



Understanding metacognitive confidence: Insights from judgment-of-learning justifications



Radka Jersakova^{a,*}, Richard J. Allen^a, Jonathan Booth^b, Céline Souchay^c, Akira R. O'Connor^b

^a School of Psychology, University of Leeds, Leeds LS2 9JT, UK

^b School of Psychology and Neuroscience, University of St Andrews, St Andrews KY16 9JP, UK

^c Laboratoire de Psychologie et Neurocognition CNRS UMR 5105, Université Grenoble Alpes, Grenoble 38040, France

ARTICLE INFO

Article history:

Received 19 January 2017

Revision received 31 July 2017

Keywords:

Metacognition
Judgments-of-learning
Episodic memory
Confidence
Linguistics

ABSTRACT

This study employed the delayed judgment-of-learning (JOL) paradigm to investigate the content of metacognitive judgments; after studying cue-target word-pairs, participants predicted their ability to remember targets on a future memory test (cued recognition in Experiments 1 and 2 and cued recall in Experiment 3). In Experiment 1 and the confidence JOL group of Experiment 3, participants used a commonly employed 6-point numeric confidence JOL scale (0–20–40–60–80–100%). In Experiment 2 and the binary JOL group of Experiment 3 participants first made a binary *yes/no* JOL prediction followed by a 3-point verbal confidence judgment (*sure-maybe-guess*). In all experiments, on a subset of trials, participants gave a written justification of why they gave that specific JOL response. We used natural language processing techniques (latent semantic analysis and word frequency [*n*-gram] analysis) to characterize the content of the written justifications and to capture what types of evidence evaluation uniquely separate one JOL response type from others. We also used a machine learning classification algorithm (support vector machine [SVM]) to quantify the extent to which any two JOL responses differed from each other. We found that: (i) participants can justify and explain their JOLs; (ii) these justifications reference cue familiarity and target accessibility and so are particularly consistent with the two-stage metacognitive model; and (iii) JOL confidence judgements do not correspond to *yes/no* responses in the manner typically assumed within the literature (i.e. 0–40% interpreted as *no* predictions).

© 2017 Elsevier Inc. All rights reserved.

Introduction

Cognitive processes are accompanied by states of awareness that guide evaluation of their function and content (Fleming, Dolan, & Frith, 2012; Nelson & Narens, 1990; Overgaard & Sandberg, 2012). This metacognitive awareness (or monitoring) is understood as an inferential process, relying on cues derived from the task at hand to construct judgments about performance (Koriat, 2000), that has behavioral consequences (Koriat, Ma'ayan, & Nussinson, 2006; Metcalfe & Finn, 2008a). As such, understanding the basis on which these metacognitive judgments are made is crucial. While there have been numerous paradigms developed for the study of metacognition, subjective report from participants remains a vital method for tapping into metacognitive and related processes (Jersakova, Moulin, & O'Connor, 2016; Overgaard & Fazekas, 2016). Confidence in particular is the hallmark of metacognitive judgments and the most commonly used

paradigm for investigating metacognition across domains, ranging from decision making and reasoning (Ackerman & Thompson, 2014; Fletcher & Carruthers, 2012; Yeung & Summerfield, 2012) to perceptual judgments (Fleming et al., 2015; Peters & Lau, 2015; Rahnev, Koizumi, McCurdy, D'Esposito, & Lau, 2015) and memory evaluations (Dunlosky, Serra, Matvey, & Rawson, 2005; Finn & Metcalfe, 2007; Koriat & Levy-Sadot, 2001).

Metacognitive confidence is often interpreted as corresponding to quantity and quality of some (internal) evidence gathered toward the judgment being made (e.g. ease of reading as evidence that an item has been sufficiently learned and will later be remembered; Rhodes & Castel, 2008) and reflecting the probability that the given judgment is correct (Kepecs & Mainen, 2012). Whereas metacognitive research has tended to focus on examining which variables lead to general shifts in confidence (e.g. Alban & Kelley, 2013; Castel, McCabe, & Roediger, 2007; Koriat & Levy-Sadot, 2001; Rhodes & Castel, 2008), there is less understanding of what expressed metacognitive confidence *means*. This includes considerations of what differentiates one confidence level (e.g. 40% confidence) from another (e.g. 60% confidence) and whether

* Corresponding author.

E-mail address: r.jersakova@hotmail.com (R. Jersakova).

confidence judgments simply rank items against each other or whether they can be further interpreted (e.g. as *yes/no* predictions). Understanding this has implications for both theory and practice and our ability to interpret participant behavior in the laboratory. In this study we focused on metacognitive judgments made about memory (metamemory) to investigate what expressed metacognitive confidence represents.

We employed the delayed judgments-of-learning (JOL) paradigm; a prediction of whether recently learned information would be successfully retrieved in the future (Nelson & Dunlosky, 1991). In a typical delayed JOL experiment, participants study cue-target word pairs following which they are again presented with the studied cues and asked to make a prediction about whether they think they would retrieve the target on the subsequent memory test. These predictions are usually made on a numeric confidence scale expressed as percentages; e.g. 0–20–40–60–80–100%. This study evaluated how participants construct and justify their delayed JOLs by asking them to provide written reports alongside their JOLs. Participants were given no instructions on how to write their justifications, as we wanted to see what features would be referenced spontaneously. We used natural language processing techniques to investigate the type of information and explanation that characterizes each JOL and differentiates one JOL from another (e.g. 20% vs. 40%), as well as to quantify the extent to which any two JOLs are justified with reference to different types of evidence.

The experiments presented here draw on research investigating retrospective confidence in contents of memory retrieval, which has established that probing participants for explanations and justifications of their answers is a powerful tool for characterizing processes underlying cognition and metacognition. For example, Koriat, Lichtenstein, and Fischhoff (1980) asked participants to list reasons for and against their chosen answer to a general knowledge question. They observed that confidence was influenced by the amount of evidence accessed in support of the given answer, lending support to the idea that confidence is a result of a process of evaluation of different sources of evidence. More recently, Selmecky and Dobbins (2014) asked participants to justify their confidence in recognition judgements. Analysis of these justifications showed a pattern of results consistent with dual-process accounts of recognition memory (see Yonelinas, 2002); for example, the presence of ‘remembering’ characterized high confidence *old* responses and its absence corresponded to high confidence *new* responses. In other words, this quantitative analysis of subjective reports lent support to one side of an on-going debate in recognition memory.

Furthermore, these results were obtained without explicit instructions or theory-laden manipulations from the experimenters, who did not highlight specific experiences or types of evidence for participants to focus on. This is in contrast to classic metamemory research, which relies largely on explicitly asking participants about access to specific types of information relating to the studied items (e.g. the degree to which they can remember partial characteristics, such as the first letter, of the target item, Koriat, 1993). Such an approach allows for the evaluation of how access to specific features of the studied items influences confidence judgments. However, it leaves open the question whether participants would rely on this type of information in their judgments if their attention was not drawn to it by asking (see Hertzog, Fulton, Sinclair, & Dunlosky, 2014). Overall, studies that have asked participants to justify their responses (see also Gardiner, Ramponi, & Richardson-Klavehn, 1998; Urquhart & O’Connor, 2014; Williams, Conway, & Moulin, 2013) indicate that much can be learned from the relatively infrequent practice of asking participants to explain their metacognitive judgments. In the present study we adopted and developed the analytical approach

to participant justifications pioneered by Selmecky and Dobbins (2014) to gain insight into processes underlying JOL confidence.

Turning specifically to the theoretical issues that could be informed by this approach, there is a debate about how numeric confidence JOL responses relate to a binary (*yes/no*) sub-classification of the scale. The idea that low confidence JOL predictions should equate to a rejection of future retrieval makes sense probabilistically (i.e. a 40% predicted success rate should correspond to a 60% predicted failure rate). Correspondingly, it is common practice to interpret confidence as representing success probabilities as evidenced by the use of calibration measures (e.g. Finn & Metcalfe, 2007; Koriat, Sheffer, & Ma’ayan, 2002; Serra & England, 2012). Similarly, in cases where binary data is required for analysis purposes, confidence responses are commonly split equally into a binary (*yes/no*) sub-classification (e.g. Hanczakowski, Zawadzka, Pasek, & Higham, 2013; Masson & Rotello, 2009).

There is some theoretical support for the idea that participants explicitly make a *yes/no* sub-classification in their interpretation of the confidence scale, which has been suggested for a range of metacognitive tasks (Dunlosky et al., 2005; Liu, Su, Xu, & Chan, 2007). For example, Dunlosky et al. (2005) observed that when participants were asked to make a confidence judgment about the accuracy of their JOL prediction (a second-order judgment; SOJ), a plot of the SOJ magnitude against JOL confidence yielded a U-shaped function – participants were most confident in the predictions that lay on the extremes of the JOL scale (with least confidence at the mid-range of the scale). Dunlosky et al. (2005) interpreted the SOJ function minimum as the point where *yes* and *no* predictions diverge and suggested that one possible interpretation is that JOLs could be viewed as a two-step process that consists first of a *yes* or *no* judgment, directly followed by an assignment of confidence.

While the binary *yes/no* sub-classification seems intuitive and plausible, it has not yet been directly tested. Further, its relationship to confidence, such as whether we can split the numerical scale into equal proportion of *yes* and *no* responses, is poorly understood. This is crucial since it is a theory-laden interpretation of JOLs that is commonly employed in the literature when analysing confidence data and yet one that lacks explicit support. This absence of verification of widely held interpretations of how participants respond highlights the need to understand better how confidence and binary *yes/no* judgments relate to each other.

An alternative approach to understanding JOL confidence has been to investigate the underlying processes that shape the formation of JOLs. The early literature focused on explaining JOLs as a result of single process (target retrieval) evaluations (e.g. Nelson & Dunlosky, 1991). In this view, it was assumed that participants accrue one type of evidence (the degree to which the target is accessible) toward their JOL—the more evidence they collect, the higher their JOL. According to this view, different JOLs (e.g. 60% as compared to 80%) express different degrees of access to the target. An alternative two-stage view has proposed a quick pre-retrieval stage driven by cue-familiarity followed by an effortful memory search (target accessibility evaluation; Benjamin, 2005). Metcalfe and Finn (2008b) further elaborated this view, suggesting the first stage can result in: (i) a quick *don’t know* decision driven by lack of cue familiarity (expressed as responding with the lowest point on the JOL scale); or (ii) the initiation of the second effortful retrieval stage. In this case, there should be qualitatively different processes that underlie the lowest confidence JOL (i.e. 0%) and distinguish it from all others. More specifically, it should be a cue-driven evaluation as compared to a target-based judgment. If this holds, we would expect participants to refer to these different types of evidence in their justifications and to observe a qualitative

متن کامل مقاله

دریافت فوری ←

ISIArticles

مرجع مقالات تخصصی ایران

- ✓ امکان دانلود نسخه تمام متن مقالات انگلیسی
- ✓ امکان دانلود نسخه ترجمه شده مقالات
- ✓ پذیرش سفارش ترجمه تخصصی
- ✓ امکان جستجو در آرشیو جامعی از صدها موضوع و هزاران مقاله
- ✓ امکان دانلود رایگان ۲ صفحه اول هر مقاله
- ✓ امکان پرداخت اینترنتی با کلیه کارت های عضو شتاب
- ✓ دانلود فوری مقاله پس از پرداخت آنلاین
- ✓ پشتیبانی کامل خرید با بهره مندی از سیستم هوشمند رهگیری سفارشات