COMPUTATION

The Effect of Working Memory Capacity Limitations on the Intuitive Assessment of Correlation: Amplification, Attenuation, or Both?

Sorel Cahan and Yaniv Mor
The Hebrew University of Jerusalem

This article challenges Yaakov Kareev’s (1995a, 2000) argument regarding the positive bias of intuitive correlation estimates due to working memory capacity limitations and its adaptive value. The authors show that, under narrow window theory’s primacy effect assumption, there is a considerable between-individual variability of the effects of capacity limitations on the intuitive assessment of correlation, in terms of both sign and magnitude: Limited capacity acts as an amplifier for some individuals and as a silencer for others. Furthermore, the average amount of attenuation exceeds the average amount of amplification, and the more so, the smaller the capacity. Implications regarding the applicability and contribution of the bias notion in this context and the evaluation of the adaptive value of capacity limitations are discussed.

Keywords: correlation coefficient, sampling distribution, intuitive assessment, bias, narrow window theory

According to narrow window theory (NWT), introduced by Yaakov Kareev 10 years ago (Kareev, 1995b; see also Kareev, 1995a, 2000, 2003, 2005; Kareev, Lieberman, & Lev, 1997), intuitive estimates of population correlations overestimate population parameters. This amplification effect (Kareev, 1995b) is due to the small samples (7 ± 2 items) on which people rely in the calculation of intuitive estimates due to the limited capacity of working memory (WM). For this magnitude of sample size, “sample-based measures of correlation provide a [positively] biased estimate of the population parameter” (Kareev, 2000, p. 398). Furthermore, “...the bias is stronger when the sample is smaller ... The bias strength is substantial for samples of the size likely to be considered by humans” (Kareev, 2003, p. 263).

Moreover, Kareev (2000) has raised the counterintuitive and intriguing argument (Anderson, Doherty, Berg, & Friedrich, 2005; Juslin & Olsson, 2005) that capacity limitation has an adaptive advantage. By increasing the chances of overestimating the population parameter, capacity limitations “maximize the chances for the early detection of strong and useful relationships” (Kareev, 2000, p. 397). Furthermore, according to Kareev, the provision of an amplified picture of the strength of the relationships between variables in the environment may well be the very developmental reason for the severe capacity limitations of the human WM: “By providing such a biased picture, capacity limitations may have evolved so as to protect organisms from missing strong correlations and to help them handle the daunting tasks of induction” (Kareev, 2000, p. 401).

NWT’s original publications (Kareev, 1995a, 1995b), as well as later publications (e.g., Kareev, 2000, 2003, 2005), offer no formal definition and measure of the bias of the sample r as an estimator of the population correlation ρ (ρ). In fact, the only justification provided in support of the positive bias claim consists of the comparison between the median and the mode of the sampling distribution of r, on the one hand, and ρ, on the other hand, for selected values of ρ and sample size n. For example, for ρ = 0.60 and n = 4, the median of the distribution of r is .63 and its mode is around .75, whereas with n = 4, the median is .69, and the mode is around .97 (Kareev, 1995b, p. 266).

However, the numerical values of the median and the mode of the sampling distribution of a statistic (e.g., r) and the difference between them, on the one hand, and the corresponding parameter (e.g., ρ), on the other hand, are not known and accepted definitions or measures of the sampling and estimation bias of the statistic. Furthermore, the differences “median (r) – ρ” and “mode (r) – ρ” are also incompatible with NWT’s probabilistic formulation of the positive bias of r. For example, “people are more likely [italics added] to encounter a sample indicating a more extreme relationship than the actual relationship in the population” (Kareev, 1995b, pp. 265–266) or “more often than not” (Kareev, 2005, p. 280). Neither the difference “median (r) – ρ” nor “mode (r) – ρ” can be meaningfully interpreted in probabilistic terms.

The lack of a formal definition of bias is particularly critical in view of the fact that NWT’s positive bias argument contradicts the well-documented statistical fact that the sample r is negatively biased, and the more so, the smaller the sample (e.g., Soper, Young, Cave, Lee, & Pearson, 1917). Hence, if NWT’s argument is to be validly interpreted in terms of positive bias, the implicit definition of bias underlying it has to differ from the standard one in statistics. The purpose of this article is twofold:

1. To specify the implicit statistical definition of the bias underlying NWT’s argument regarding the positive bias of r in small samples, to point to the difference between it and the standard
statistical definition, and to explain the contradictory conclusions regarding the bias of $r$ based on them.

2. To discuss the psychological implications of the bias of $r$ regarding the effects of limited WM capacity on the intuitive assessment of correlation, to discuss the applicability and contribution of the bias notion in this context, and to discuss the adaptive value of capacity limitations.

For the sake of simplicity and clarity of presentation, throughout the presentation we shall relate only to positive $\rho$ values.

A Statistical Analysis of the Sampling Error Bias of $r$ in Small Samples

Bias is an attribute of errors. Errors are considered to be biased (i.e., systematic) in a given population or subpopulation if and only if their mean differs from zero. This definition of bias applies to measurement error—the difference between the observed and the true scores (e.g., Feldt & Breman, 1989; Lord & Novick, 1968; Stanley, 1971)—as well as to prediction error, that is, the difference between predicted and criterion scores (e.g., Cleary, 1968). In both cases, bias is positive if and only if the mean difference is positive and negative if and only if the mean difference is negative.

The Bias of the Sampling Error ($e$) of $r$

The same definition of bias also applies to sampling errors, that is, to the difference between the sample statistic and the population parameter (e.g., Mood, Graybill, & Boes, 1987). In the particular case of $r$, for any given sample size $n$, the bias of the sampling error of $r$ ($e = r - \rho$) is defined as follows:

$$E(e) = E(r - \rho) = E(r) - \rho. \quad (1)$$

As pointed out almost 100 years ago (Soper, 1913) and repeatedly restated since (e.g., Fisher, 1915; Hotelling, 1953; Soper et al., 1917), $e$ is negatively biased, that is, $E(r) < \rho$, and the more so, the smaller $n$ and the stronger $\rho$. Graph A in Figure 1 (based on Soper et al.'s (1917) study, and (B) narrow window theory’s dichotomous sampling error $\epsilon$, $E(\epsilon)$, which is based on David’s (1954) study, as a function of sample size ($n$) for selected positive $\rho$ values.

Figure 1. (A) The bias of the sampling error $e = r - \rho$, $E(e)$, which is based on Soper et al.’s (1917) study, and (B) narrow window theory’s dichotomous sampling error $\epsilon$, $E(\epsilon)$, which is based on David’s (1954) study, as a function of sample size ($n$) for selected positive $\rho$ values.

(Kareev, 1995a, 1995b) nor in latter publications (e.g., Kareev, 2000, 2003, 2005). A third possibility is that the error underlying the positive bias argument is not the sampling error $e = r - \rho$, but some other (still unidentified) error. We suggest that this is, indeed, the case. Consequently, the valid interpretation of NWT’s argument in terms of positive bias (which is explicitly included in the very title of Kareev’s, 1995a, article as well as in latter publications) necessarily requires the specification of this error.

The Dichotomous Sampling Error ($\epsilon$) of $r$ and Its Bias

We submit that the conceptualization of NWT’s amplification effect in terms of positive bias implicitly relies on a dichotomous function of the original sampling error $e$.

$$\epsilon = g(e); \epsilon = +0.50 \text{ if } e > 0, \text{ and } \epsilon = -0.50 \text{ if } e < 0. \quad (2)$$

where epsilon ($\epsilon$) distinguishes only between positive and negative sampling errors, assumed to be of equal magnitude and assigned the numerical values $+0.5$ and $-0.5$, respectively; that is, $\epsilon$ is a
symmetrical directional sampling error.\textsuperscript{1} In line with the general definition, the bias of $\varepsilon$ is defined as follows:

$$E(\varepsilon) = -0.50 \times P(\varepsilon < 0) + 0.50 \times P(\varepsilon > 0) = 0.50 \times [P(\varepsilon > 0) - P(\varepsilon < 0)].$$

That is, $\varepsilon$ is unbiased [$E(\varepsilon) = 0$] if and only if $P(\varepsilon < 0) = P(\varepsilon > 0)$. An alternative definition of $E(\varepsilon)$, which directly allows for its computation on the basis of the statistical tables of the sampling distribution of $r$ (e.g., David, 1954), is given by Equation 4:

$$E(\varepsilon) = 0.50 - Fr(p),$$

where

1. $Fr(p) = P(\varepsilon < 0)$ is the cumulative relative frequency of $p$ in the distribution of $r$ for sample size $n$ and parameter $p$ [i.e., $P(r \leq p)$], and
2. $0.50$ is the value of $Fr(p)$ when $\varepsilon$ is unbiased [i.e., $E(\varepsilon) = 0$].

We submit that $E(\varepsilon)$ is the (missing) implicit definition and measure of NWT’s amplification effect in terms of positive bias. According to Equation 4, $E(\varepsilon)$ can be meaningfully interpreted as a proportion: the proportion of samples in which $r > p$ that is added to 0.50 because of small sample size. This interpretation is consistent with the probabilistic formulation of NWT’s positive bias effect (e.g., “more often than not”; Kareev, 2005, p. 280). Graph B in Figure 1, based on David’s (1954) tables according to Equation 4, gives the size of $E(\varepsilon)$ as a function of $n$ for selected positive values of $p$. As illustrated by the diagram, $E(\varepsilon)$ is positive (congruent with NWT’s assertion) albeit small. For example, for $p = 0.50$ and $5 \leq n \leq 9$, $E(\varepsilon)$ ranges between 0.04 and 0.06. Furthermore, $E(\varepsilon)$ is larger, the smaller the sample size (in accordance with NWT) and the stronger $p$.

It is interesting to note that, unlike $E(\varepsilon)$, $E(\varepsilon)$ cannot be interpreted in terms of the magnitude relation between any central tendency measure of the sampling distribution of $r$ and $p$. The fact that, for small samples, both median($r$) and mode($r$) are more extreme than $p$—offered as evidence for the amplification effect (Kareev, 1995b, p. 266)—is a result of $E(\varepsilon) > 0$ rather than its cause: Whereas [median ($r$) $- p$] $> 0$ indicates the existence of a positive bias of $\varepsilon$, it does not define or measure it.

\textit{e Versus $\varepsilon$}

To conclude, we suggest that the contradiction between NWT’s main statistical argument—the positive bias of the sample $r$ for small samples—and the well-established statistical wisdom—according to which $r$ is negatively biased, and the more so, the smaller the sample—is entirely attributable to the difference between the sampling errors on which they rely, $e$ and $\varepsilon$, respectively. This contradiction is, in fact, a part-whole contradiction, one between the conclusion based on one component of the sampling error $e$—its sign, which underlies the definition of $e$—and the conclusion based on the sampling error $e$ itself (which consists of both sign and magnitude). The negative $E(\varepsilon)$ values for small samples reflect the higher average magnitude of the negative errors, which compensate (in fact, overcompensate) for their lower relative frequency. In contrast, $E(\varepsilon)$ is positive because $\varepsilon$ ignores the difference between positive and negative errors in terms of average magnitude. The sampling distribution of $r$ for small samples can thus be validly described as being both positively biased [in the sense of $E(\varepsilon) > 0$] and negatively biased [in the sense of $E(\varepsilon) < 0$].

**Psychological Implications**

**The Biasing Effect of WM Capacity Limitations on the Intuitive Assessment of Correlation: Positive or Negative?**

The main implication of the statistical analysis in the previous section is that the formulation and interpretation of NWT’s argument regarding the effect of limited WM capacity on the intuitive assessment of correlation in terms of positive bias is valid only for a dichotomous definition of error. Clearly, however, estimation error is not only a matter of sign. Magnitude matters. The psychological result of encountering a sample $r$ that deviates $-0.30$ from $p$ is likely to be different from that of encountering a sample $r$ that deviates $-0.01$ from $p$. For example, the detection of a strong and useful regularity (e.g., $p = 0.50$) will be strongly affected by a large negative sampling error (e.g., $r = -0.20$), but entirely unaffected by a small, same sign error (e.g., $r = .49$). In contrast, it can be safely assumed that the intuitive estimation of $p$ will not be affected by small sampling errors of opposite sign (e.g., $r = .49$ and $r = .51$). A valid psychological theory of sample-based intuitive correlation assessment should consider both the sign and the magnitude of the sampling error of $r$ in small samples.

Hence, the definition of bias in terms of $e$—rather than $\varepsilon$—is preferable not only on formal grounds, but also from substantive considerations. Outwardly, adoption of this definition leads to inversion of NWT’s main argument: Reliance on small samples due to capacity limitations results in negative, rather than positive, bias in correlation assessment. Nevertheless, we suggest that this conclusion, necessarily based on the average error, gives a misleading picture of the psychological effects of WM capacity limitations on the intuitive assessment of correlation. A conclusion based on error averaging [either $E(\varepsilon)$ or $E(\varepsilon)$] would have been psychologically meaningful had the intuitive assessment of correlation by each individual been based on the average of the within-individual distribution of the sample $r$ values between the same pair of variables, $X$ and $Y$ (e.g., height and weight) observed by him or her. In this case, the (contradictory) psychological implications of $E(\varepsilon) > 0$ and $E(\varepsilon) < 0$ would have been universal: According to NWT, intuitive assessment of any $p$ by each and every individual is positively biased; according to our own analysis, bias is negative for each and every individual.

However, despite the appeal of this possibility, the within-individual distribution of sample $r$s is not involved, according to NWT, in the intuitive estimation of $p$. Rather, the theory specifi---

\textsuperscript{1} The symmetrical numerical values that can be assigned to positive and negative errors are arbitrary (e.g., $+1$ and $-1$ or $+8$ and $-8$). Note, however, that the numerical value of $E(\varepsilon)$ is specific to these arbitrary numbers and, therefore, meaningless. A metric-free measure of bias is obtained by standardizing $\varepsilon$ by its range:

$$\varepsilon' = \frac{\varepsilon}{\varepsilon_{\text{max}} - \varepsilon_{\text{min}}}$$

where $\varepsilon_{\text{max}} - \varepsilon_{\text{min}}$ are the (symmetrical) arbitrary numerical values assigned to positive and negative errors, respectively. As a result of standardization, $\varepsilon'$ is independent of the arbitrary metric of $\varepsilon$: The same numerical $\varepsilon'$ values, $+0.5$ and $-0.5$, will obtain for any symmetrical $\varepsilon_{\text{min}}$ and $\varepsilon_{\text{max}}$ values.
cally assumes that each individual’s intuitive assessment of ρ exclusively relies on the computed in the first sample encountered and is not affected by subsequent samples (e.g., Kareev, 1995b, p. 264). Hence, the sampling distribution of r for small samples is not used by NWT to model the psychological process underlying the intuitive correlation estimation of each and every individual—which, as previously mentioned, is assumed to rely on a single r value—only the across-individuals distribution of its results: the distribution of the rs between the same X and Y computed by all the individuals with a given WM capacity on the basis of the first random sample encountered by each of them. Hence, any bias argument [either E(ε) > 0 or E(ε) < 0] can only apply, in the framework of NWT, to this distribution and necessarily involves averaging sampling errors across all the individuals with the same WM capacity in the population.

Average Versus Differential Effects

Averaging errors across individuals is problematic from two perspectives. First, it has no empirical counterpart: No population-level estimate based on averaging rs across individuals is typically computed. Second, it ignores the between-individual variability of e, in terms of both sign and magnitude. Because, according to NWT, each individual estimate is based on a single sample, different individuals with the same WM capacity will encounter samples with different sampling errors: some positive and some negative, some large and some small. Consequently, the intuitive estimates of the same ρ by different individuals will differ both in the direction and the magnitude of the estimation error they include: Some individuals will overestimate ρ (by different amounts), whereas others will underestimate it (by different amounts). A realistic and valid psychological theory of intuitive correlation assessment should acknowledge these differential effects of reliance on small samples due to WM capacity limitations rather than focus on their population means.

Furthermore, measurement of these effects requires a clear conceptual distinction between (a) the effect of the unavoidable reliance on samples in the intuitive assessment of correlations and (b) the net effect of the small size of these samples, due to capacity limitations, per se—that is, the net effect of WM capacity limitations. As a result of the former, intuitive estimates of ρ (based on the first sample encountered) will necessarily include sampling error—albeit relatively small—even for large samples (e.g., 50, 100, or 400): ≈50% of the individual estimates will be greater than ρ (i.e., overestimates), and ≈50% lower than ρ (i.e., underestimates; Fisher, 1973, p. 194). Reliance on small (rather than large) samples, due to WM capacity limitations (i.e., seeing the world through a narrow window; [b] above), affects the intuitive assessment of correlation in two ways:

1. By increasing the proportion of individuals who overestimate ρ, over and above the 50% base rate for large samples (and concomitantly decreasing the proportion of individuals who underestimate ρ), congruent with NWT’s main assertion. This net effect of WM capacity limitations is validly measured by E(ε)—Graph B in Figure 1. For most combinations of n and ρ, the increase is modest (Mdn = 4%).

2. By increasing both the average amount of overestimation of ρ [by the 0.50 + E(ε)] individuals who encounter small samples in which r > ρ] and the average (absolute) amount of underestimation of ρ [by the 0.50 − E(ε)] individuals who encounter small samples in which r < ρ] —E(ε > 0), and E(ε < 0), respectively—relative to their large sample (N) values, E(ε > 0), and E(ε < 0), respectively. We refer to these effects as the amplification and attenuation effects of limited WM capacity, per se, Δ(ε−) and Δ(ε+), respectively:

\[ Δ(ε−) = E(ε > 0) − E(ε > 0)_{N} \]
\[ Δ(ε+) = E(ε < 0) − E(ε < 0)_{N} \]

Table 1 gives the numerical values of Δ(ε−) and Δ(ε+) as a function of n for selected ρ values, assuming N = 400 (the largest value of N in David’s, 1954, tables).

Table 1: The Magnitude of the Attenuation and Amplification Effects of Limited Working Memory Capacity [Δ(ε−) and Δ(ε+), respectively] as a Function of n for Selected ρ Values

<table>
<thead>
<tr>
<th>Working memory capacity</th>
<th>n = 3</th>
<th>n = 5</th>
<th>n = 7</th>
<th>n = 9</th>
</tr>
</thead>
<tbody>
<tr>
<td>ρ</td>
<td>Δ(ε−)</td>
<td>Δ(ε+)</td>
<td>Δ(ε−)</td>
<td>Δ(ε+)</td>
</tr>
<tr>
<td>0.1</td>
<td>−.064</td>
<td>.55</td>
<td>−.40</td>
<td>.36</td>
</tr>
<tr>
<td>0.3</td>
<td>−.69</td>
<td>.44</td>
<td>−.42</td>
<td>.30</td>
</tr>
<tr>
<td>0.5</td>
<td>−.71</td>
<td>.32</td>
<td>−.40</td>
<td>.22</td>
</tr>
<tr>
<td>0.7</td>
<td>−.65</td>
<td>.19</td>
<td>−.33</td>
<td>.14</td>
</tr>
</tbody>
</table>

Hence, contrary to Kareev’s (1995b, p. 268) unreserved claim that “. . . [limited capacity] acts as an amplifier, strengthening signals which may otherwise be too weak to be noticed,” our analysis points to a complex picture of the net effects of limited WM capacity (i.e., of the window’s “narrowness”) on the intuitive assessment of correlation, assuming exclusive reliance on the first (small) sample of the joint distribution of X and Y. Although limited capacity does indeed act as an amplifier for some individ-

COMMENTARY
uals, for other individuals it acts as a silencer, attenuating signals that may have otherwise been noticed. Furthermore, the (absolute) amount of attenuation always exceeds the amount of amplification.

The average-based bias notion obscures this between-individuals variability in the effects of WM capacity limitations and may be misinterpreted as indicating the existence of a single, fixed, universal individual-level effect. Awareness of this variability is particularly critical for the evaluation of the adaptive value of the effects of limited WM capacity on the intuitive assessment of correlation and of the limited capacity itself.

The Adaptive Value of the Effects of Capacity Limitations on Intuitive Correlation Assessment

The conclusions of our analysis are clearly inconsistent with Kareev’s (2000) unqualified positive evaluation of the adaptive value of the effects of limited WM capacity on intuitive correlation assessment as well as of the WM limitations themselves.\(^2\) The sizeable attenuation of \(r\) in the samples encountered by a substantial minority of individuals can hardly be defended as adaptive—that is, as protecting individuals “from missing strong correlations” (Kareev, 2000, p. 401). The overall evaluation of the adaptive value of limited capacity should not overlook this sizeable minority and concentrate exclusively on the small majority who encounter samples in which \(r\) is, indeed, amplified (i.e., \(r > r\)). As suggested in NWT’s original publication: “the present discussion involves the values of \(r_{XY}\) in samples most likely to be encountered [italics added]” (Kareev, 1995b, p. 266, Footnote 3). Under the assumptions of NWT (particularly the exclusive reliance on the first sample encountered), small capacity is good for some people and bad for others, and, for each individual, it is sometimes good and sometimes bad.

\(^2\)The import of our analysis to this issue is entirely independent of the recent debate regarding it (Anderson et al., 2005; Juslin & Olsson, 2005; Kareev, 2005). This controversy was based on the acceptance of NWT’s main assertion regarding the existence of a sizeable positive bias in correlation assessment. Opinion differences only regarded the methods by which the effect of this positive bias on the detection of correlation should be evaluated and interpretation of their results.

References


