

# Supplemental Appendix: Personality and Prosocial Behavior: A Multilevel Meta-Analysis

Reuben Kline  
Stony Brook University

Alexa Bankert  
University of Georgia

Lindsey Levitan  
Shepherd University

Patrick Kraft  
Stony Brook University

## Contents

<b>1 Detailed description of prosociality measure</b>	<b>2</b>
<b>2 Summary of the Data</b>	<b>4</b>
<b>3 Replication with Frequentist Models</b>	<b>6</b>
<b>4 Additional Information on Bayesian Models</b>	<b>8</b>
4.1 STAN code	8
4.2 Traceplots for Main Models	11
<b>5 Robustness checks and additional model results</b>	<b>13</b>
<b>6 Additional Types of Multilevel Meta-Analysis</b>	<b>13</b>

# 1 Detailed description of prosociality measure

Our main results—presented in Table 3 and Figure 1 in the main text—are derived from a model in which prosociality outcomes have been normalized across a number of different types of games and parameter values. In the context of our analyses, we call this variable *svo1*. In Table A.1 we provide detailed information about each of these normalizations.

To test the robustness of our results, we created several alternative outcome measures. It could be argued that prosocial behavior by dictators in DG, responders in UG and trustees in TG could be understood as giving at least *half* of their endowment rather than the full amount. We therefore rescaled the variable for distributive games such that the maximum value of 1 corresponds to distributing half or more of the available amount to the other player, with values below half remaining scaled between 0 and 1. We call this measure *svo2*. More formally, if  $svo1 \geq 0.5$ ,  $svo2 = 1$ , otherwise  $svo2 = 2 * svo1$  for the game types described above. The results for models 1-3 estimated with *svo2* as the dependent variable are plotted in A.1.

More generally, one might worry that the aggregation of outcome measures from separate studies with different games and experimental parameters into a common continuous (or semi-continuous) variable might require too strong assumptions about the equivalence of individual measures. In order to alleviate this concern, we created two dichotomous variables of prosociality by rounding the original measure for each study to 0 or 1 (for one specification, 0.5 was assigned the value of 0, and for the other, 0.5 was assigned the value of 1). We call these outcome measures *svoD1* and *svoD2*. These dichotomizations also apply to several additional studies that employ iterated PD, such as Hirsh and Peterson [2009] and Pothos et al. [2011]. More formally, for all games, if  $svo1 \geq 0.5$ ,  $svoD1 = 1$ , otherwise  $svoD1 = 0$ , and, if  $svo1 > 0.5$ ,  $svoD1 = 1$ , otherwise  $svoD1 = 0$ . The results for models 1-3 estimated as multi-level logistic regressions with *svoD1* and *svoD2* as the dependent variables are plotted in A.2.

Table A.1: **Normalization Procedure for Prosociality Outcome Measure**

Study	Game	Prosociality Normalization ( <b>svo1</b> )
Artinger et al. [2014]	DG	Maximum giving is 100: <b>svo1</b> = Amount given / 100
Ben-Ner and Kramer [2011]	DG	Each subject made 74 hypothetical dictator decisions dividing \$10. <b>svo1</b> = mean amount given across these 74 decisions.
Ben-Ner et al. [2004]	DG	<b>svo1</b> = Amount given / 10.
Brocklebank et al. [2011]	DG	<b>svo1</b> = mean of 6 binary forced choice “dictator games”, taken from Charness and Rabin [2002]—Berk17, Berk23, Berk29, Berk15, Berk26, Ed128. Each item is coded as 1 for the prosocial choice and 0 for the selfish choice. This is the same measure of prosociality that is used by the authors.
Fischbacher et al. [2001]	PGG	Linear PGG with 20 tokens maximum contribution: <b>svo1</b> = contribution / 20
Gunnthorsdottir et al. [2002]	TG	Binary choice TG. P1 can choose to send nothing, then payoff is 10,10. If P1 sends, then P2 can choose either 0,40 or 15,25. So, we only look at 2’s payoffs. If 2’s payoff is 40, then <b>svo1</b> = 0, if 2’s payoff is 25, then <b>svo1</b> = 1. Otherwise it is missing.
Hilbig and Zettler [2009]	DG	Maximum giving is 100: <b>svo1</b> = Amount given / 100
Hilbig et al. [2012a]	DG	<b>svo1</b> = mean of 9 binary or ternary forced choice “dictator games”, each with 3 options, distinguishing among competitive, individualistic and prosocial preferences, with the former two choices being coded as selfish.
Hilbig et al. [2012b]	PGG	There are two experiments/studies. In each, subjects play a (hypothetical) standard linear public goods game, each with max contribution of 100. <b>svo1</b> = sum of contributions / 200.
Hirsh and Peterson [2009]	PD	Iterated PD. Each player made 10 decisions, with standard PD coding. <b>svo1</b> = mean of 10 PD decisions. We dropped the 25 x 10 “decisions” made by the 25 confederates.
Koole et al. [2001]	CPRD	12 round CPRD, with maximum extraction 10 in each round (most selfish). <b>svo1</b> = 1 - average extraction.
Kurzban and Houser [2001]	PGG	Linear PGG with maximum contribution of 50: <b>svo1</b> = contribution/50.
Pothos et al. [2011]	PD	Study uses 12 variations of PD. For 6 “cooperation” is prosocial, and for another 6 “defection” is prosocial. We adjust the latter to be coded as defect = 1 and <b>svo1</b> = mean of these 12 decisions.
Schmitt et al. [2004]	UG	Authors use strategy method to elicit minimum acceptable offers (MAO) from responders. Total amount to split is \$15: <b>svo1</b> = (1-MAO)/15)
Swope et al. [2008]	DG, TG, PD	Authors use between subjects design to look at 4 games (no MAO data for UG so we could not use it). <b>ProsocDG</b> = amount given / 15. <b>ProsocPD</b> = 1 for cooperation and 0 for defection. For TG we just use the trustee’s data. Money sent by trustor (max \$10) is tripled and given to trustee: <b>ProsocTG</b> = amount returned / (3 x money given)

Notes: CPRD = common pool resource dilemma; DG = dictator game; PD = prisoner’s dilemma; PGG = public goods game; TG = trust game UG = ultimatum game. **svo1** is our prosociality outcome measure.

## 2 Summary of the Data

In tables A.2, A.3, and A.4 we present some descriptive statistics for the variables in our three datasets.

Table A.2: Data for Model 1 (12 Studies)

Statistic	N	Mean	St. Dev.	Min	Max
study	2,235	5.510	2.910	1	12
extro	2,235	0.608	0.171	0.000	1.000
open	2,235	0.615	0.178	0.000	1.000
agree	2,235	0.530	0.177	0.000	1.000
conscien	2,235	0.609	0.169	0.000	1.000
payment	2,235	0.350	0.477	0	1
hilbig	2,235	0.487	0.500	0	1
benner	2,235	0.232	0.422	0	1
mbti	2,235	0.114	0.317	0	1
coop	2,235	0.349	0.477	0	1
svo1	2,235	0.462	0.342	0.000	1.000
svo2	2,235	0.568	0.382	0.000	1.000
svoD1	2,235	0.490	0.500	0	1
svoD2	2,235	0.392	0.488	0	1

Table A.3: Data for Model 2 (10 Studies)

Statistic	N	Mean	St. Dev.	Min	Max
study	1,981	4.674	2.226	1	10
extro	1,981	0.622	0.152	0.000	1.000
open	1,981	0.629	0.161	0.000	1.000
agree	1,981	0.544	0.162	0.000	1.000
conscien	1,981	0.617	0.155	0.000	1.000
neuro	1,981	0.523	0.166	0.000	1.000
payment	1,981	0.267	0.443	0	1
hilbig	1,981	0.550	0.498	0	1
benner	1,981	0.261	0.440	0	1
coop	1,981	0.369	0.483	0	1
svo1	1,981	0.449	0.344	0.000	1.000
svo2	1,981	0.535	0.375	0.000	1.000
svoD1	1,981	0.462	0.499	0	1
svoD2	1,981	0.374	0.484	0	1

Table A.4: Data for Model 3 (15 Studies)

Statistic	N	Mean	St. Dev.	Min	Max
study	2,482	6.658	3.494	1	15
payment	2,482	0.415	0.493	0	1
hilbig	2,482	0.439	0.496	0	1
benner	2,482	0.209	0.406	0	1
coop	2,482	0.340	0.474	0	1
svo1	2,482	0.450	0.344	0.000	1.000
svo2	2,482	0.556	0.383	0.000	1.000
svoD1	2,482	0.466	0.499	0	1
svoD2	2,482	0.375	0.484	0	1

### 3 Replication with Frequentist Models

Here we replicate the results in table 3 with the more standard frequentist approach to multilevel modeling. In place of the 95% credible intervals we present in table 3 we report 95% *confidence* intervals to be consistent with the presentation in table 3. At the individual level (level 1), the point estimates and even the standard errors for the personality traits are essentially unchanged—in sum, openness and agreeableness are still associated with significantly increased prosocial behavior, whereas the other three personality traits are not. At level 2 however, as Stegmueller [2013] shows, both the point estimates and the standard errors exhibit more variability between the two methods. In general, and as to be expected, the standard errors for the level 2 variables are much smaller here than in the Bayesian analysis reported in the text. As a result, in model 1 but not model 2, the estimate for Hilbig is different from zero. However, these standard errors are likely underestimated, and therefore the Bayesian estimates are more credible. The frequentist models were estimated using the `lme4` package in R.

Table A.5: Frequentist replication of main models (multilevel regression)

	<i>Dependent variable:</i>		
		Prosociality	
	(1)	(2)	(3)
Extraversion	-0.017 (-0.096, 0.062)	-0.061 (-0.155, 0.034)	
Openness	0.136 (0.053, 0.219)	0.233 (0.134, 0.332)	
Agreeableness	0.123 (0.039, 0.208)	0.161 (0.061, 0.262)	
Conscientiousness	-0.053 (-0.134, 0.029)	-0.072 (-0.168, 0.025)	
Neuroticism		-0.054 (-0.144, 0.036)	
Payment	0.181 (-0.025, 0.387)	0.178 (-0.040, 0.397)	0.052 (-0.113, 0.217)
Author: Hilbig	0.254 (0.017, 0.491)	0.236 (-0.041, 0.513)	
Author: Ben-Ner	-0.045 (-0.282, 0.192)	-0.083 (-0.384, 0.218)	
MBTI Personality Measure	0.169 (-0.074, 0.413)		
Cooperative	0.076 (-0.022, 0.174)	0.036 (-0.198, 0.271)	0.071 (-0.021, 0.163)
(Global) Intercept	0.124 (-0.087, 0.335)	0.143 (-0.159, 0.446)	0.359 (0.224, 0.495)
Observations	2,235	1,981	2,482

## 4 Additional Information on Bayesian Models

### 4.1 STAN code

For all three main models, the following STAN code was used:

```
data {
  int<lower=0> n;          // number of obs
  int<lower=0> m;          // number of studies
  int<lower=0> nX;         // number of individual-level parameters
  int<lower=0> nZ;         // number of study-level parameters
  matrix[n,nX] X;        // predictor matrix individual level
  matrix[m,nZ] Z;        // predictor matrix study level
  vector[n] svo;         // dependent variable
  int<lower=0> study[n];  // study indicator
}
parameters {
  vector[m] beta0;
  vector[nX] beta;
  vector[nZ] gamma;
  real<lower=0> sigma_beta0;
  real<lower=0> sigma_beta;
  real<lower=0> sigma_svo;
}
model {
  vector[n] svo_hat;

  for (i in 1:n)
    svo_hat[i] <- beta0[study[i]] + X[i] * beta;
  beta0 ~ normal(Z * gamma, sigma_beta0);
  svo ~ normal(svo_hat, sigma_svo);

  beta ~ normal(0,1);
  gamma ~ normal(0,1);
  //sigma_beta0 ~ cauchy(0,5);
  //sigma_svo ~ cauchy(0,5);
}
```



For the varying-slopes model, the STAN code was adjusted as follows:

```
data {
  int<lower=0> n;          // number of obs
  int<lower=0> m;          // number of studies
  int<lower=0> nX;         // number of individual-level parameters
  int<lower=0> nZ;         // number of study-level parameters
  matrix[n,nX] X;        // predictor matrix individual level
  matrix[m,nZ] Z;        // predictor matrix study level
  vector[n] svo;         // dependent variable
  int<lower=0> study[n]; // study indicator
}
parameters {
  vector[m] beta0;
  vector[m] beta_extro;
  vector[m] beta_open;
  vector[m] beta_agree;
  vector[m] beta_conscien;
  vector[4] mu_beta;
  vector<lower=0>[4] sigma_beta;
  vector[nZ] gamma;
  real<lower=0> sigma_beta0;
  real<lower=0> sigma_svo;
}
model {
  vector[n] svo_hat;

  for (i in 1:n)
    svo_hat[i] <- beta0[study[i]] + X[i,1] * beta_extro[study[i]]
      + X[i,2] * beta_open[study[i]] + X[i,3] * beta_agree[study[i]]
      + X[i,4] * beta_conscien[study[i]];
  beta0 ~ normal(Z * gamma, sigma_beta0);
  svo ~ normal(svo_hat, sigma_svo);

  beta_extro ~ normal(mu_beta[1],sigma_beta[1]);
  beta_open ~ normal(mu_beta[2],sigma_beta[2]);
  beta_agree ~ normal(mu_beta[3],sigma_beta[3]);
  beta_conscien ~ normal(mu_beta[4],sigma_beta[4]);
  mu_beta ~ normal(0,1);
  gamma ~ normal(0,1);
  //sigma_beta0 ~ cauchy(0,5);
  //sigma_beta ~ cauchy(0,5);
  //sigma_svo ~ cauchy(0,5);
}
```

The multilevel logit, on the other hand, was specified as follows:

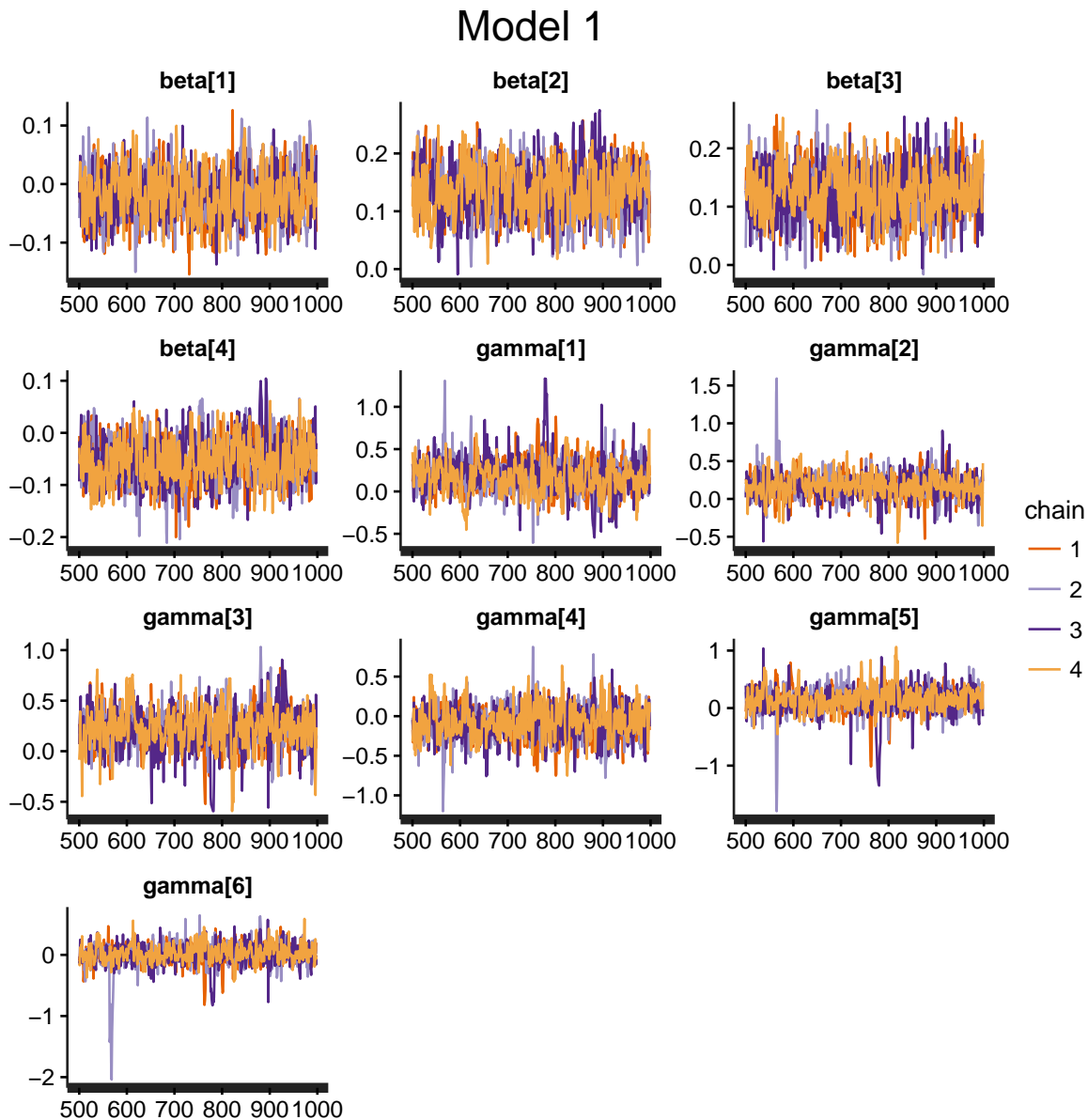
```
data {
  int<lower=0> n;          // number of obs
  int<lower=0> m;          // number of studies
  int<lower=0> nX;         // number of individual-level parameters
  int<lower=0> nZ;         // number of study-level parameters
  matrix[n,nX] X;        // predictor matrix individual level
  matrix[m,nZ] Z;        // predictor matrix study level
  int<lower=0,upper=1> svo[n]; // dependent variable
  int<lower=0> study[n]; // study indicator
}
parameters {
  vector[m] beta0;
  vector[nX] beta;
  vector[nZ] gamma;
  real<lower=0> sigma_beta0;
  real<lower=0> sigma_beta;
}
model {
  vector[n] svo_hat;

  for (i in 1:n)
    svo_hat[i] <- beta0[study[i]] + X[i] * beta;
  beta0 ~ normal(Z * gamma, sigma_beta0);
  svo ~ bernoulli_logit(svo_hat);

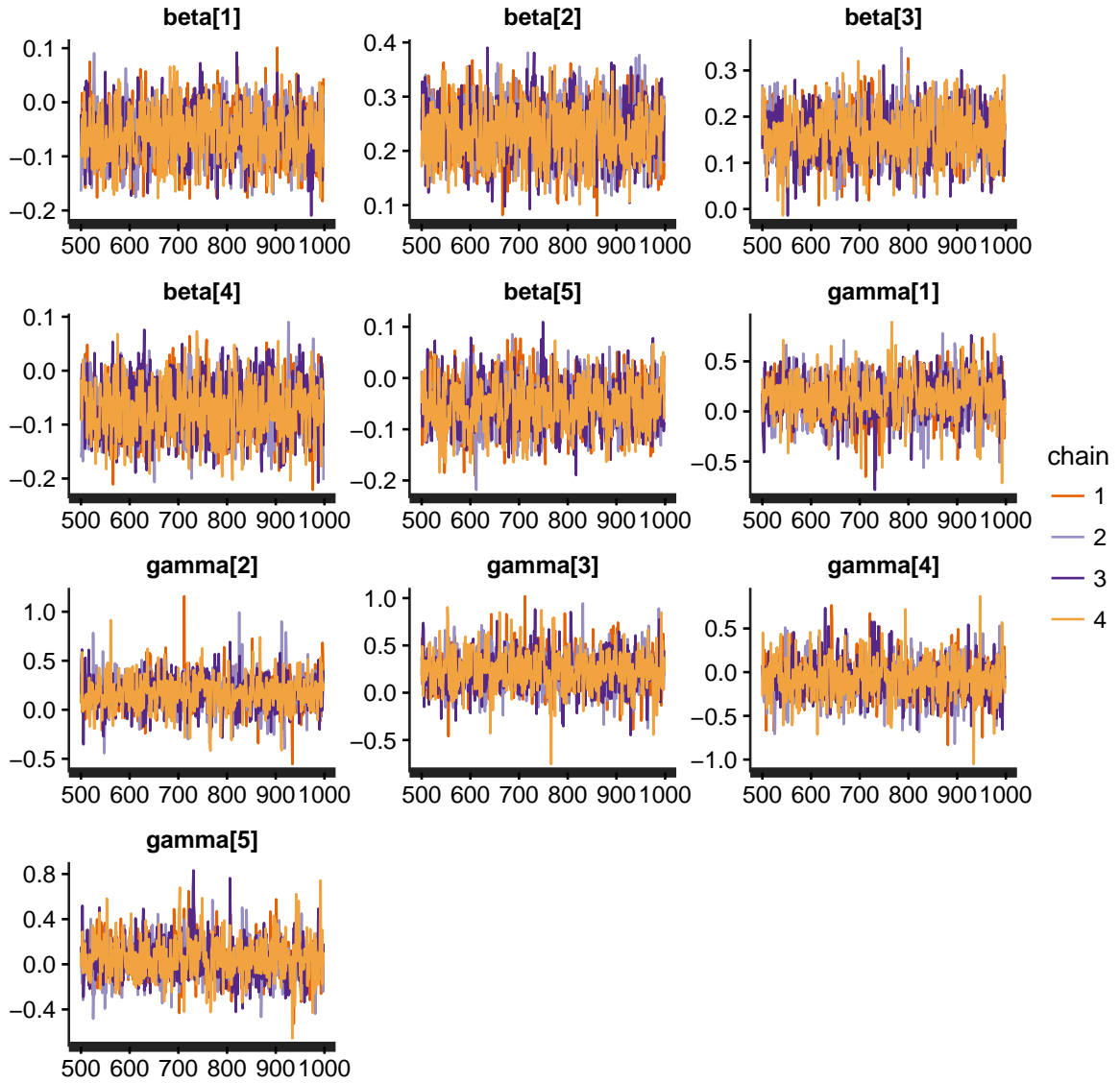
  beta ~ normal(0,5);
  gamma ~ normal(0,5);
  //sigma_beta0 ~ cauchy(0,5);
}
```

## 4.2 Traceplots for Main Models

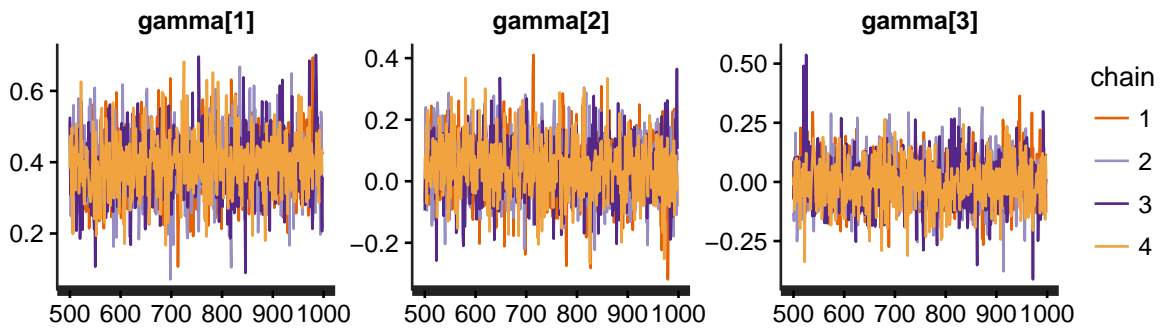
Here we present trace plots for our main models to monitor convergence. These figures display draws from the posterior distribution generated by the Hamiltonian Monte Carlo implemented in STAN. While the individual chain length is longer, the sampler is set to save only a subsample of draws (i.e. the chains are thinned) in order to reduce total memory usage. In total, we save 1000 draws for each parameter, where the first half of the chain is discarded as ‘burn-in’ or adaptation draws. The numbering of the parameters follows the order they are entered in the models above, with the “betas” representing the level 1 coefficients, and the “gammas” representing the level 2 coefficients.



## Model 2



## Model 3



## 5 Robustness checks and additional model results

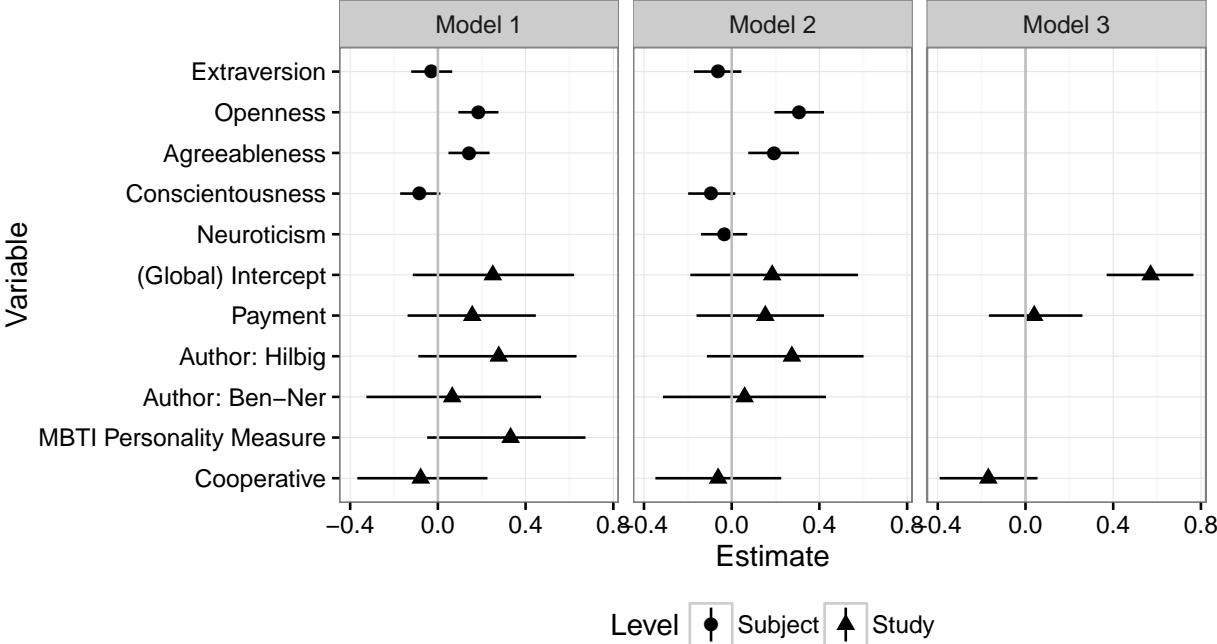


Figure A.1: **Multilevel Meta-Analysis Results: Predicting Prosocial Behavior, alternative specification of the dependent variable.** Dependent variable is *svo2*.  $svo2 = 1$  if  $svo1 \geq 0.5$ , otherwise  $svo2 = 2 * svo1$ . For each model and each parameter, the dot represents the posterior mean and the bar represents the 95% credible interval.

## 6 Additional Types of Multilevel Meta-Analysis

The approach described in point 3 in the text, namely the incorporation of study-level summary statistics such as treatment effects in the MLMA framework, is analogous to traditional meta-analysis. However, MLMA allows us to explicitly model the multilevel structure of the data by incorporating both the within- and between-studies error variance [Thompson et al., 2001]. We do not employ this approach, but nonetheless think it is important to introduce this type of MLMA because it shows how conventional meta-analytic approaches can be viewed as a special case of MLMA. Since it explicitly models the multilevel structure of the data, however, MLMA yields more efficient estimates. Efficiency is especially important in meta-analyses because they often have to rely on a small number of studies.

In a conventional meta-analysis, one would typically employ the following functional form:

$$\theta_j \sim N(\mu, \tau^2 + \sigma_j^2) \tag{1}$$

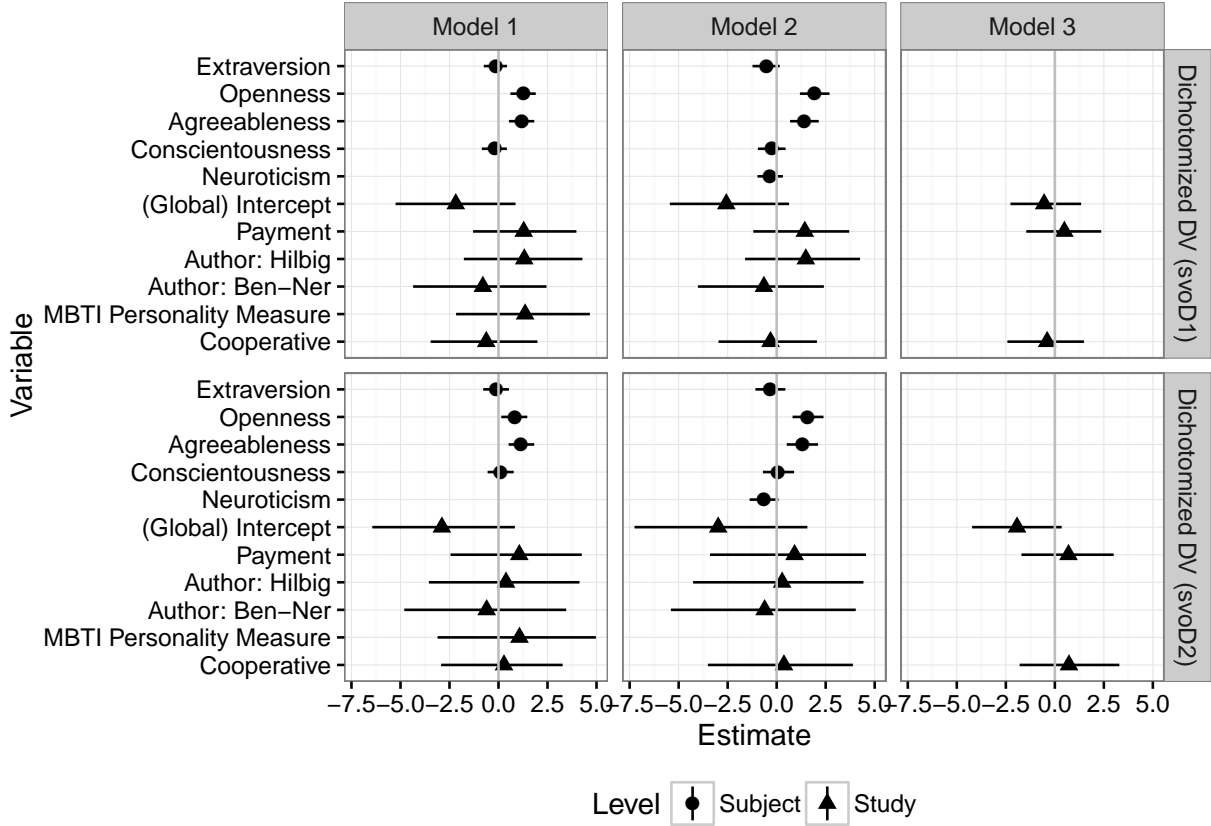


Figure A.2: **Multilevel Meta-Analysis Results: Predicting Prosocial Behavior, multilevel logit.** *svoD1* and *svoD2* are the dependent variables for the top and bottom panels, respectively.  $svoD1 = 1$  if  $svo1 \geq 0.5$ , otherwise  $svoD1 = 0$ .  $svoD2 = 1$  if  $svo1 > 0.5$ , otherwise  $svoD2 = 0$ . For each model and each parameter, the dot represents the posterior mean and the bar represents the 95% credible interval.

Where  $\theta_j$  is the observed summary statistic in study  $j$  (e.g. a reported treatment effect),  $\mu$  is the average effect across all studies,  $\sigma_j^2$  represents the within-study sampling variability, and  $\tau^2$  denotes the between study variability. This model, however, can be described in terms of a multilevel framework.

For level 1 (within study):

$$\theta_j \sim N(\mu_j, \sigma_j^2) \quad (2)$$

And for level 2 (between studies):

$$\mu_j \sim N(\mu, \tau^2), \quad (3)$$

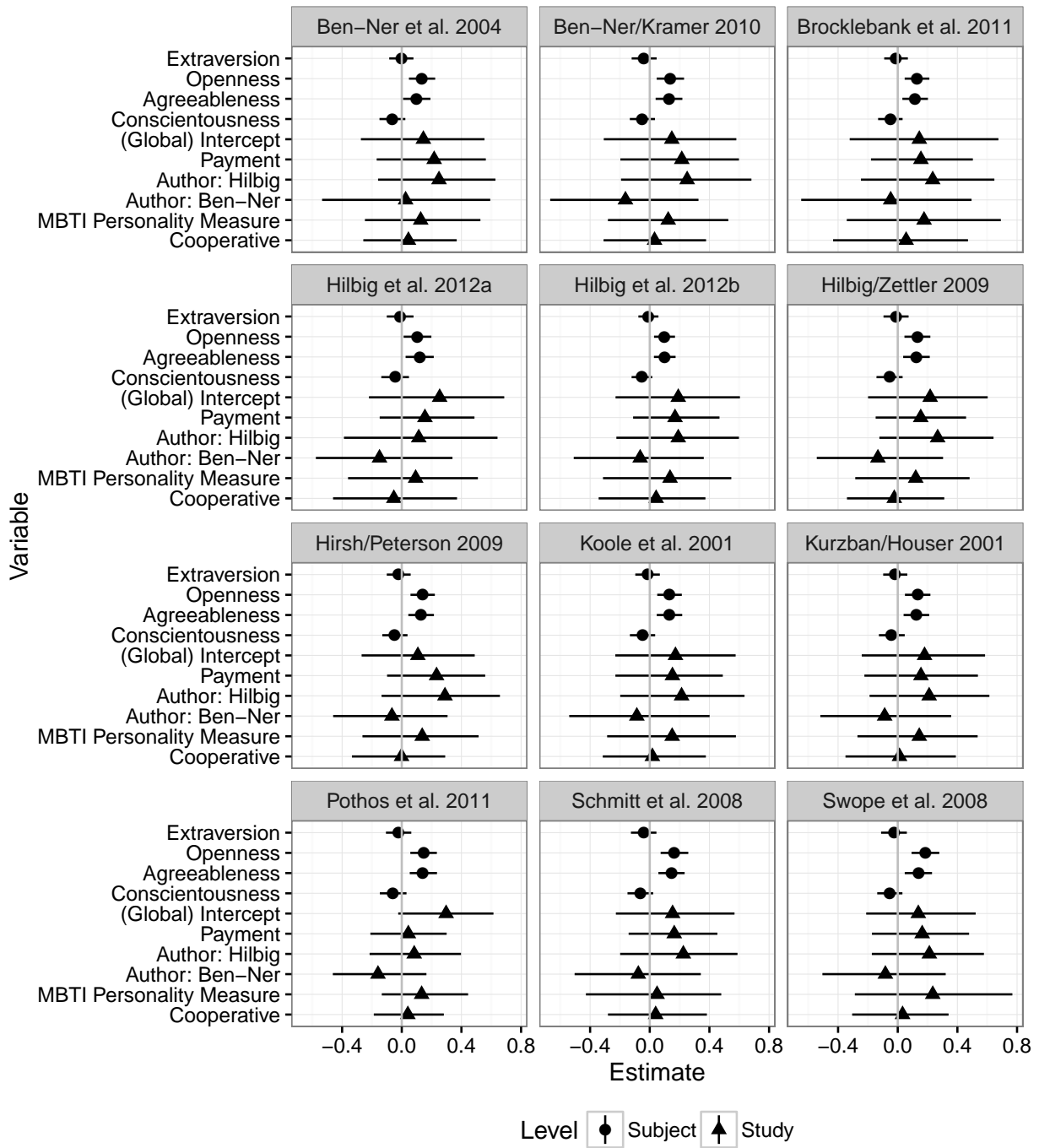


Figure A.3: **Multilevel Meta-Analysis Results: Predicting Prosocial Behavior (study jackknife)**. Each model excludes one of the studies from the estimation (for Model 1 in the main text), as indicated on the subfigure titles. For each model and each parameter, the dot represents the posterior mean and the bar represents the 95% credible interval.

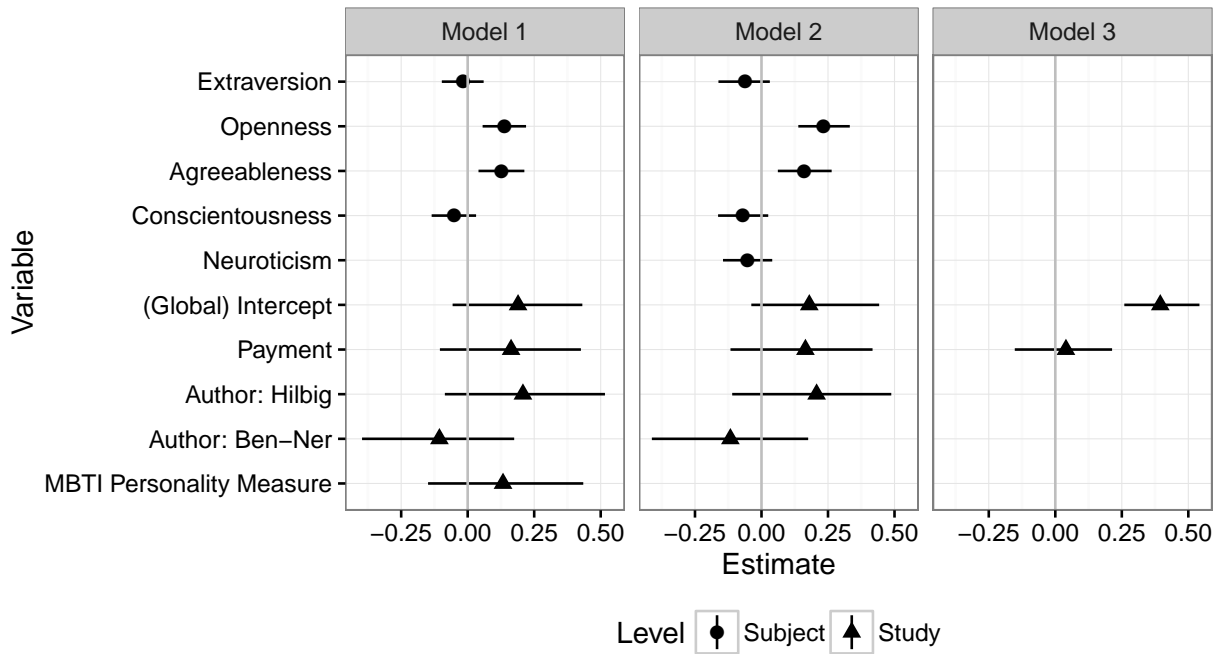


Figure A.4: **Multilevel Meta-Analysis Results: Predicting Prosocial Behavior (excluding cooperative dummy)**. Specification identical to main text, but excluding “cooperative” game type dummy. For each model and each parameter, the dot represents the posterior mean and the bar represents the 95% credible interval.



## References

- Florian Artinger, Filippos Exadaktylos, Hannes Koppel, and Lauri Sääksvuori. In others' shoes: do individual differences in empathy and theory of mind shape social preferences? *PloS one*, 9(4):e92844, 2014.
- Avner Ben-Ner and Amit Kramer. Personality and altruism in the dictator game: Relationship to giving to kin, collaborators, competitors, and neutrals. *Personality and Individual Differences*, 51(3):216–221, 2011.
- Avner Ben-Ner, Louis Putterman, Fanmin Kong, and Dan Magan. Reciprocity in a two-part dictator game. *Journal of Economic Behavior & Organization*, 53(3):333–352, 2004.
- Sean Brocklebank, Gary J Lewis, and Timothy C Bates. Personality accounts for stable preferences and expectations across a range of simple games. *Personality and Individual Differences*, 51(8):881–886, 2011.
- Gary Charness and Matthew Rabin. Understanding social preferences with simple tests. *The Quarterly Journal of Economics*, 117(3):817–869, 2002.
- Urs Fischbacher, Simon Gächter, and Ernst Fehr. Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters*, 71(3):397–404, 2001.
- Anna Gunnthorsdottir, Kevin McCabe, and Vernon Smith. Using the machiavellianism instrument to predict trustworthiness in a bargaining game. *Journal of Economic Psychology*, 23(1):49–66, 2002.
- Benjamin E Hilbig and Ingo Zettler. Pillars of cooperation: Honesty–humility, social value orientations, and economic behavior. *Journal of Research in Personality*, 43(3):516–519, 2009.
- Benjamin E Hilbig, Ingo Zettler, and Timo Heydasch. Personality, punishment and public goods: Strategic shifts towards cooperation as a matter of dispositional honesty–humility. *European Journal of Personality*, 26(3):245–254, 2012a.
- Benjamin E Hilbig, Ingo Zettler, Morten Moshagen, and Timo Heydasch. Tracing the path from personality via cooperativeness to conservation. *European Journal of Personality*, 2012b.
- Jacob B Hirsh and Jordan B Peterson. Extraversion, neuroticism, and the prisoners dilemma. *Personality and Individual Differences*, 46(2):254–256, 2009.
- Sander L Koole, Wander Jager, Agnes E van den Berg, Charles AJ Vlek, and Willem KB Hofstee. On the social nature of personality: Effects of extraversion, agreeableness, and feedback about collective resource use on cooperation in a resource dilemma. *Personality and Social Psychology Bulletin*, 27(3):289–301, 2001.
- Robert Kurzban and Daniel Houser. Individual differences in cooperation in a circular public goods game. *European Journal of Personality*, 15(S1):S37–S52, 2001.

- Emmanuel M Pothos, Gavin Perry, Philip J Corr, Mervin R Matthew, and Jerome R Busemeyer. Understanding cooperation in the prisoners dilemma game. *Personality and Individual Differences*, 51(3):210–215, 2011.
- Pamela Schmitt, Robert Shupp, Kurtis J Swope, and Justin Mayer. Personality preferences and pre-commitment: Behavioral explanations in ultimatum games. *Retrieved September, 13:2007*, 2004.
- Daniel Stegmueller. How many countries for multilevel modeling? a comparison of frequentist and bayesian approaches. *American Journal of Political Science*, 57(3):748–761, 2013.
- Kurtis J Swope, John Cadigan, Pamela M Schmitt, and Robert Shupp. Personality preferences in laboratory economics experiments. *The Journal of Socio-Economics*, 37(3):998–1009, 2008.
- Simon G Thompson, Rebecca M Turner, and David E Warn. Multilevel models for meta-analysis, and their application to absolute risk differences. *Statistical methods in medical research*, 10(6):375–392, 2001.