Can working memory be non-conscious?

Timo Stein, Daniel Kaiser, and Guido Hesselmann

– Supplementary material –

Two studies on non-conscious WM computed measures of "detection sensitivity" for the memory cue to demonstrate that objectively assessed awareness of the memory cue did not differ from chance and that no correlation existed between detection sensitivity for the memory cue and memory test performance (Dutta et al., 2014; Soto et al., 2011). Although these analyses are described as being derived from signal detection theory (SDT) and the resulting indices for "detection sensitivity" are labeled "*d'*" and "*A'*", thus borrowing the terminology from SDT, these analyses are, in fact, not grounded in SDT and do not represent valid, response criterion-free measures of detection sensitivity. In the following, we first briefly describe the SDT approach for measuring detection sensitivity and contrast this with the computation of detection sensitivity in these two studies on non-conscious WM. We then show that the uncommon way of estimating detection sensitivity in these studies is not invariant to bias, but varies as a function of the observer's decision criterion, and underestimates true detection sensitivity for relatively conservative observers.

In SDT difficult perceptual decisions are modeled by assuming two overlapping distributions representing noise and signal-plus-noise (e.g., Macmillan & Creelman, 2005). These distributions form a continuum of evidence along a decision axis on which observers set a criterion to respond "signal absent" (or "no stimulus visibility", e.g., rating "1" on a 4-point scale, as in Soto et al., 2011) or "signal present" (or a rating indicating some stimulus visibility, e.g., ratings "2", "3", and "4" on a 4-point scale, as in Soto et al., 2011). Because the distributions overlap, errors occur, resulting in four types of responses in the case of a simple present-absent experiment: Hits, misses, false alarms, and correct rejections (see Table S1). For the studies on non-conscious WM a standard SDT analysis of cue sensitivity would calculate the hit rate as the proportion of cue-present trials that were categorized as such (e.g., the proportion of cue-present trials with visibility ratings 2–4 in Soto et al., 2011), and the false alarm rate as the proportion of cue-absent trials that were incorrectly categorized as cue-present trials (e.g., the proportion of cue-absent trials with visibility

ratings 2–4 in Soto et al., 2011). Thus, response frequencies from a 2 × 2 contingency table (such as Table S1) are required for SDT analysis. The difference between the z-transformed hit and false alarm rates yields a measure of the separation between the noise and the signal-plus-noise distribution in units of their standard deviation. In the studies on non-conscious WM the sensitivity measure *d'* computed in this way would have provided a measure of how well observers were able to discriminate between the absence and the presence of the memory cue. This sensitivity measure *d'* has the attractive property that it is unchanged by response bias, meaning that *d'* remains the same independent of where on the decision axis the observer is setting the criterion.

The studies on non-conscious WM did not follow the SDT framework to compute their measure of detection sensitivity. Rather, only a subset of the frequency data necessary to compute *d'* was included in their analyses. To illustrate: In these studies (Dutta et al., 2014; Soto et al., 2011), the memory cue was present in 50% of the trials and absent in the other 50% of the trials. For the computation of the sensitivity index only those trials in which participants indicated no subjective awareness of the memory cue were included (trials with the rating "1", subjectively unaware trials). This selection of a subset of trials contingent on the participants' responses is incompatible with SDT analysis. For these subjectively unaware trials, the proportion of trials in which the memory cue was present (in SDT terms, misses) was compared to the proportion of trials in which the memory cue was absent (in SDT terms, correct rejections). Thus, only the frequencies of misses and correct rejections were considered, and their frequencies relative to the total number of subjectively unaware trials were taken as the "miss rate" and the "correct rejection rate", respectively.[1]. The "miss rate" and the "correct rejection rate" therefore always added up to 1 (see Table S1). The z-transformed "miss rate" was then subtracted from the z-transformed "correct rejection rate" to yield an index the authors labeled "*d'*". SDT, however, requires not only the frequencies of misses and correct rejections but also the frequencies of hits and false alarms to compute actual miss and correct rejection rates, and

[1] The authors adopted a different terminology, labeling trials in which the memory cue was present "false alarms" rather than misses and trials in which the memory cue was absent "hits" rather than correct rejections. Here we decided to use the standard SDT terminology.

to derive the actual "*d*'" score. Thus, the analyses of detection sensitivity in these papers on non-conscious WM is not derived from SDT, and the resulting index that the authors labeled "*d*'" is unrelated to the specific meaning of the sensitivity measure *d'* in SDT.

	Response present (Visibility ratings 2-4)	Response absent (Visibility rating 1)	
Stimulus present	Hits (H)	Misses (M)	M rate in SDT = $M/(M+H)$ M rate in non-consc. WM = $M/(M+CR)$
Stimulus absent	False Alarms (FA)	Correct Rejections (CR)	CR rate in SDT = $CR/(CR+FA)$ CR rate in non-consc. WM = $CR/(CR+M)$

Table S1. A 2 × 2 contingency table for standard SDT analysis, illustrating the calculation of miss and correct rejection rates in SDT vs. the two studies on non-conscious WM (Dutta et al., 2014; Soto et al., 2011). Note that the analysis in the two studies on non-conscious WM considers only the frequencies of misses and correct rejections, although SDT requires hit and false alarm frequencies to compute miss and correct rejection rates.

However, the fact that the new index – pseudo-*d'* – used in these papers is not derived from SDT does not imply that the new index is invalid for measuring detection sensitivity for the memory cue. To test whether this newly introduced pseudo-*d'* represents a valid, bias-free measure of stimulus detectability, we computed pseudo-*d'* for varying response criteria at fixed values of the SDT index for sensitivity *d'*. The response criterion *c* is computed as −0.5 × (z-transformed hit rate + z-transformed false alarm rate), such that negative values reflect a more liberal criterion (tendency to respond "stimulus present") and positive values reflect a more conservative criterion (tendency to respond "stimulus absent"). When *c* equals 0, pseudo-*d'* equals *d'.* This is because when *c* equals 0 the hit rate is equal to the false alarm rate, such that the omission of the hit rate and the false alarm rate in the calculation of pseudo-*d'* is of no consequence. However, as can be seen in Figure 2, whenever *c* is smaller or larger than 0, pseudo-*d'* deviates from *d'*. This means that pseudo-*d'* is not independent of the location of an observer's response criterion, but varies as a function of response biases. Thus, pseudo-*d'* does not represent a bias-free measure of

detection sensitivity. For negative values of *c*, pseudo-*d'* is larger than *d'*, and for positive values of *c*, pseudo-*d'* is smaller than *d'*. For example, for a true *d'* of 2, pseudo-*d'* could vary between a value below 0.25 and values above 4, depending on the observer's response criterion. Although the two studies on non-conscious WM report no response criteria, it is possible that observers adopt a relatively conservative response criterion in difficult perceptual tasks such as those used in these studies (e.g., reflecting underconfidence, see Björkmann, Juslin, & Winman, 1993). In this case, pseudo-*d'* would have underestimated true sensitivity. In summary, our analysis demonstrates that the new index used in these studies does not represent a valid, bias-free measure of detection sensitivity and thus cannot be used to provide evidence for non-conscious WM. Future studies need to adopt the standard SDT framework to calculate objective, bias-free measures of detection sensitivity.

References

Björkmann, M., Juslin, P., & Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. *Perception & Psychophysics, 54*, 75–81.

Dutta, A., Shah, K., Silvanto, J., & Soto, D. (2014). Neural basis of non-conscious visual working memory. *Neuroimage, 91*, 336–343.

Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). Mahwah (NJ): Erlbaum.

Soto, D., Mäntylä, T., & Silvanto, J. (2011). Working memory without consciousness. *Current Biology, 21*, R912–R913.