Bayesian explorations in mathematical psychology

Matzke, D.

Chapter 8

Accounting for Measurement Error and the Attenuation of the Correlation: A Bayesian Approach

This chapter is currently in preparation as:
Dora Matzke, Alexander Ly, Ravi Selker, Wouter D. Weeda, Benjamin Scheibehenne, Michael D. Lee, and Eric-Jan Wagenmakers
Accounting for measurement error and the attenuation of the correlation: A Bayesian approach.

Abstract

The Pearson product-moment correlation coefficient can be severely underestimated when the observations are subject to measurement noise. Various approaches exist to correct the estimation of the correlation in the presence of measurement error, but none are routinely applied in psychological research. Here we outline a Bayesian correction method for the attenuation of correlations proposed by Behseta et al. (2009) that is conceptually straightforward and easy to apply. We illustrate the Bayesian correction with two empirical data sets; in each data set, we first estimate posterior distributions for the uncorrected and corrected correlation coefficient and then compute Bayes factors to quantify the evidence that the data provide for the presence of an association. We demonstrate that correcting for measurement error can substantially increase the correlation between noisy observations.

8.1 Introduction

The Pearson product-moment correlation coefficient is a well-known and frequently used measure to assess the linear relationship between two variables. Its popularity in psychological research is illustrated by the fact that we found that 42% of the 67 articles in the 2012 volume of the *Journal of Experimental Psychology: General* (JEP:G) report at least one Pearson correlation coefficient, with 152 correlations in total, and an average of 5.43 reported correlations per article. Also well-known, at least among statisticians, is that measurement error decreases the observed correlation between the variables (e.g., Charles, 2005; Spearman, 1904).

Although it is generally recognized that most —if not all— psychological constructs are measured only imperfectly, few researchers seem to acknowledge explicitly that the observed correlation often underestimates the true correlation between two variables. This neglect is especially puzzling because various approaches are now available to correct the correlation for the measurement error that affects the observations. Attempts to remedy the problem of the attenuation of the correlation date back to Spearman (1904), who proposed to correct the measured correlation using the reliability with which the observations were obtained. Spearman's method, however, suffers from a number of shortcomings. First, Spearman's correction formula can produce corrected correlation coefficients in excess of 1.00. Second, Spearman's method assumes homogenous error variances, an assumption that is likely to be violated in many real-life applications.

As an alternative, Behseta et al. (2009) proposed a Bayesian correction method that does not suffer from the above limitations. Contrary to Spearman's (1904) formula, the Bayesian method yields corrected correlations that are naturally bounded by -1.00 and 1.00 and is not limited to situations with homogenous error variances. Behseta et al.'s approach is conceptually straightforward and their simulations demonstrated that it is superior to Spearman's method in terms of the accuracy of the recovered corrected correlation.

Despite the availability of methods to correct the correlation coefficient for attenuation —be it Spearman's (1904) traditional method or Behseta et al.'s (2009) Bayesian approach— psychologists seldom attempt to adjust their correlations for measurement noise. Indeed, out of the 28 JEP:G articles in the 2012 volume which reported one or more correlations, only one acknowledged the deleterious effects of measurement error and corrected the observed correlation. This situation is unfortunate; as demonstrated by Behseta et al. —and as we will demonstrate again shorty— correcting the correlation for attenuation may substantially increase the association between noisy observations. Of course, correction is only possible if the magnitude of the measurement error is known. Luckily, our JEP:G literature review suggests that the uncertainty of the observations may be often estimated from the data, for example, when each observation is derived as the average of multiple trials in a repeated measures design. Specifically, we found that for 25% (38/152) of the reported correlations, the measurement error could have been estimated and corrected for. For 42% (16/38) of these correlations, correction was possible for both variables; for the remaining 58% (22/38), correction was possible for only one of the variables.

The goal of the present article is therefore to facilitate the correction of attenuated correlations with Behseta et al.'s (2009) Bayesian approach. To this end, we will first illustrate the consequences of measurement error for the computation of the correlation and present Spearman's (1904) traditional attenuation formula. We will then describe Behseta et al.'s Bayesian correction method in detail. Finally, we will illustrate the use of the Bayesian correction with two empirical data sets: one focusing on the correlation between parameters of cumulative prospect theory (Tversky & Kahneman, 1992), the other focusing on the correlation between general intelligence and the drift rate parameter of the Ratcliff diffusion model (Ratcliff, 1978; Wagenmakers, 2009).

8.2 Attenuation of the Correlation and Spearman's Correction

In this section, we first show why the presence of measurement error decreases the observed correlation between two variables. We then discuss Spearman's (1904) method for correcting the attenuation. Let θ and β be two independent random variables and let $\hat{\theta}$ and $\hat{\beta}$ be the observed, noise-contaminated measurements:

$$\hat{\theta} = \theta + \epsilon_\theta \tag{8.1}$$

and

$$\hat{\beta} = \beta + \epsilon_\beta, \tag{8.2}$$

where ϵ_θ and ϵ_β are the measurement errors associated with θ and β, respectively. The correlation between the *unobserved* variables θ and β is given by

$$\text{Cor}(\theta, \beta) = \frac{\text{Cov}(\theta, \beta)}{\sqrt{\text{Var}(\theta) \times \text{Var}(\beta)}}. \tag{8.3}$$

Assuming that the measurement errors are uncorrelated with θ and β and with each other, the correlation between the *observed* variables $\hat{\theta}$ and $\hat{\beta}$ is given by

$$\text{Cor}(\hat{\theta}, \hat{\beta}) = \text{Cor}(\theta + \epsilon_\theta, \beta + \epsilon_\beta) = \frac{\text{Cov}(\theta, \beta)}{\sqrt{(\text{Var}(\theta) + \text{Var}(\epsilon_\theta)) \times (\text{Var}(\beta) + \text{Var}(\epsilon_\beta))}}. \tag{8.4}$$

It immediately follows from Equation 8.3 and Equation 8.4 that the observed correlation $\text{Cor}(\hat{\theta}, \hat{\beta})$ is always lower than the unobserved true correlation $\text{Cor}(\theta, \beta)$.

To remedy the problem of attenuation, Spearman (1904) proposed to correct the observed correlation coefficient using the reliabilities of the measurements:

$$r_{\theta\beta} = \frac{r_{\hat{\theta}\hat{\beta}}}{\sqrt{r_{\hat{\theta}\hat{\theta}}} \times \sqrt{r_{\hat{\beta}\hat{\beta}}}}, \tag{8.5}$$

where $r_{\theta\beta}$ is the corrected sample correlation coefficient, $r_{\hat{\theta}\hat{\beta}}$ is the observed sample Pearson correlation coefficient, and $r_{\hat{\theta}\hat{\theta}}$ and $r_{\hat{\beta}\hat{\beta}}$ are the reliabilities of $\hat{\theta}$ and $\hat{\beta}$, respectively. The reliabilities can be computed using the sample variances, s^2_θ and s^2_β, and the measurement error variances, $\sigma^2_{\epsilon_\theta}$ and $\sigma^2_{\epsilon_\beta}$, as follows:

$$r_{\hat{\theta}\hat{\theta}} = \frac{s^2_{\hat{\theta}} - s^2_{\epsilon_\theta}}{s^2_{\hat{\theta}}} \tag{8.6}$$

and

$$r_{\hat{\beta}\hat{\beta}} = \frac{s^2_{\hat{\beta}} - s^2_{\epsilon_\beta}}{s^2_{\hat{\beta}}}. \tag{8.7}$$

Confidence intervals for the corrected correlation coefficient $r_{\theta\beta}$ are outlined, for example, in Charles (2005) and Winnie and Belfry (1982).

Spearman's (1904) correction for measurement error is related to errors-in-variables models. Errors-in-variables models (e.g., Buonaccorsi, 2010; Cheng & Van Ness, 1999; Fuller, 1987; for Bayesian solutions, see Congdon, 2006; Gilks et al., 1996; Gustafson, 2004; Lunn et al., 2012) are extensions of standard regression models that —similar to Spearman's method— aim at correcting the bias in parameter estimates that results from measurement error. If the criterion and the response variables are both assumed to be measured with noise, Spearman's method and the correction within standard linear regression models result in the same disattenuation (see Behseta et al., 2009, Appendix). Ratcliff and Strayer (2014) outline an alternative method to deal with the adverse consequences of measurement error using Monte Carlo simulations. Cole and Preacher (2014) discuss solutions to deal with measurement uncertainty in the context of path analysis.

Example

Consider the following example. A researcher sets out to investigate the association between excitability (θ) and depression (β); she hypothesizes that people with increased responsiveness to threatening stimuli are likely to report more depressive symptoms. The researcher measures excitability using mean response time (RT) to pictures of threatening images and measures depression using a standard depression questionnaire. The researcher then correlates the observed mean RTs ($\hat{\theta}$) and depression scores ($\hat{\beta}$) of the $i = 1, ..., 10$ participants and obtains a Pearson correlation coefficient $r_{\hat{\theta}\hat{\beta}}$ of -0.57. The mean RTs and depression scores of the 10 hypothetical participants are reported in Table 8.1.

Aware of the fact that neither excitability nor depression is measured with perfect reliability, the researcher decides to use Spearman's (1904) method to correct the correlation coefficient for the unreliability of the observations. Assuming that the reliability of mean RT $r_{\hat{\theta}\hat{\theta}}$ is .65 and the reliability of the depression questionnaire $r_{\hat{\beta}\hat{\beta}}$ is .39, she obtains the following corrected correlation coefficient using Spearman's formula:

$$r_{\theta\beta} = \frac{-0.57}{\sqrt{0.65} \times \sqrt{0.39}} = -1.13. \tag{8.8}$$

Table 8.1 Data for the Hypothetical Experiment on the Relationship between Mean RT and Depression.

Participant i	Mean RT ($\hat{\theta}$)	Depression score ($\hat{\beta}$)
1	435	27
2	491	24
3	448	33
4	363	16
5	402	8
6	390	19
7	394	18
8	375	12
9	468	18
10	428	25
Observed variance (s^2)	1732.04	54.67
Error variance (s^2_ϵ)	605.73	33.56
Reliability	.65	.39

Note. The reliabilities are computed with the observed variances ($s^2_{\hat{\theta}}$ and $s^2_{\hat{\beta}}$) and the error variances ($s^2_{\epsilon_\theta}$ and $s^2_{\epsilon_\beta}$) using Equation 8.6 and Equation 8.7.

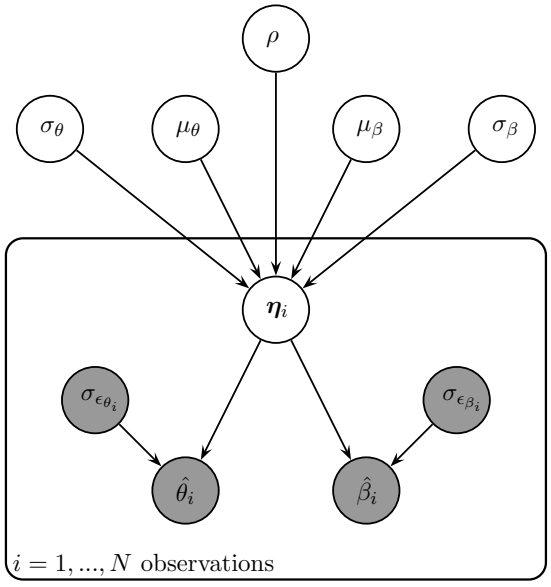
Our hypothetical experiment illustrates two shortcomings of Spearman's (1904) correction for attenuation. First, Spearman's method can produce corrected correlation coefficients in excess of 1.00 or -1.00, implying —as shown in Equation 8.3 and Equation 8.4— that true score variance is larger than the total observed variance (i.e., true score variance + error variance; see also Muchinsky, 1996). Second, Spearman's correction assumes homogenous error variances. Consider, however, our illustrative RT experiment; mean RT is unlikely to be measured with the same

precision for each participant. In fact, in most psychological investigations, error variance is likely to differ across the observations, as the observations usually represent different participants.

In the next section, we describe Behseta et al.'s (2009) Bayesian correction for attenuation, an alternative to Spearman's (1904) method that does not suffer from the above mentioned shortcomings: The Bayesian method yields corrected correlations that are bounded by -1 and 1 and can be applied to situations with heterogeneous error variances. Behseta et al.'s approach is conceptually straightforward and brings along the benefits of Bayesian modeling, such as easy-to-use statistical software and a coherent inferential framework.

8.3 Bayesian Correction for the Attenuation of the Correlation

Behseta et al.'s (2009) correction for the attenuation of the correlation is based on Bayesian multilevel (or hierarchical) modeling (e.g., Farrell & Ludwig, 2008; Gelman & Hill, 2007; M. D. Lee, 2011; Matzke & Wagenmakers, 2009; Rouder et al., 2005) and estimates the posterior distribution of the corrected correlation coefficient using Markov chain Monte Carlo sampling (MCMC; Gamerman & Lopes, 2006; Gilks et al., 1996). As Behseta et al. showed through a series of simulation studies, the Bayesian procedure is superior to Spearman's (1904) method in terms of the accuracy of the recovered corrected correlation and the coverage of the confidence interval, especially when the assumption of normality is violated. In this section, we describe Behseta et al.'s Bayesian approach in detail. The graphical representation of the Bayesian correction method is shown in Figure 8.1.



$$\mu_\theta, \mu_\beta \sim \text{Normal}(0, 1000)$$

$$\sigma_\theta, \sigma_\beta \sim \text{Uniform}(0, 100)$$

$$\rho \sim \text{Uniform}(-1, 1)$$

$$\boldsymbol{\eta}_i \sim \text{MultivariateNormal}\left((\mu_\theta, \mu_\beta), \begin{bmatrix} \sigma_\theta^2 & \rho\sigma_\theta\sigma_\beta \\ \rho\sigma_\theta\sigma_\beta & \sigma_\beta^2 \end{bmatrix}\right)$$

$$\hat{\theta}_i \sim \text{Normal}(\eta_{1_i}, \sigma^2_{\epsilon_{\theta i}})$$

$$\hat{\beta}_i \sim \text{Normal}(\eta_{2_i}, \sigma^2_{\epsilon_{\beta i}})$$

Figure 8.1 *Graphical model for Behseta et al.'s (2009) Bayesian correction for the attenuation of the correlation.* Observed variables are represented by shaded nodes and unobserved variables are represented by unshaded nodes. The graph structure indicates dependencies between the nodes (e.g., M. D. Lee, 2008). The normal distributions are parameterized in terms of the variance σ^2. $\eta_{1_i} = \theta_i$; $\eta_{2_i} = \beta_i$.

Level I: Modeling the Observed Data

The shaded nodes in the horizontal panel in Figure 8.1 represent the observed variables in the model. As before, let θ and β represent the true values, $\hat{\theta}$ and $\hat{\beta}$ the corresponding observed values, and ϵ_θ and ϵ_β the errors associated with θ and β, respectively. For each observation i, $i = 1, ..., N$, $\hat{\theta}_i$ and $\hat{\beta}_i$ are given by

$$\hat{\theta}_i = \theta_i + \epsilon_{\theta_i} \tag{8.9}$$

and

$$\hat{\beta}_i = \beta_i + \epsilon_{\beta_i}. \tag{8.10}$$

The error terms ϵ_{θ_i} and ϵ_{β_i} are assumed to be drawn from independent zero-centered normal distributions with variance $\sigma_{\epsilon_i^2}$:

$$\epsilon_{\theta_i} \sim \text{Normal}(0, \sigma^2_{\epsilon_{\theta_i}}) \tag{8.11}$$

and

$$\epsilon_{\beta_i} \sim \text{Normal}(0, \sigma^2_{\epsilon_{\beta_i}}). \tag{8.12}$$

The error variances are assumed to be known a priori or are estimated from data. Note that contrary to Spearman's (1904) correction, the Bayesian approach does not assume homogenous error variances across the N observations —each observation i has its own error variance.

Level II: Modeling Unobserved Means, Variances, and Correlations

The unshaded nodes in Figure 8.1 represent the unobserved variables in the model. For each observation i, $i = 1, ..., N$, $\boldsymbol{\eta}_i = (\theta_i, \beta_i)$ is assumed to follow a bivariate normal distribution, with mean $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\eta}_i \sim \text{Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \tag{8.13}$$

where

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_\theta \\ \mu_\beta \end{pmatrix} \tag{8.14}$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma^2_\theta & \rho\sigma_\theta\sigma_\beta \\ \rho\sigma_\theta\sigma_\beta & \sigma^2_\beta \end{pmatrix}. \tag{8.15}$$

Here ρ is the corrected correlation between θ and β —the correlation of interest that is uncontaminated by measurement error.

The means $\boldsymbol{\mu}$ and the three elements (i.e., σ_θ, σ_β, and ρ) of the variance-covariance matrix $\boldsymbol{\Sigma}$ are estimated from the data and require prior distributions. In the present article, we will use the following prior set-up:

$$\mu_\theta, \mu_\beta \sim \text{Normal}(0, 1000), \tag{8.16}$$

$$\sigma_\theta, \sigma_\beta \sim \text{Uniform}(0, 100), \tag{8.17}$$

and

$$\rho \sim \text{Uniform}(-1, 1). \tag{8.18}$$

Note that Behseta et al. (2009) used a slightly different prior set-up, where $\boldsymbol{\Sigma}$ was assumed to follow an Inverse-Wishart distribution. We chose to model the individual components in $\boldsymbol{\Sigma}$ rather than $\boldsymbol{\Sigma}$ itself because this provides an intuitive specification that allows users to adapt the range of the uniform prior for σ_θ and σ_β to the measurement scale of their variables in a straightforward manner.

Bayesian Parameter Estimation

In Bayesian parameter estimation, the prior distributions for the model parameters are updated by the incoming data to obtain posterior distributions. The posterior distribution often cannot be derived analytically; rather it must be approximated using numerical sampling techniques, such as MCMC sampling (Gamerman & Lopes, 2006; Gilks et al., 1996). The posterior distribution quantifies the uncertainty of the parameter estimates. The central tendency of the posterior, such as the mean or median, is often used as a point estimate. The dispersion of the posterior, such as the standard deviation or the percentiles, quantifies the precision of the parameter estimate: the larger the dispersion, the greater the uncertainty in the estimated parameter. For example, the 95% Bayesian credible interval for the corrected correlation ρ ranges from the 2.5^{th} to the 97.5^{th} percentile of its posterior distribution, indicating that we can be 95% confident that the true value of ρ lies within this range.

Parameter estimation for the present approach may proceed using standard Bayesian statistical software, such as WinBUGS (Bayesian inference Using Gibbs Sampling for Windows; Lunn et al., 2012; for an introduction for psychologists, see Kruschke, 2010b, and M. D. Lee & Wagenmakers, 2013). The WinBUGS script is presented in Appendix E.1. Note that the WinBUGS script requires minimal, if any, modification to run under OpenBUGS (Lunn et al., 2009) or JAGS (Plummer, 2003, 2013). The Stan project (Stan Development Team, 2012) provides yet another alternative to obtain posterior distributions for the parameters.

Bayesian Hypothesis Testing

Once the model parameters are estimated, we can formally assess the presence of an association using Bayes hypothesis testing. Throughout the article, we will rely on the Bayes factor —a popular Bayesian model selection measure— to quantify the probability of the data under the null hypothesis (H_0: $\rho = 0$) relative to the probability of the data under the alternative hypothesis (H_1: $\rho \neq 0$). For instance, BF_{01} of 10 indicates that the data are 10 times more likely under the null hypothesis than under the alternative hypothesis. Alternatively, BF_{01} of $\frac{1}{10}$ indicates that the data are 10 times more likely under the alternative hypothesis than under the null hypothesis (e.g., Jeffreys, 1961; Kass & Raftery, 1995).

We will compute two-sided Bayes factors using the Savage-Dickey density ratio method (e.g., Dickey & Lientz, 1970; Wagenmakers et al., 2010; Wetzels, Grasman, & Wagenmakers, 2010), assuming a uniform prior distribution for the correlation ρ parameter. The Savage-Dickey density ratio is an intuitive and flexible approach for the computation of Bayes factors in nested model comparison. Applied to the present situation, BF_{01} is given by the ratio of the height of the posterior and the prior distribution of ρ under the alternative hypothesis at $\rho = 0$. The height of the prior distribution is calculated by evaluating the uniform probability density function on -1.00 and 1.00 at $\rho = 0$; the height of the uniform prior distribution at $\rho = 0$ equals $\frac{1}{2}$. The height

of the posterior distribution is calculated as follows. We first obtain samples from the posterior distribution of ρ using WinBUGS. We then fit to the posterior samples a scaled beta distribution with parameters α and β. Lastly, we evaluate the height of the scaled beta distribution at $\rho = 0$ using the obtained α and β parameters. One-sided (i.e., order-restricted) Bayes factors will be computed as recommended by Morey and Wagenmakers (2014), namely by correcting the two-sided Bayes factor using the proportion of posterior samples that is consistent with the order-restriction. The R script (R Core Team, 2012) for the computation of the Bayes factor is available in the supplemental materials at `http://dora.erbe-matzke.com/publications.html`.

8.4 Empirical Examples

In this section, we illustrate the use of Behseta et al.'s (2009) Bayesian correction for attenuation with two empirical data sets. In the first example, we assessed the correlation between parameters of cumulative prospect theory (CPT; Tversky & Kahneman, 1992) measured at two different time points. In the second example, we assessed the correlation between general intelligence and the drift rate parameter of the Ratcliff diffusion model (Ratcliff, 1978). In order to apply the graphical model shown in Figure 8.1, we first estimated posterior distributions for the CPT and diffusion model parameters for each participant separately. We then computed posterior distributions for the uncorrected[1] and the corrected correlation coefficients using the mean of the posterior distribution of the individual model parameters as point estimate.[2] In the corrected analysis, we used the variance of the posterior distribution of the individual model parameters as estimate for the participant-specific error variance. Finally, we formally assessed the presence of an association using Bayes hypothesis testing.

Example 1: Inference for the Correlation between Parameters of CPT

As our first example, we computed the uncorrected and corrected correlation in a data set obtained from a decision making experiment reported in Glöckner and Pachur (2012). The 64 participants were instructed to choose between monetary gambles in two experimental sessions. The two sessions were separated by one week and each featured 138 two-outcome gambles. The observed choice data were modeled using cumulative prospect theory (CPT; Tversky & Kahneman, 1992). CPT has a number of free parameters that reflect specific individual differences. Here we focus on the δ parameter that governs how individual decision makers weight the probability information of the gambles: higher values of δ indicate high risk aversion, whereas lower values of δ indicate less risk aversion.

As in other models in the decision making literature, the CPT parameters are assumed to be relatively stable across short periods of time. Here we therefore examined the association between the δ parameter measured at the two experimental sessions. The CPT was fit to the data of each individual participant, separately for the two measurement occasions. Model parameters were obtained using Bayesian parameter estimation with JAGS (Plummer, 2013), by adapting an existing model by Nilsson et al. (2011).[3] The prior for δ was uninformative across possible

[1] In the uncorrected Bayesian analysis, the bivariate normal distribution in Equation 8.13 was placed directly on the observed data.

[2] As we will discuss later, in the Bayesian framework, we are not limited to the two-step procedure outlined in this section; Bayesian hierarchical modeling allows for the *simultaneous* estimation of the individual parameters and the group-level means and covariances.

[3] The CPT account of performance in the Glöckner and Pachur (2012) data set is merely an illustration; we do not suggest that the CPT with the present parameter setting provides the best, or even an adequate, description

parameter values found in previous research but excluded theoretically implausible values.

We computed posterior distributions for the uncorrected and corrected correlation between the δ_1 and the δ_2 parameters. We treated the mean of the posterior distribution of the individual δ_1 and δ_2 parameters as observed data. In the corrected analysis, we used the variance of the posterior distribution of the individual δ_1 and δ_2 parameters as estimate for the participant-specific error variance. Lastly, we computed one-sided Bayes factors for the uncorrected and corrected correlation to quantify the evidence that the data provide for H_0 ($\rho = 0$) relative to H_1 ($\rho > 0$).

Results

The results are shown in Figure 8.2. The top-row panels show scatterplots of the observed $\hat{\delta}_{i,1}$ and $\hat{\delta}_{i,2}$ parameters and the standard deviation of the measurement errors (i.e., $\sigma_{\epsilon_{\delta_{i,1}}}$ and $\sigma_{\epsilon_{\delta_{i,2}}}$). The bottom left panel shows a scatterplot of the mean of the posterior distribution of the corrected δ parameters (i.e., $\boldsymbol{\eta}_i = (\delta_{i,1}, \delta_{i,2})$) and the corresponding posterior standard deviations. The bottom right panel shows the posterior distribution of the uncorrected and the corrected correlation (i.e, ρ).

As shown in the upper left panel of Figure 8.2, the uncorrected Pearson correlation between the observed $\hat{\delta}_1$ and $\hat{\delta}_2$ parameters is .62. If we take into account the uncertainty of the observations, the correlation increases substantially: The bottom left panel shows that the posterior means of the corrected $\delta_{i,1}$ and $\delta_{i,2}$ parameters are associated very highly; the bottom right panel shows that the posterior distribution of the corrected ρ parameter is shifted to the right relative to the posterior of the uncorrected correlation. In fact, after correcting for the noise in $\hat{\delta}_1$ and $\hat{\delta}_2$, the mean of the posterior distribution of ρ increases from .61 to .92.

One-sided Bayes factors indicate decisive evidence (Jeffreys, 1961) for the presence of an association for the corrected as well as the uncorrected ρ parameter; in both cases, the data are more than 4,000,000 times more likely under H_1 than under H_0. This result is visually evident from the fact that the posterior distributions are located away from zero such that their height at $\rho = 0$ is all but negligible.

Note that the dramatic increase in the correlation observed in the present data set is not unusual. Figure 8.3 shows the results of a simulation study where we investigated the magnitude of the expected attenuation for different values of the latent correlation in a parameter setting that resembles the one found in the present data set. We conducted five sets of simulations, each with a different true "latent" correlation: .92 (i.e., the posterior mean of the corrected ρ in the present data set), .21, 0, -.21, and -.92. For each set of simulations, we generated 1,000 synthetic data sets with $N = 64$, using the error variances $\sigma^2_{\epsilon_{\delta_{i,1}}}$ and $\sigma^2_{\epsilon_{\delta_{i,2}}}$ obtained from fitting the CPT to the data, and the posterior mean of the μ_{δ_1}, μ_{δ_2}, σ_{δ_1}, and σ_{δ_2} parameters estimated with the Bayesian correction method.[4] In each synthetic data set, we then computed the attenuated "observed" correlation. The gray violin plots in Figure 8.3 show the distribution of the predicted attenuated correlations for the five values of the latent correlation.

Two results are noteworthy. First, all else being equal, the larger the absolute value of the true latent correlation, the larger the attenuation. This relationship is also evident from Equation 8.3 and Equation 8.4. Second, the observed attenuated correlation in the empirical data (i.e., .62; horizontal dashed line) is slightly higher than expected, but is well within the 2.5^{th} and 97.5^{th} percentile of the attenuated correlations predicted by the model with the present parameter setting.

of the data of the individual participants. Note also that Glöckner and Pachur reported the results from fitting a slightly different model than the one used in the present article.

[4]Note that this procedure is not the same as the posterior predictive assessment of model fit (e.g., Gelman & Hill, 2007; Gelman et al., 1996).

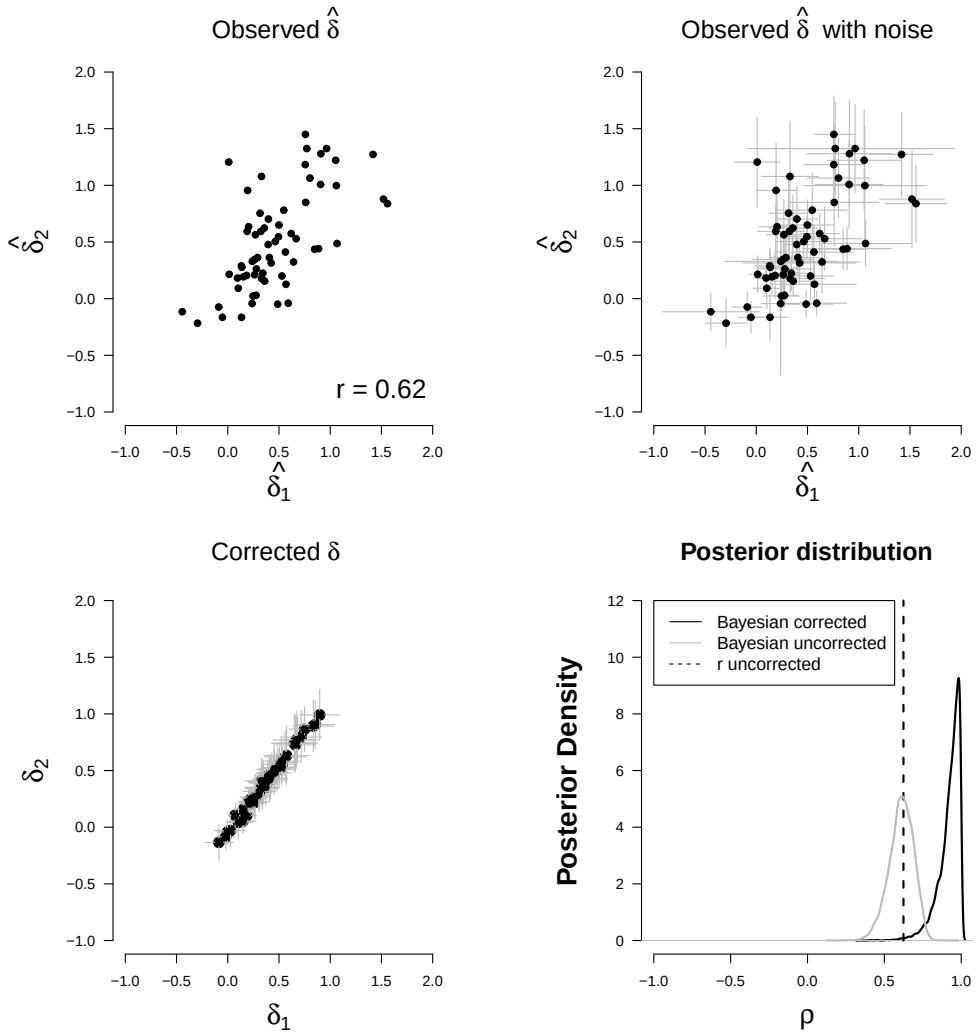


Figure 8.2 *Corrected and uncorrected correlation between parameters of cumulative prospect theory (CPT).* The top left panel shows a scatterplot of the observed $\hat{\delta}_{i,1}$ and $\hat{\delta}_{i,2}$ parameters. The top right panel shows a scatterplot of the observed $\hat{\delta}_{i,1}$ and $\hat{\delta}_{i,2}$ parameters and the standard deviation of the measurement errors (i.e., $\sigma_{\epsilon_{\delta_{i,1}}}$ and $\sigma_{\epsilon_{\delta_{i,2}}}$; gray lines). The $\hat{\delta}_{i,1}$, $\hat{\delta}_{i,2}$, $\sigma_{\epsilon_{\delta_{i,1}}}$, and $\sigma_{\epsilon_{\delta_{i,2}}}$ values were obtained from fitting the CPT to the data. The bottom left panel shows a scatterplot of the posterior mean of the corrected $\delta_{i,1}$ and $\delta_{i,2}$ parameters (i.e., $\boldsymbol{\eta}_i = (\delta_{i,1}, \delta_{i,2})$). The gray lines show the standard deviation of the posterior distribution of the parameters. The bottom right panel shows the posterior distribution of the uncorrected (gray density line) and the corrected correlation (i.e, ρ; black density line). The dashed vertical line shows the Pearson correlation coefficient computed with the observed $\hat{\delta}_{i,1}$ and $\hat{\delta}_{i,2}$ parameters. r = Pearson correlation coefficient.

In sum, correcting the correlation for measurement noise resulted in a dramatic increase in the correlation between the CPT parameters; the mean of the posterior distribution of the correlation parameter increased from .62 to .92. Despite this increase, the Bayes factor indicated decisive evidence for the presence of an association in the corrected as well as the uncorrected analysis.

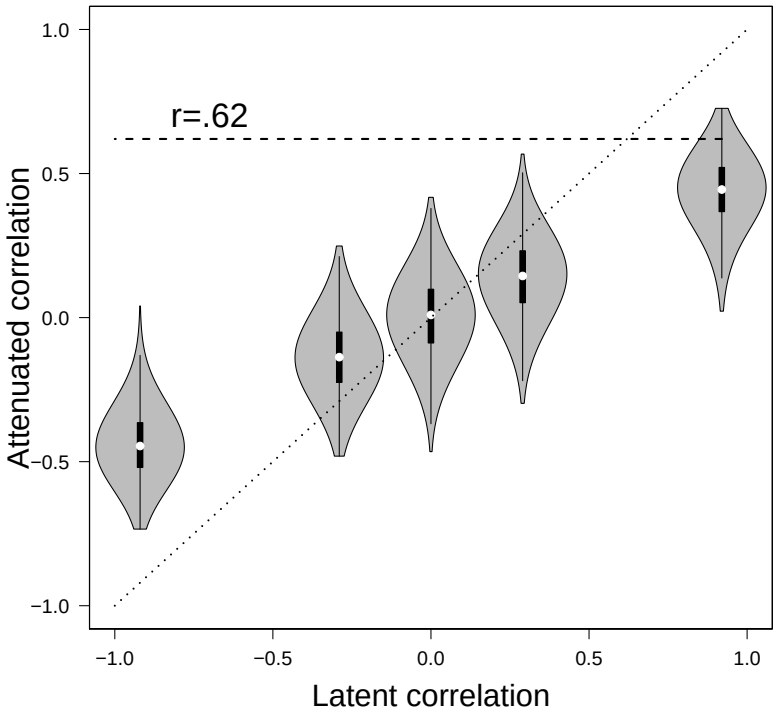


Figure 8.3 *Expected attenuation of the correlation between the δ parameters of cumulative prospect theory (CPT).* The figure shows that the attenuation of the correlation increases with the absolute value of the latent correlation. The gray violin plots show the distribution of the predicted attenuated correlations for five values of the latent correlation. Violin plots (e.g., Hintze & Nelson, 1998) combine information available from density plots with information about summary statistics in the form of box plots. The black boxplot in each violin plot ranges from the 25^{th} to the 75^{th} percentile of the attenuated correlations predicted by the model, where the white circle represents the median of the predictions. The predicted correlations were generated using the error variances $\sigma^2_{\epsilon_{\delta_{i,1}}}$ and $\sigma^2_{\epsilon_{\delta_{i,2}}}$ obtained from fitting the CPT to the data, and the posterior mean of the μ_{δ_1}, μ_{δ_2}, σ_{δ_1}, and σ_{δ_2} parameters estimated with the Bayesian correction method. The dashed line shows the observed attenuated correlation in the empirical data. r = Pearson correlation coefficient.

Example 2: Inference for the Correlation between General intelligence and Diffusion Model Drift Rate

As a second example, we computed the uncorrected and corrected correlation in a data set collected by Weeda and Verouden (unpublished data). The data set featured response times (RT) and accuracy data from 51 participants performing a two-choice RT task. The stimuli were borrowed from the π-paradigm (Vickers, Nettelbeck, & Willson, 1972; Jensen, 1998) and consisted of a series of configurations, each with one horizontal and two vertical lines (i.e.; two legs) that together formed the letter π, with one of the vertical lines longer than the other. The task was to indicate by means of a button press whether the left or the right leg of the π was longer. Task difficulty was manipulated on three levels —easy, medium, and difficult— by varying the difference between

the length of the two legs.

The RT and accuracy data were modeled with the Ratcliff diffusion model (Ratcliff, 1978; Wagenmakers, 2009). The diffusion model provides a theoretical account of performance in speeded two-choice tasks. The four key parameters of the diffusion model correspond to well-defined psychological processes (Ratcliff & McKoon, 2008; Voss et al., 2004), such as response caution (a), a priori bias (z), the time taken up by processes unrelated to decision making (e.g., encoding and motor processes; T_{er}), and —the parameter of interest in the present article— the rate of information accumulation drift rate (v).[5]

The drift rate parameter of the diffusion model has been repeatedly associated with higher cognitive functions and reasoning (i.e., Ratcliff, Schmiedek, & McKoon, 2008; Ratcliff, Thapar, & McKoon, 2010; Schmiedek et al., 2007; van Ravenzwaaij, Brown, & Wagenmakers, 2011), and this is why we focus here on the correlation between drift rate and general intelligence. The four key diffusion model parameters were estimated from the RT and accuracy data for each participant separately using Bayesian parameter estimation with the diffusion model JAGS module (Wabersich & Vandekerckhove, 2014). We used uninformative prior distributions based on parameter values reported in Matzke and Wagenmakers (2009). As drift rate is known to decrease with increasing task difficulty (e.g., Ratcliff & McKoon, 2008), we used the following order-restriction: $v_{difficult} < v_{medium} < v_{easy}$. The remaining parameters were constrained to be equal across the conditions, and we set $z = \frac{a}{2}$.[6] General intelligence was measured by the total score of the 20-min version of the Raven Progressive Matrices Test (Hamel & Schmittmann, 2006; Raven, Raven, & Court, 1998).

We computed posterior distributions for the uncorrected and corrected correlation between the mean of the drift rate parameters across the three task difficulty conditions ($\bar{v}$) and the Raven total score (g). For mean drift rate $\bar{v}$, we treated the mean of the posterior distribution of the individual $\bar{v}$ parameters as observed data.[7] In the corrected analysis, we used the variance of the posterior distribution of the individual $\bar{v}$ parameters as estimate for the participant-specific error variance. For the Raven total score g, we assumed homogenous error variance and —for illustrative purposes— investigated how the extent of the correction varies as a function of the amount of measurement noise assumed in the data. Specifically, we examined three scenarios: we assumed that 5%, 25%, and 55% of the total variance in Raven scores is attributable to measurement error, corresponding to excellent, acceptable, and very poor reliability, respectively. Lastly, we computed one-sided Bayes factors for the uncorrected and corrected correlation to quantify the evidence that the data provide for H_0 ($\rho = 0$) relative to H_1 ($\rho > 0$) under each scenario.

[5] In addition to these key parameters, the diffusion model also features parameters that describe the trial-to-trial variability of the key parameters.

[6] The diffusion model account of performance in the Weeda and Verouden data set (unpublished data) is merely an illustration; we do not suggest that the diffusion model with the present parameter constraints provides the best, or even an adequate, description of the data of the individual participants.

[7] Note that the scale of both drift rate and general intelligence are bounded: $\bar{v}$ can take on values between 0 and 5.86 (i.e., prior range) and the Raven total score can take on values between 0 and 36. The use of the bivariate normal group-level distribution shown in Figure 8.1 is therefore theoretically unjustified. As a solution, we may transform the individual $\bar{v}$ parameters and the Raven scores g to the real line using a probit transformation. Additional analyses not reported here confirmed that using the transformed $\bar{v}$ and g values yields results that are very similar to the ones obtained using the untransformed drift rates and Raven scores. For simplicity, in the present article, we will report the results of modeling the untransformed $\bar{v}$ and Raven g values.

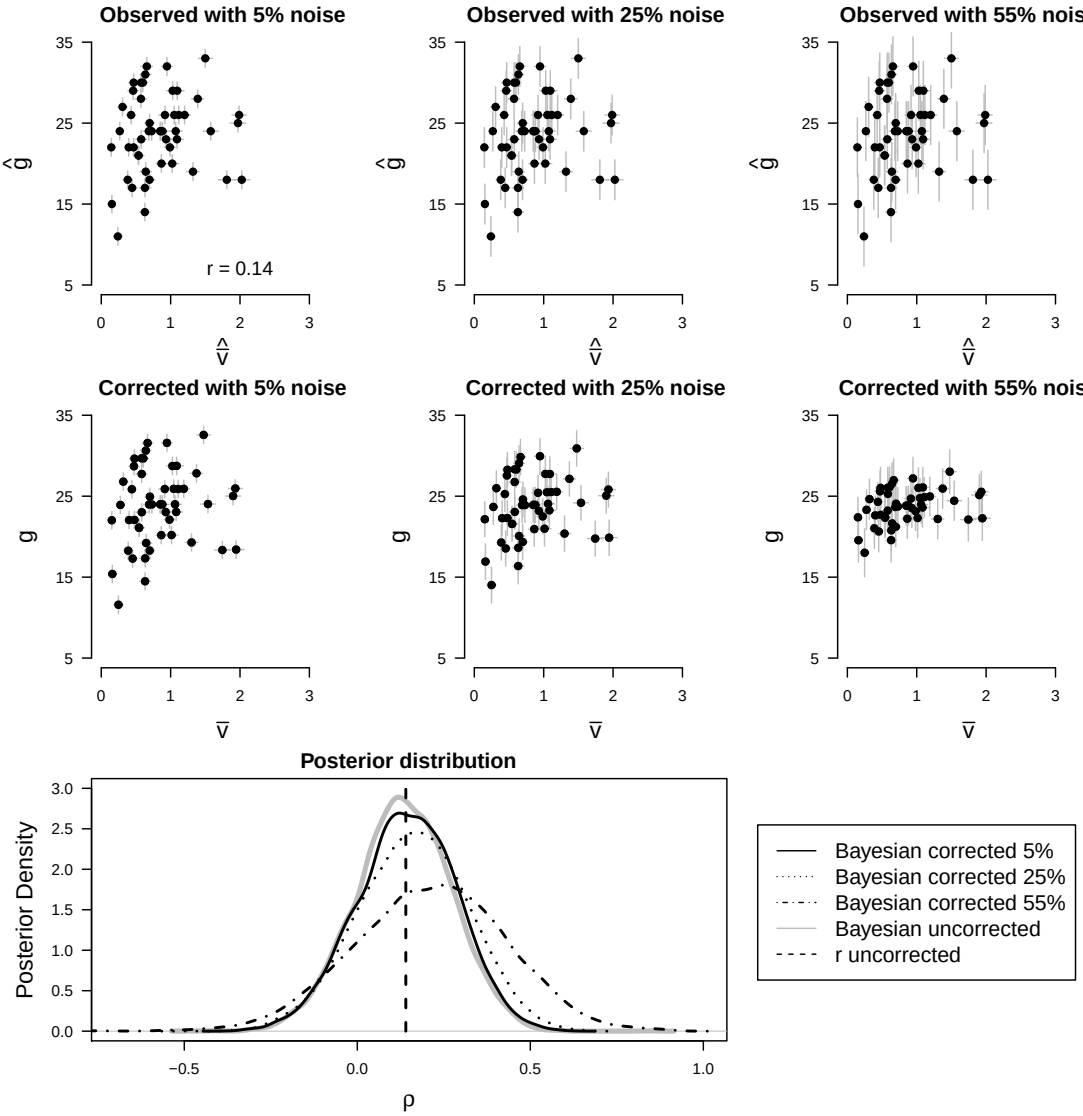


Figure 8.4 *Corrected and uncorrected correlation between mean drift rate $\bar{v}$ and Raven total score g.* The top panels show scatterplots of the observed mean drift rates $\hat{\bar{v}}_i$ and Raven total scores $\hat{g}_i$ with the standard deviation of the measurement errors (i.e., $\sigma_{\epsilon_{\bar{v}_i}}$ and σ_{ϵ_g}; gray lines) for the 5%, 25%, and 55% measurement noise scenarios. The $\hat{\bar{v}}_i$ and the $\sigma_{\epsilon_{\bar{v}_i}}$ values were obtained from fitting the diffusion model to the data. The middle panels show scatterplots of the posterior mean of the corrected $\bar{v}_i$ and g_i parameters (i.e., $\boldsymbol{\eta}_i = (\bar{v}_i, g_i)$). The gray lines show the standard deviation of the posterior distribution of the parameters. The bottom panel shows the posterior distribution of the uncorrected (gray density line) and the corrected correlations (i.e, ρ) for the 5% (solid black density line), 25% (dotted black density line), and the 55% (dotted-dashed black density line) measurement noise scenarios. The dashed vertical line shows the Pearson correlation coefficient computed with the observed mean drift rates $\hat{\bar{v}}_i$ and Raven total scores $\hat{g}_i$. r = Pearson correlation coefficient.

Results

The results are shown in Figure 8.4. The top-row panels show scatterplots of the observed mean drift rates $\hat{\bar{v}}$ and Raven total scores $\hat{g}$ with the standard deviation of the measurement errors (i.e., $\sigma_{\epsilon_{\bar{v}_i}}$ and σ_{ϵ_g}) for the 5%, 25%, and 55% noise scenarios. The middle-row panels show scatterplots of the mean of the posterior distribution of the corrected $\bar{v}$ and g parameters (i.e., $\boldsymbol{\eta}_i = (\bar{v}_i, g_i)$) and the corresponding posterior standard deviations for each scenario. The bottom panel shows the posterior distribution of the uncorrected and the corrected correlation (i.e, ρ) for each scenario.

As shown in the upper left panel of Figure 8.4, the uncorrected Pearson correlation between the observed $\hat{\bar{v}}$ parameters and the observed Raven scores $\hat{g}$ is .14. As expected, the magnitude of the correction for attenuation increases with increasing error variance: The middle-row panels show that the association between the posterior means of the corrected $\bar{v}_i$ and g_i parameters becomes stronger; the bottom panel shows that the mean of the posterior distribution of the corrected ρ parameter is progressively shifted to higher values. Note, however, that the correction is modest even if we assume that the Raven total score is an extremely unreliable indicator of general intelligence. The mean of the posterior distribution of ρ equals .13 in the uncorrected analysis, .14 in the corrected analysis with 5% noise, .16 in the corrected analysis with 25% noise, and .21 in the corrected analysis with 55% noise. Note also that the posterior of ρ tends to be quite spread out, a tendency that becomes more pronounced with increasing error variance.

One-sided Bayes factors indicate evidence *against* the presence of an association between mean drift rate and Raven total score. The evidence, however, is "worth no more than a bare mention" (Jeffreys, 1961). The BF_{01} decreases from 2.13 in the uncorrected analysis to 2.00 in the corrected analysis with 5% noise, to 1.75 in the corrected analysis with 25% noise, and to 1.32 in the corrected analysis with 55% noise. Even with extremely unreliable Raven scores, the data are thus still more likely to have occurred under H_0 than under H_A. Note however that BF_{01} of 1.32 —or even BF_{01} of 2.13— constitutes almost perfectly ambiguous evidence, indicating that the data are not sufficiently diagnostic to discriminate between H_0 and H_A. Inspection of the posterior distribution of the ρ parameters suggests a similar conclusion: ρ is estimated quite imprecisely (i.e., the posteriors are spread out) in all four analyses.

Figure 8.5 shows the results of a simulation study where we investigated the magnitude of the expected attenuation for different values of the latent correlation in a parameter setting that resembles the one found in the present data set. Throughout the simulations, we assumed that 55% of the total variance of the Raven scores is attributable to error variance. We conducted five sets of simulations, each with a different value of the true "latent" correlation: .92, .21 (i.e., the posterior mean of the corrected ρ in the present data set), 0, -.21, and -.92. For each set of simulations, we generated 1,000 synthetic data sets with $N = 51$, using the $\sigma^2_{\epsilon_{\bar{v}_i}}$ parameters obtained from fitting the diffusion model to the data, and the posterior means of the $\mu_{\bar{v}}$, μ_g, $\sigma_{\bar{v}}$, and σ_g parameters estimated with the Bayesian correction method. In each synthetic data set, we then computed the attenuated "observed" correlation. The gray violin plots in Figure 8.5 show the distribution of the predicted attenuated correlations for the five values of the latent correlation.

As pointed out earlier, all else being equal, the larger the absolute value of the true latent correlation, the larger the attenuation. Moreover, considering the relatively low corrected correlation in the present data set, an attenuation of only $.21 - .14 = .07$ is perfectly reasonable. In fact, the median of the attenuated correlations predicted by the model with the present parameter setting very closely approximates the observed attenuated correlation in the empirical data (i.e., .14; horizontal dashed line).

In sum, correcting for measurement noise resulted in negligible increase in the correlation between drift rate and general intelligence; even with unrealistically unreliable Raven scores, the

posterior mean of the correlation parameter increased from .13 only to .21. Regardless of the type of analysis (uncorrected or uncorrected) and regardless of the magnitude of the error variance, the Bayes factor indicated evidence against the presence of an association between drift rate and general intelligence. The evidence for the null hypothesis was, however, only anecdotal, a result that is attributable to the substantial uncertainty in the estimated correlation parameters.

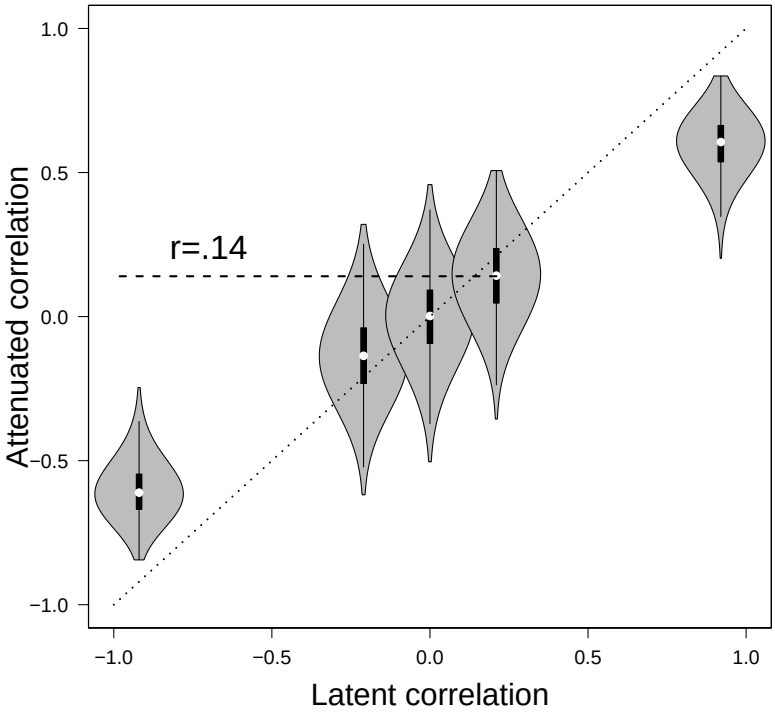


Figure 8.5 *Expected attenuation of the correlation between mean drift rate and general intelligence.* The figure shows that the attenuation of the correlation increases with the absolute value of the latent correlation. The gray violin plots show the distribution of the predicted attenuated correlations for five values of the latent correlation. The black boxplot in each violin plot ranges from the 25^{th} to the 75^{th} percentile of the attenuated correlations predicted by the model, where the white circle represents the median of the predictions. The predicted correlations were generated assuming 55% error variance in the Rave scores, using the $\sigma^2_{\epsilon_{\bar{v}_i}}$ parameters obtained from fitting the diffusion model to the data, and the posterior means of the $\mu_{\bar{v}}$, μ_g, $\sigma_{\bar{v}}$, and σ_g parameters estimated with the Bayesian correction method. The dashed line shows the observed attenuated correlation in the empirical data. r = Pearson correlation coefficient.

8.5 Discussion

Although various approaches are available to correct the correlation in the presence of measurement error, such corrections are presently the exception rather than the rule in experimental psychology. The goal of the present paper was therefore to demonstrate the application of the Bayesian correction of attenuated correlations (Behseta et al., 2009). We illustrated the use of the Bayesian

method with two empirical data sets; in each data set, we first estimated posterior distributions for the uncorrected and corrected correlation and then computed Bayes factors to quantify the evidence that the data provide for the presence of the association.

In our first example, we computed the uncorrected and corrected correlation between parameters of cumulative prospect theory (Tversky & Kahneman, 1992) and demonstrated that correcting for measurement noise can result in a dramatic increase in the correlation: the mean of the posterior distribution of the correlation parameter increased from .61 to .92. The Bayes factor indicated decisive evidence for the presence of an association in the corrected as well as the uncorrected analysis. In our second example, we assessed the correlation between general intelligence and the drift rate parameter of the diffusion model (Ratcliff, 1978; Wagenmakers, 2009), where we examined three scenarios: we assumed that 5%, 25%, and 55% of the total variance in Raven scores is attributable to measurement error, corresponding to excellent, acceptable, and very poor reliability, respectively. Correcting for measurement noise resulted in negligible increase in the correlation; even with extremely unreliable Raven scores, the posterior mean of the correlation parameter increased from .13 to only .21. In all analyses, we obtained anecdotal evidence against the presence of an association between drift rate and general intelligence, a result that is attributable to the substantial uncertainty in the estimated correlation parameters.

Behseta et al.'s (2009) Bayesian correction for attenuation is easy-to-use and conceptually straightforward. In fact, the present approach can be viewed as a simple Bayesian structural equation model with two latent variables, each with a single indicator (see, for example, S.-Y. Lee, 2007; Song & Lee, 2012). The original formulation of the Bayesian correction method relies on a slightly different prior set-up than the one used in the present article. Specifically, Behseta et al. used an Inverse-Wishart distribution to model the variance-covariance matrix of the corrected observations, whereas we chose to model the standard deviations and the correlation separately using uniform distributions. We feel that the present specification is more intuitive and allows users to adapt the range of the uniform prior for the standard deviations to the measurement scale of their variables in a straightforward manner. Note also that Bayesian parameter estimation is insensitive to the choice of the prior as long as sufficiently informative data are available (e.g., Edwards et al., 1963; Gill, 2002; M. D. Lee & Wagenmakers, 2013).

Correcting the correlation for measurement noise is, of course, impossible unless the error variance of the observations is known or can be estimated from the data. Our investigation of the extent of the correction as a function of the amount of measurement noise in the Raven scores served merely as an illustration. In real-life applications, the magnitude of the error variance should not be cherry-picked to obtain the desired (higher) correlation; rather it should be estimated from the data. Bayesian inference is particularly suitable for modeling measurement error because the resulting posterior distributions can be automatically used to quantify the uncertainty of the parameter estimates.[8] Accordingly, in our two examples, we treated the mean of the posterior distribution of the CPT and diffusion model parameters as observed data and used the variance of the posterior distributions as estimate for the participant-specific error variance. Note also that within the Bayesian framework, we are not limited to the two-step procedure illustrated in the present article; Bayesian hierarchical modeling allows for the *simultaneous* estimation of the individual parameters and the group-level means and covariances, where the correlation is automatically adjusted for the uncertainty of the individual estimates. The Bayesian estimation of covariance structures is illustrated, for example, in Gelman and Hill (2007), Klauer (2010), Matzke et al. (in press), Rouder et al. (2008), and Rouder et al. (2007).

[8] Naturally, using the variance of the posterior distribution as estimate for the error variance is only sensible if the posteriors are approximately normally distributed.

Our literature review showed that nearly 50% of the articles published in the 2012 volume of the *Journal of Experimental Psychology: General* reported at least one Pearson product-moment correlation coefficient. Despite the wide-spread use of correlations, most researchers do not acknowledge explicitly that the observed correlation often underestimates the true correlation if the variables are measured with noise. Here we illustrated the use of a Bayesian correction procedure and showed that its application can dramatically increase the estimated correlation. Of course, estimating the uncertainty of the observations is not always feasible. Also, our simulations confirmed that for relatively low true correlations, the correction is likely to have only a negligible effect. We nevertheless urge researchers to carefully consider the issue of the attenuation and whenever possible correct the observed correlation for the uncertainty of the measurements.