A hierarchical Bayesian model for measuring individual-level and group-level numerical representations

Thomas J. Faulkenberry
Department of Psychological Sciences, Tarleton State University, USA

Abstract
A popular method for indexing numerical representations is to compute an individual estimate of a response time effect, such as the SNARC effect or the numerical distance effect. Classically, this is done by estimating individual linear regression slopes and then either pooling the slopes to obtain a group-level slope estimate, or using the individual slopes as predictors of other phenomena. In this paper, I develop a hierarchical Bayesian model for simultaneously estimating group-level and individual-level slope parameters. I show examples of using this modeling framework to assess two common effects in numerical cognition: the SNARC effect and the numerical distance effect. Finally, I demonstrate that the Bayesian approach can result in better measurement fidelity than the classical approach, especially with small samples.

Keywords: hierarchical regression, Bayesian modeling, numerical representations, SNARC effect, distance effect

The purpose of this paper is to introduce a hierarchical Bayesian model for measuring two types of mental representations that are often assessed in numerical cognition: the SNARC effect (Dehaene, Dupoux, & Mehler, 1990) and the numerical distance effect (Moyer & Landauer, 1967). The SNARC effect (Spatial-Numerical Associations of Response Codes) refers to the finding that response times (RTs) for small numerical stimuli are faster with the left hand, whereas RTs for large numbers are faster with the right hand. The numerical distance effect (NDE) refers to the finding that RTs decrease as the numerical distance between stimulus numbers increases. Both effects are thought to reflect particular aspects of mental representation of number. As the SNARC effect occurs even when number magnitude is not salient to a given task (e.g., asking whether a given number is odd or even), the SNARC effect is often taken as an index of automatic processing of number magnitude (Wood, Willmes, Nuerk, & Fischer, 2008). Similarly, the NDE is thought to rep-
resent a noisy, analog representation of number (Dehaene, 1992), and thus is purported to occur because as numerical distance increases, the amount of representational overlap (and hence response competition) decreases, resulting in faster RTs (Gevers, Verguts, Reynvoet, Caessens, & Fias, 2006).

As indices of numerical representation, both the SNARC effect and the NDE have been used to study individual differences in number representation (Holloway & Ansari, 2009; Viarouge, Hubbard, & McCandliss, 2014). For example, Viarouge et al. (2014) showed that participants with stronger SNARC effects tended to exhibit slower mental rotation processing speed. Similarly, Pinhas, Shaki, and Fischer (2014) found that the SNARC effect was positively associated with spatial operational momentum. The NDE has been used in a similar manner. Holloway and Ansari (2009) found that children with a larger NDE tended to score lower on standardized assessments of mathematical fluency and calculation skill. Fazio, Bailey, Thompson, and Siegler (2014) showed a similar pattern, namely that the NDE was negatively correlated with mathematics achievement, but not reading achievement. As such, it is clear that accurate measurement of the SNARC effect and NDE is a desirable objective for a wide range of research paradigms within mathematical cognition.

The classical method for measuring the SNARC effect and NDE for individuals is based on a procedure known as regression coefficient analysis (RCA; Lorch & Myers, 1990). In this procedure, a dependent measure such as RT is regressed against a predictor for each individual participant. The estimated slope and intercept for each individual is then recorded, at which point any number of analyses may be performed. One of the first examples of this procedure in numerical cognition comes from Fias (1996), who adapted this method to measure the SNARC effect. Participants’ left-hand and right-hand RTs for each number stimulus (e.g., 1, 2, 8, 9) were aggregated via the median. Then, the difference between right- and left-hand RTs ($dRT$) was computed for each number. The SNARC effect was then presented as a negative correlation between number and $dRT$. That is, for small numbers (e.g., 1, 2), the left hand is faster, so $dRT > 0$. For large numbers, the right hand is faster, so $dRT < 0$. This negative correlation is captured by a linear regression, which is used to calculate the slope on $dRT$ for each participant. Analyses of the NDE are similar, but instead of using $dRT$, one typically regresses RT against a predictor such as numerical distance (e.g., Holloway & Ansari, 2009) or ratio (e.g., Fazio et al., 2014). As with the SNARC effect, the NDE typically presents as a negative slope. These slopes can then be submitted to a variety of analyses, including group-level tests (e.g., does the SNARC/NDE effect vary by group?) or individual-level correlations (i.e., is the SNARC/NDE effect correlated with mathematical performance?).

In the following sections, I describe a hierarchical Bayesian approach to measuring the SNARC effect and the NDE. Certainly, a complete treatment of Bayesian methods is beyond the scope of this paper, but the interested reader is advised consult the excellent books by Gelman, Carlin, Stern, and Rubin (2013), McElreath (2015), or Kruschke (2015) for more details. However, I will briefly outline some advantages to using such an approach to measuring numerical representations. One such advantage is that in a hierarchical model, variability from participants and items is modeled simultaneously, resulting in in both participant-level and group-level parameter estimates. As such, the hierarchical model allows the researcher to avoid aggregating single estimates across individuals, which can be problematic (Haider & Frensch, 2002; Heathcote, Brown, & Mewhort, 2000).
Pratte, Rouder, & Morey (2010). Another advantage is both philosophical and pragmatic. Bayesian modeling provides a principled means to combine prior knowledge with new data via Bayes’ theorem. The result is a quantification of posterior belief in terms of a probability distribution that reflects uncertainty after seeing data. From this distribution, one can estimate a variety of summary statistics, including the posterior mode (i.e., the parameter estimate with highest density), and Bayesian credible intervals, which give a range of values containing a particular probability mass. One particular type of credible interval used in a Bayesian context is the highest posterior density interval (HPDI), which is the narrowest interval containing a given probability mass.

Generally, the models will be constructed as follows. As with the classical method described above, the goal is to estimate a slope on \(dRT\) (for the SNARC effect) or \(RT\) (for the NDE). However, the proposed models are hierarchical, which means that individual-level slopes are drawn from a group-level distribution that specifies how the slopes are distributed in the population. Thus, individual-level and group-level estimates are modeled simultaneously, which improves upon the noisy slope estimates usually obtained from classical regression with small samples (Gelman & Hill, 2013). Further, since the model is Bayesian, one can specify group-level prior distributions that will moderate extreme individual-level parameter estimates (i.e., shrinkage), which will result in increased measurement accuracy for our parameter estimates.

**Model 1: estimating the SNARC effect**

The model is a hierarchical Bayesian linear regression model, where \(dRT\) is predicted by stimulus number. A graphical representation of the model can be seen in Figure 1. Formally, we define

\[
dRT_{ij} = \alpha_i + \beta_i(j) + \varepsilon_{ij}
\]

(1)

where \(i = \text{subject number}\) and \(j = \text{stimulus number}\). The residuals \(\varepsilon_{ij}\) are assumed to be normally distributed with mean 0 and precision \(1/\sigma^2\). Thus, we can express the likelihood for the data as

\[
dRT_{ij} \sim \text{Normal}(\mu_{ij}, 1/\sigma^2)
\]

(2)

where \(\mu_{ij} = \alpha_i + \beta_i(j)\). The prior for each individual-level intercept \(\alpha_i\) is set to be uniformly distributed between -200 and 200. The hierarchical structure is instantiated on slope. First, I set the prior for each individual-level slope \(\beta_i\) to be normally distributed with two hyperparameters: mean \(b\) and precision \(1/\sigma^2\). In turn, this requires priors on the hyperparameters – the prior for \(b\) is uniformly distributed between -20 and 20. Note that this range is based on inspection of the slopes obtained in Viarouge et al. (2014). The prior for group-level slop precision \(1/\sigma^2\) (as well as group-level residual precision \(1/\sigma^2\)) is set as a Gamma distribution with both shape and scale equal to 0.01.

---

1Bayesian credible intervals are not the same thing as frequentist confidence intervals, which are defined as a range of parameters that would contain a population parameter in some specified percentage of a large number of repeated samples. While people often misinterpret confidence intervals using probability statements (Hoekstra, Morey, Rouder, & Wagenmakers, 2014), Bayesian credible intervals can be directly interpreted as probability statements.
**Figure 1.** Graphical model of the hierarchical Bayesian linear regression model for the SNARC effect. Following the convention of Lee and Wagenmakers (2014), nodes represent variables of interest (observed=shaded, latent=unshaded), with dependencies represented via the graph structure.

### Fitting the model

**Data.** I fit the model to data collected from 35 participants in a number parity task. The numbers 1, 2, 8, and 9 were presented in the center of a computer screen, after which participants were asked to quickly indicate via a button press whether the number was even or odd. The procedure mirrored that of Experiment 1 of Pinhas et al. (2014). Participants completed 112 trials of the task under two counterbalanced response rules (either even=left or even=right). This resulted in a collection of 3,920 trials. We removed 259 error trials and an additional 12 trials for which RT exceeded 3 seconds (a total of 6.9% of trials). The remaining 3,649 trials were collapsed into $35 \times 8 \times 2$ cells by computing median RT for each of the conditions defined by crossing the factors of subject, stimulus number, and response hand. Then, $dRT_{ij}$ was computed for each combination of subject and number by subtracting left-hand RT from right-hand RT.

**Results.** The regression model parameters were estimated using R (R Core Team, 2016) and JAGS (Plummer, 2017). Posterior sampling consisted of 3 MCMC chains, each containing 100,000 draws. The first 5000 draws of each chain were discarded as “burn-in” samples, leaving 285,000 samples remaining. These remaining samples were thinned by a factor of 10, leaving a final sample of 28,500 posterior draws for each parameter. Visual inspection of trace plots indicated that all chains converged appropriately. Additionally, the Gelman-Rubin statistic $\hat{R} = 1.001$ for each parameter, indicating that the Markov chains for each parameter converged to the appropriate stationary distribution (Gelman et al., 2013; Gelman & Rubin, 1992).

Since the model is hierarchical, I was able to estimate posterior distributions for each
Figure 2. Posterior distribution of the group-level slope $b$ for Model 1, indicating a substantial SNARC effect across the group.

The posterior distribution of the group-level slope $b$ is depicted in Figure 2. As can be seen in the figure, the mass of the distribution is centered over the posterior mode $b = -11.5$. Further, the 95% HPDI for the slope $b$ is $[-15.6, -7.3]$. Finally, since the posterior distribution is a probability distribution, we can compute the probability that $b < 0$ (that is, the probability that there is a non-zero SNARC effect). This probability is greater than 0.999. Other Bayesian tools can be applied to this model, such as computing a Bayes factor comparing a null-SNARC model to our obtained model via the Savage-Dickey density ratio (Wagenmakers, Lodewyckx, Kuriyal, & Grasman, 2010). Briefly, this method amounts to computing the density of the value $b = 0$ in the posterior distribution divided by the density of $b = 0$ in the prior (i.e., Uniform($-20, 20$)). Doing this resulted in a Bayes factor of approximately 280,000 to 1 in favor of the alternative hypothesis, indicating overwhelming support for an overall SNARC effect.
Model 2: estimating the numerical distance effect

As with Model 1, this model is also a hierarchical Bayesian linear regression model, where \( RT \) is predicted by numerical distance. For this specific instantiation of the model, I will index numerical distance via the ratio between compared numbers. For example, Fazio et al. (2014) used four ratio bins as predictors of RT. A graphical representation of the model can be seen in Figure 3. Formally, we define

\[
RT_{ij} = \alpha_i + \beta_i(j) + \epsilon_{ij} \tag{3}
\]

where \( i \) = participant number and \( j = 1, \ldots, 4 \) equals ratio bin number. The residuals \( \epsilon_{ij} \) are again assumed to be normally distributed with mean 0 and precision \( 1/\sigma^2 \). The likelihood for the data is

\[
RT_{ij} \sim \text{Normal}(\mu_{ij}, 1/\sigma^2) \tag{4}
\]

where \( \mu_{ij} = \alpha_i + \beta_i(j) \). The prior for each individual-level intercept \( \alpha_i \) is set to be uniformly distributed between 0 and 2000. The prior for each individual-level slope \( \beta_i \) to be normally distributed with two hyperparameters: mean \( b \) (with uniform prior between -100 and 100) and precision \( 1/\sigma^2_b \) (with a Gamma(0.01, 0.01) prior). Finally, the prior for the group-level residual precision is Gamma(0.01, 0.01).

Fitting the model

Data. I fit the model to the RT data from Fazio et al. (2014). Fifty-five 5th graders completed a symbolic number comparison task, in which they were asked to choose which
Figure 4. Posterior distribution of the group-level slope $b$ for Model 2, indicating a substantial numerical distance effect across the group.

Results. The regression model was fit using the same procedure as in Model 1. As before, visual inspection of trace plots indicated that all chains converged appropriately, with $\hat{R} = 0.001$ for all parameters.

The posterior distribution of the group-level slope $b$ is depicted in Figure 4. As can be seen in the figure, the mass of the distribution is centered over the posterior mode $b = -65.9$, with 95% HPDI for the slope $b$ equal to $[-83.4, -50.7]$. In addition to estimating the group-level slope, we can test the existence of the numerical distance effect via a Bayes factor, which as before I computed using the Savage-Dickey density ratio. This Bayes factor was approximately 1.7 million to 1 in favor of the alternative hypothesis, indicating overwhelming support for an overall numerical distance effect.
Comparing the model to the classical approach

In this section, I will demonstrate that the hierarchical Bayesian model developed in this paper can provide better measurement fidelity for the group-level regression slopes that are desired when assessing group-level SNARC effects or NDEs, particularly for small samples. As mentioned earlier, the reason for this increased accuracy is because of a property that is unique to Bayesian inference – namely the property of shrinkage. When one specifies a group-level prior distribution for the slope (e.g., the Uniform(-20, 20) distribution that I used in Model 1), this prior is then combined with the data likelihood via Bayes theorem by multiplication. In our case, extreme individual slope estimates (those beyond -20 and 20) vanish by virtue of being multiplied by the prior probability of obtaining those estimates (i.e., probability = 0). The resulting posterior distribution is then shrunk away from these parameter estimates.

To demonstrate this, I conducted a simulation. I randomly generated $d_{RT}$ values for a small samples of $n = 15$ simulated participants as follows. First, I generated 15 random slopes $b_i$, where $b_i \sim \text{Gaussian}(-10, 1)$. That is, I assumed that individual slopes are drawn from a normal distribution centered at -10, with standard deviation 1. Similarly, I randomly generated 15 random intercepts $a_i$, where $a_i \sim \text{Uniform}(-200, 200)$. Then, I generated

$$d_{RT_{ij}} = a_i + b_i(j) + \varepsilon_{ij}$$

where $i = 1, \ldots, 15$, $j = 1, 2, 8, 9$, and $\varepsilon_{ij} \sim \text{Gaussian}(0, 100)$. Then, I computed a classical linear regression for each participant $i = 1, \ldots, 15$, recording each estimated slope $\hat{b}_i$. Finally, I fit the hierarchical Bayesian model (Model 1) to the overall set of data to compute the posterior distribution of $b$.

The results can be seen in Figure 5. Notice that while both the frequentist 95% confidence interval and the Bayesian 95% highest posterior density interval contain the population mean $b = -10$, the Bayesian estimate is much less variable. This is because compared to the classical linear regression method (e.g., Lorch & Myers, 1990), the extreme parameter estimates are shrunk toward the posterior mean. The same story is repeated with hypothesis testing, as well. Indeed, performing a $t$-test on the individual regression slopes (as in Fias, 1996) results in a non-significant SNARC effect, $t(14) = -1.82$, $p = 0.09$, which results in a Type II error. However, a Bayesian hypothesis test (the Savage-Dickey method) yields a Bayes factor that favors the SNARC effect by a factor of 75 to 1.

Discussion

The purpose of this paper was to introduce a hierarchical Bayesian linear regression framework for measuring group-level and individual-level numerical representations. The models were then applied to two common markers of numerical representations: the SNARC effect, which indexes spatial-numerical associations, and the numerical distance effect (NDE), which indexes representations of numerical magnitude. The models developed here represent a straightforward Bayesian extension of the classical linear regression approach to measuring these effects that was introduced by Lorch and Myers (1990) and first applied to the SNARC effect by Fias (1996).

Both models were built from a hierarchical Bayesian framework, which is advantageous for several reasons. First, the models assume that the regression slopes (the primary
Figure 5. Slope estimates from a simulated data set of \( n = 15 \) participants. The solid line represents the posterior density of \( b \), the group-level slope estimate for the SNARC effect. The histogram represents the slope estimates obtained from separate classical linear regressions for each individual.

object of interest in these analyses) are drawn from a group-level distribution. This hierarchical definition allows both group-level and individual-level slope estimates to be modeled simultaneously from the data. As such, we can answer questions at either the group-level (i.e., is there a SNARC effect?) or participant-level (i.e., are the slopes associated with other measures, such as math achievement?). For example, one might use this framework as a stepping stone to a more complex model, where the group-level slopes \( b \) are hypothesized to differ by some independent variable. A concrete instance of this comes from studies in which the authors found group differences in the SNARC effect (e.g., Cipora et al., 2015; Fischer, Mills, & Shaki, 2010). One could perform similar studies by building an additional linear model on these group-level parameters, with priors appropriate to such effects (e.g., a Cauchy prior, as in Rouder, Speckman, Sun, Morey, & Iverson, 2009). Importantly, since the inference would be done in a Bayesian framework, it is possible to measure evidence for null effects too (Wagenmakers, 2007), so the model could be used to test for invariances as well as differences.

Another advantage to using a hierarchical Bayesian model for numerical representations is that such models tend to have better measurement fidelity. Indeed, one of the advantages of the Bayesian framework in general is the notion of shrinkage, where extreme parameter estimates are “shrunk” toward the mean by virtue of the prior (Gelman et al., 2013). This property is particularly salient for small samples, where classical frequentist
methods tend to perform poorly. I demonstrated exactly this phenomenon in the simulation above: because of some extreme individual-level slope estimates, the frequentist confidence interval was quite wide, and as a result, we could not detect the SNARC effect. However, the Bayesian highest posterior density interval provided a much more accurate estimate of the true population slope. Critically, in this simulation, only the Bayesian method produced the correct inference.

A final advantage to this modeling framework that I will mention is that all assumptions of the model are made explicit. This may be a new approach to some, especially those who are accustomed to classical inferential software packages for which the statistical assumptions are kept “under the hood.” However, I think the approach presented in this paper can be very useful to a wide variety of problems in cognitive psychology, as the researcher can take the models presented here and modify them for any desired context. Indeed, the prior knowledge of a given field can be easily and coherently integrated into the model without too much work.

In summary, the models developed in this paper will provide a flexible tool that can be used to estimate group-level and individual-level numerical representations quickly and coherently. Further, the methods developed are not specific to numerical cognition, so application to a wide variety of problems should not be too difficult. Regardless of the application, the hierarchical Bayesian linear regression framework presented here provides a powerful, coherent, and accurate measurement model for applied work in cognitive psychology.

References

Cipora, K., Hohol, M., Nuerk, H.-C., Willmes, K., Brożek, B., Kucharzyk, B., & Nęcka, E. (2015). Professional mathematicians differ from controls in their spatial-numerical associations. Psychological Research, 80(4), 710–726. doi: 10.1007/s00426-015-0677-6

Dehaene, S. (1992). Varieties of numerical abilities. Cognition, 44(1-2), 1–42. doi: 10.1016/0010-0277(92)90049-n

Dehaene, S., Dupoux, E., & Mehler, J. (1990). Is numerical comparison digital? Analogical and symbolic effects in two-digit number comparison. Journal of Experimental Psychology: Human Perception and Performance, 16(3), 626–641. doi: 10.1037/0096-1523.16.3.626

Fazio, L. K., Bailey, D. H., Thompson, C. A., & Siegler, R. S. (2014). Relations of different types of numerical magnitude representations to each other and to mathematics achievement. Journal of Experimental Child Psychology, 123, 53–72. doi: 10.1016/j.jecp.2014.01.013

Fias, W. (1996). The importance of magnitude information in numerical processing: Evidence from the SNARC effect. Mathematical Cognition, 2(1), 95–110. doi: 10.1080/135467996387552

Fischer, M. H., Mills, R. A., & Shaki, S. (2010). How to cook a SNARC: Number placement in text rapidly changes spatial–numerical associations. Brain and Cognition, 72(3), 333–336. doi: 10.1016/j.bandc.2009.10.010

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2013). Bayesian data analysis (3rd ed.). Chapman & Hall/CRC, Boca Raton, FL.

Gelman, A., & Hill, J. (2013). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. Statistical Science, 7(4), 457–472. doi: 10.1214/ss/1177011136

Gevers, W., Verguts, T., Reynvoet, B., Caessens, B., & Fias, W. (2006). Numbers and space: A computational model of the SNARC effect. Journal of Experimental Psychology: Human Perception and Performance, 32(1), 32–44. doi: 10.1037/0096-1523.32.1.32
Haider, H., & Frensch, P. A. (2002). Why aggregated learning follows the power law of practice when individual learning does not: Comment on Rickard (1997, 1999), Delaney et al. (1998), and Palmeri (1999). *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(2), 392–406. doi: 10.1037//0278-7393.28.2.392

Heathcote, A., Brown, S., & Mewhort, D. J. K. (2000). The power law repealed: The case for an exponential law of practice. *Psychonomic Bulletin & Review, 7*(2), 185–207. doi: 10.3758/bf03212979

Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (2014). Robust misinterpretation of confidence intervals. *Psychonomic Bulletin & Review, 21*(5), 1157–1164. doi: 10.3758/s13423-013-0572-3

Holloway, I. D., & Ansari, D. (2009). Mapping numerical magnitudes onto symbols: The numerical distance effect and individual differences in children’s mathematics achievement. *Journal of Experimental Child Psychology, 103*(1), 17–29. doi: 10.1016/j.jecp.2008.04.001

Kruschke, J. K. (2015). *Doing Bayesian data analysis* (2nd ed.). Academic Press, London.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge university press.

Lorch, R. F., & Myers, J. L. (1990). Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16*(1), 149–157. doi: 10.1037//0278-7393.16.1.149

McElreath, R. (2015). *Statistical rethinking*. Chapman & Hall/CRC, Boca Raton, FL.

Moyer, R. S., & Landauer, T. K. (1967). Time required for judgements of numerical inequality. *Nature, 215*, 1519-1520. doi: 10.1038/2151519a0

Pinhas, M., Shaki, S., & Fischer, M. H. (2014). Heed the signs: Operation signs have spatial associations. *The Quarterly Journal of Experimental Psychology, 67*(8), 1527–1540. doi: 10.1080/17470218.2014.892516

Plummer, M. (2017). JAGS: Just Another Gibbs Sampler [Computer software manual]. Retrieved from https://mcmc-jags.sourceforge.net

Pratte, M. S., Rouder, J. N., & Morey, R. D. (2010). Separating mnemonic process from participant and item effects in the assessment of ROC asymmetries. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*(1), 224–232. doi: 10.1037/a0017682

R Core Team. (2016). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian t tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review, 16*(2), 225–237. doi: 10.3758/pbr.16.2.225

Viarouge, A., Hubbard, E. M., & McCandliss, B. D. (2014). The cognitive mechanisms of the SNARC effect: An individual differences approach. *PLoS ONE, 9*(4), e95756. doi: 10.1371/journal.pone.0095756

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of p values. *Psychonomic Bulletin & Review, 14*(5), 779–804. Retrieved from https://doi.org/10.3758/bf03194105

Wagenmakers, E.-J., Lodewyckx, T., Kuriyal, H., & Grasman, R. (2010). Bayesian hypothesis testing for psychologists: A tutorial on the Savage-Dickey method. *Cognitive Psychology, 60*(3), 158–189. doi: 10.1016/j.cogpsych.2009.12.001

Wood, G., Willmes, K., Nuerk, H.-C., & Fischer, M. H. (2008). On the cognitive link between space and number: A meta-analysis of the SNARC effect. *Psychology Science Quarterly, 50*(4), 489–525.