Behavior Research Methods
https://doi.org/10.3758/s13428-019-01295-1

Comment on "Quantifying the informational value of classification images": A miscomputation of the *infoVal* metric

Mathias Schmitz[1] · Marine Rougier[1] · Vincent Yzerbyt[1]

Abstract

Brinkman et al. (2019) recently introduced an innovative metric—*infoVal*—to assess the informational value of classification images (CIs) relative to a random distribution. Although this measure constitutes a valuable tool to distinguish random from nonrandom CIs, we identified two noteworthy discrepancies between the mathematical formalization of the *infoVal* metric and the authors' computation. Specifically, the computation was based on the *one norm* instead of the *Euclidean norm*, and the *k* constant was omitted in the denominator of the ratio that produces *infoVal*. Accordingly, the simulations and experimental results reported by Brinkman et al. do not build on the correct *infoVal* computation but on a biased index. Importantly, this discrepancy in the computation affects the statistical power and Type I and error rate of the metric. Here we clarify the nature of the discrepancies in the computation and run Brinkman et al.'s Simulation 1 anew with the correct values, to illustrate their consequences. Overall, we found that relying on the miscomputed *infoVal* metric can lead to misguided conclusions, and we urge researchers to use the correct values.

Keywords *infoVal* · Reverse correlation · Classification images · rcicr package · R

In their recent article "Quantifying the Informational Value of Classification Images," Brinkman et al. (2019) introduced an objective informational value metric—*infoVal*—as a means to statistically assess the degree to which signal rather than noise is present in facial representations (i.e., classification images or CIs) obtained from reverse-correlation (RC) experiments (see Dotsch, Wigboldus, Langner, & van Knippenberg, 2008; Mangini & Biederman, 2004). Indeed, as was argued by the authors, one of the main caveats of RC experiments is that they will always yield a CI, whether that image builds on random or on meaningful responses. Researchers or external judges may thus erroneously interpret the CI as containing signal when there is in fact none. In light of this limitation, this innovative measure aims to distinguish CIs containing noise from those containing signal. To ascertain the quality of their new metric, Brinkman et al. thoroughly simulated and empirically tested and validated the internal and convergent validity of *infoVal*. On the basis of their findings, they recommend the use of *infoVal* scores as a new standard of good practice for RC studies.

As much as *infoVal* may prove to be a valuable tool, closer inspection of the manuscript led us to notice that the computation of *infoVal* in the R code available from the authors (Brinkman et al., 2019; see https://osf.io/v3y5e/) differed from the metric's formal conceptualization in two ways. To be in a position to gauge the impact of the discrepancies between the "correct" and "incorrect" computations of the *infoVal* metric, we first reran the Simulation 1 proposed by Brinkman et al. Concretely, we independently wrote an R script building upon the information provided in Brinkman et al.'s manuscript.[1] Note that although we restricted the present analysis to their Simulation 1, the impact of these discrepancies on the other analyses presented by Brinkman et al. should be similar.

In this simulation, participants' responses to a typical RC task are simulated in such a way that they vary from completely random (i.e., pure noise) to completely accurate (i.e., pure signal). As can be seen in Table 1 and Fig. 1, the data obtained

Electronic supplementary material The online version of this article (https://doi.org/10.3758/s13428-019-01295-1) contains supplementary material, which is available to authorized users.

✉ Mathias Schmitz
mathias.schmitz@uclouvain.be

[1] Université catholique de Louvain, Louvain-la-Neuve, Belgium

[1] All our computations were performed with R version 3.5.3 (R Core Team, 2015). The R scripts can also be run with Microsoft R Open (Version 3.5.3, Microsoft & R Core Team, 2017; see https://mran.microsoft.com/) for improved performance. The noise pattern (i.e., the noise matrix) was generated with the developmental version of the rcicr package (Dotsch, 2017).

Published online: 25 October 2019

Table 1 Proportions of H_0 rejections (H_0 = CI generated by a random process) out of 1,000 simulation runs per level of *P*[random] for two different critical values (1.96 and 3) and for the "correct" (Euclidean norm and $k = 1.4826$) versus "incorrect" (one norm and $k = 1$) computation of the *infoVal* metric

P[random]	*infoVal* > 1.96		*infoVal* > 3	
	Correct Computation	Incorrect Computation	Correct Computation	Incorrect Computation
.000	1.000	1.000	1.000	1.000
.125	1.000	1.000	1.000	1.000
.250	1.000	1.000	1.000	1.000
.375	1.000	1.000	1.000	1.000
.500	1.000	1.000	.983	1.000
.625	.878	.970	.543	.866
.750	.302	.553	.060	.282
.875	.032	.125	.002	.028
1.000	.006	.041	.000	.005

from our simulation match within rounding errors those presented by Brinkman et al. (2019; see their Fig. 3 and Table 1). Our data, R code, and materials are available on the Open Science Framework (see the Open Practices Statement section).

According to Brinkman et al. (2019), the *infoVal* metric can be interpreted as a *z* score. It compares the observed CI's vector length to a reference distribution of simulated vector lengths based on random responses. The vector length, or norm, is defined as the squared root of the sum of squares (i.e., the *Euclidean norm*) over all the pixels of the CI ($\overline{p_i}$), as in

$$x = \sqrt{\sum_{i=1}^{I} \overline{p_i}^2} \quad (1)$$

As for the *infoVal* metric, Brinkman et al. define it as

$$infoVal = \frac{x_{\text{obs}} - \widetilde{x}_{\text{sim}}}{\sigma_{\text{sim}}} \quad (2)$$

where x_{obs} is the vector length of the empirically observed CI, $\widetilde{x}_{\text{sim}}$ is the median of the lengths of the simulated vectors, and σ_{sim} is the approximated standard deviation of the lengths of the simulated vectors, as in

$$\sigma_{\text{sim}} = k \cdot MAD_{\widetilde{x}_{\text{sim}}} \quad (3)$$

where k is a fixed constant equal to 1.4826, and $MAD_{\widetilde{x}_{\text{sim}}}$ is the median absolute deviation of the lengths of the simulated vectors.

The aforementioned formal definition of *infoVal*, as presented in Brinkman et al. (2019), differs from the actual computation in two remarkable ways. First, although the vector lengths, or norms, of the observed or simulated CIs are conceptualized as the Euclidean norm [i.e., `norm(x, type = "F")` in R; see Eq. 1], they were in fact computed as the *one norm* [i.e., `norm(x)` or, equivalently, `norm(x, type = "O")` in R]. The difference here is that the *one norm* is computed as the sum of the absolute values of each pixel from the CI, whereas the Euclidean norm is computed as the square root of the sum of squared CI's pixels.[2] Although the centrality (e.g., mean or median) and dispersion (e.g., standard deviation or median absolute deviation) of the distributions of empirical or simulated CIs' norms are relatively larger when the one norm rather than the Euclidean norm is computed, both distributions are highly correlated ($r = .995$). As a consequence, the effect of relying on one or the other norm—that is, the *Euclidean* versus the *one norm*—should be very small if not inconsequential for the *infoVal* metric (see the supplementary materials for the norm distributions and the effect of the computed norm on *infoVal*).

A second discrepancy is that the standard $k = 1.4826$ constant was omitted in Brinkman et al.'s (2019) computation (see Eq. 3). Given that k is introduced as a denominator of the difference between the observed norm (x_{obs}) and the median of the simulated CIs' norms ($\widetilde{x}_{\text{sim}}$), the value of *infoVal* computed with k should be 32.55% times smaller than the one computed without k (or, equivalently, with $k = 1$). Accordingly, there should be a loss of statistical power but also a reduction in Type I error rate when using $k = 1.4826$. Also, because *infoVal* is close to zero when the CI contains no or very little signal, and grows rapidly as the signal rate increases, the loss of statistical power should be more pronounced than the reduction of Type I error rate. Using the

[2] For more information about the computation of the norm in R, see https://www.rdocumentation.org/packages/base/versions/3.6.0/topics/norm.

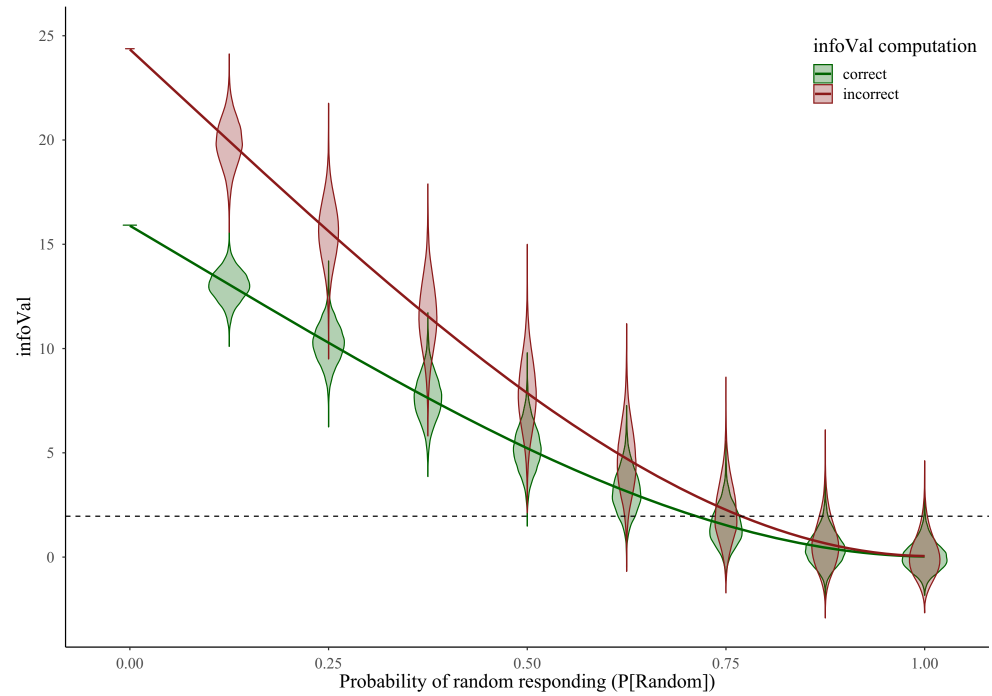

Fig. 1 Relationship between the *infoVal* metric and the probability of random responses, *P*[random], as a function of the "correct" (Euclidean norm and $k = 1.4826$) versus "incorrect" (one norm and $k = 1$) computation of this metric. The dashed line represents 1.96 units of approximated standard deviation.

same logic, the *k* constant should also narrow the spread of the *infoVal* distribution by 32.55% (see the supplementary materials for an illustration of the effect of *k* on the *infoVal* metric).

To visualize the effect of these two discrepancies in the computation of the *infoVal* metric, we compared the "correct" computation of *infoVal* (i.e., as formalized by Brinkman et al., 2019; i.e., with the Euclidean norm and with $k = 1.4826$) to the "incorrect" computation (i.e., as computed by Brinkman et al., 2019; i.e., with the *one norm* and $k = 1$). We expected the effects of both miscomputations to be additive while mainly being driven by the difference in the *k* constant. The results from our Simulation 1 are illustrated in Fig. 1 (see also Brinkman et al., 2019, Fig. 3) and Table 1 (see also Brinkman et al., 2019, Table 1).

Visual inspection of Fig. 1 confirms that the "correct" computation of *infoVal* is less powerful than the "incorrect" computation, especially as the level of *P*[random] responses approximates 0% (i.e., as the signal rate increases). At the same time, the statistical power of *infoVal* remains at 100% in both cases for *P*[random] ≤ 50% (see Table 1). The plot (see Fig. 1) and the actual statistics (see Table 1) also suggest that the "correct" computation has a smaller Type I error rate for *P*[random] responses approaching 100%, as one would expect. Finally, we observe that the dispersion is lower for the "correct" *infoVal* distributions than for the "incorrect" distributions, which may yield more consistent *infoVal* values. This is an important result, given that more consistent statistical tests could positively contribute to the replicability issue in psychological sciences (see Simmons, Nelson, & Simonsohn, 2011).

In sum, our examination of the discrepancies between the formalization and the actual computation of Simulation 1 proposed by Brinkman et al. (2019) shows that the "correct" computation of *infoVal* is more conservative than its miscomputation. The bias stems mainly from the omission of the *k* constant. The "correct" computation of this metric is therefore relatively less powerful than expected, while also having a lower Type I error (false positive) rate than the "incorrect" computation. The miscomputation of *infoVal* has direct implications for researchers, since they might erroneously discard meaningful CIs (i.e., Type II errors) or read signal in CIs when there is fact none (i.e., Type I errors), leading them to misguided conclusions.

In spite of the fact that the present commentary points out issues with the way Brinkman et al. (2019) implemented their measure, we converge with them in stressing the importance of such a measure to assess the informational value of CIs. We thus recommend that

[3] The first author of this comment (M.S.) submitted a pull request on the GitHub repository for the rcicr package from Dotsch (2017) to correct for computation of the *infoVal* metric. The request was approved by Ron Dotsch (see https://github.com/rdotsch/rcicr/issues/96, and https://github.com/rdotsch/rcicr/pull/97). For all practical purposes, researchers can thus use the developmental version of the rcicr R package (available at https://github.com/rdotsch/rcicr/) for the "correct" computation of the *infoVal* metric.

future work on reverse-correlation experiments rely on the "correct" computation of this metric, as presented formally by Brinkman et al.[3]

Acknowledgements This work was supported by the Fonds de la Recherche Scientifique (FNRS), grant 1.A393.17, awarded to M.S., and by FNRS grant 1.B347.19, awarded to M.R.

Open Practices Statement The data, R code, and material are publicly available on Open Science Framework (https://osf.io/fbduk/?view_only=e3c04e549f844facb109200dcba7a0db).

References

Brinkman, L., Goffin, S., van de Schoot, R., van Haren, N. E. M., Dotsch, R., & Aarts, H. (2019). Quantifying the informational value of classification images. *Behavior Research Methods*. Advance online publication. doi:https://doi.org/10.3758/s13428-019-01232-2

Dotsch, R. (2017). rcicr: Reverse-correlation image-classification toolbox (R package version 0.4.0). Retrieved from https://rdrr.io/cran/rcicr/

Dotsch, R., Wigboldus, D. H. J., Langner, O., & van Knippenberg, A. (2008). Ethnic out-group faces are biased in the prejudiced mind. *Psychological Science*, *19*, 978–980. doi:https://doi.org/10.1111/j.1467-9280.2008.02186.x

Mangini, M., & Biederman, I. (2004). Making the ineffable explicit: Estimating the information employed for face classifications. *Cognitive Science*, *28*, 209–226. doi:https://doi.org/10.1016/j.cogsci.2003.11.004

Microsoft & R Core Team. (2017). Microsoft R Open (Version 3.5.3). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://mran.microsoft.com/

R Core Team. (2015). R: A language and environment for statistical computing (Version 3.5.3). Vienna, Austria: R Foundation for Statistical Computing. Retrieved from www.R-project.org

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22* 1359–1366. doi:https://doi.org/10.1177/0956797611417632