decision model. The paradox arises when there is a discrepancy between the true category of a stimulus (target or lure) and the subjective category of the report (seen previously or new). The paradox is part of a broader category of reversal paradoxes of which the most famous is Simpson's paradox. Our explicit model for generating confidence reports allows us to show that the behavioral reports are entirely consistent with reports based on the mathematically defined notion of confidence, the probability of being correct given the evidence.

Acknowledgments

We thank Henry L. Roediger and K. Andrew DeSoto for kindly sharing their data and Tony Zador for his careful reading of the New York Times.

References

Hangya, B., Sanders, J. I., & Kepecs, A. (2016). A Mathematical Framework for Statistical Decision Confidence. *Neural Computation*, *28*(9), 1840–1858.

Roediger, H. L., & DeSoto, K. A. (2014). Confidence and memory: assessing positive and negative correlations. *Memory (Hove, England)*, *22*(1), 76–91.

Wixted, J. T., & Wells, G. L. (2017). The Relationship Between Eyewitness Confidence and Identification Accuracy: A New Synthesis. *Psychological Science in the Public Interest*, *18*(1), 10–65.