# Personality and Prosocial Behavior: A Multilevel Meta-Analysis*

Reuben Kline

Stony Brook University
Dept. of Political Science &
Center for Behavioral Political Economy

Alexa Bankert

University of Georgia
Dept. of Political Science

Lindsey Levitan

Shepherd University
Dept. of Psychology

Patrick W. Kraft

Stony Brook University
Dept. of Political Science

*We would like to thank the study authors who generously shared their data with us. We also thank two anonymous reviewers and the editors for their helpful suggestions which greatly improved the paper.

**Abstract**

We investigate the effect of personality on prosocial behavior in a Bayesian multilevel meta-analysis (MLMA) of 15 published, interdisciplinary experimental studies. With data from the 15 studies constituting nearly 2,500 individual observations, we find that the Big Five traits of Agreeableness and Openness are significantly and positively associated with prosocial behavior, while none of the other three traits are. These results are robust to a number of different model specifications and operationalizations of prosociality, and they greatly clarify the contradictory findings in the literature on the relationship between personality and prosocial behavior. Though previous research has indicated that incentivized experiments result in reduced prosocial behavior, we find no evidence that monetary incentivization of participants affects prosocial tendencies. By leveraging individual observations from multiple studies and explicitly modeling the multilevel structure of the data, MLMA permits the simultaneous estimation of study- and individual-level effects. The Bayesian approach allows us to estimate study-level effects in an unbiased and efficient manner, even with a relatively small number of studies. We conclude by discussing the limitations of our study and the advantages and disadvantages of the MLMA method.

# Introduction

In this study we investigate the relationship between personality and prosocial behavior. The relevant literature on this topic—spanning the disciplines of Economics, Psychology and Political Science—is a good example of what we call "pan-experimentalism," a trend within Political Science that consists of an attempt to recognize, reconcile, and utilize the diverse methodological traditions that inform experimental work in the discipline and understand these traditions in a unified, integrated way (Druckman et al., 2006; Morton and Williams, 2010). Our understanding of social and political processes is broadened and enriched by a more diverse set of perspectives. On the other hand, synthesizing a broad and diverse set of findings that employ distinct methodologies is a challenge.

To make sense of the interdisciplinary and often contradictory evidence on this subject, we rely on a technique—*multilevel* meta-analysis (MLMA)—which originated in the epidemiological (Turner et al., 2000) and educational (Goldstein et al., 2000) literatures. While traditional meta-analyses simply rely on the aggregation of study-level effects, MLMA combines individual-level data from multiple sources and directly takes into account their hierarchical structure. As such, this flexible technique allows for more efficient estimates of individual-level relationships of interest by controlling for study-specific variables and study-level error variance. MLMA is simply multilevel regression applied to meta-analysis such that the groups that define the higher-level units are the studies themselves (rather than for example, countries, states, or classrooms, as is more typically the case). The relationship between personality and prosociality is particularly suited to MLMA analysis because it is a politically important area of active research in both Psychology and Economics. Moreover, the relevant literature is characterized by findings that are inconsistent from one study to the next, and it is in precisely such a situation that meta-analysis can be particularly useful.

MLMA was developed in the epidemiology and education fields, but we believe it has been underutilized in Political Science. The major advantages of MLMA are that 1) it can clarify inconsistent or contradictory results; 2) it can directly control for individual level variables and model (cross-level) interactions, even those that are not included in the original results; 3) it can control for, and estimate the effects of, different research design choices across studies that address the same empirical question; and, 4) it can determine whether the inconsistencies within a given literature are the result of any differences in methods or other study level attributes. In short, MLMA can be fruitfully applied to many literatures within Political Science.

## Prosociality, Personality, and Politics

The concept of prosociality is of fundamental importance to Political Science, as it underlies motivations for political participation, social cooperation, charitable giving, voluntarism and redistributive preferences. Many studies focus on the question of prosociality and political behavior, without reference to personality traits. Fowler (2006) demonstrates that altruism predicts voter turnout. Further investigations (Fowler and Kam, 2007; Sautter et al., 2007; Bekkers, 2005) differentiate between general altruism and in-group favoritism in political participation. Similarly, Loewen (2010) examines the roles of affinity and antipathy in the turnout decision while Edlin et al. (2007) incorporate altruistic preferences into utility functions to predict significant voter turnout.

"Prosociality" has many definitions (see Beilin and Eisenberg 2013 for at least half a dozen) and often includes both altruistic and cooperative components. We will not wade into the debate about how exactly to define prosocial behavior, but a common thread that runs through most all of the definitions is that "prosociality" is the opposite of "selfishness." Therefore, we define prosociality for the purposes of this study as exactly that: we define behavior as prosocial to the extent that it goes against ones selfish interests *and* as a result potentially increases the payoffs to another person. In the context of the standard game forms that we consider here, this means either cooperation in social dilemmas such as the prisoner's dilemma and the public goods game, or generosity (i.e. altruism) in distributional games such as the dictator game or the trustee's role in the trust game (all of the games we consider will be described in more detail below). In each of these cases, we will consider behavior to be prosocial to the extent that it deviates from the selfish choice prescribed by the Nash equilibrium strategy–in other words, prosociality is characterized as cooperation in the cooperative dilemmas and as generosity in the distributive games.

There is some evidence that prosocial tendencies are related to a person's underlying personality traits. Personality should predict prosocial tendencies and prosocial political behaviors. Alford and Hibbing (2007) and Denny and Doyle (2008) find only a weak relationship between personality traits and general prosocial behavior, though they conclude that personality traits are much better predictors of prosocial *political* behavior. We will address the literature on how personality relates to prosocial tendencies, as measured by economic games, after we introduce our hypotheses.

Political scientists have long recognized the importance of personality in politics. Since Lasswell (1930) examined the role of personality in politics by applying psychoanalytic theory, personality research has been a persistent domain of Political Science research. Researchers have examined the role of the authoritarian personality type (Adorno et al., 1950; Altemeyer, 1988), the role of Big Five personality traits, and more recently the role of personality in mediating genetic influences in understanding politics (Alford et al., 2005; Fowler and Dawes, 2008). The Big Five personality traits in particular have helped improve our understanding of partisanship, ideology, political sophistication, and political participation (see Gerber et al. (2010), for a review, Mondak (2010) for detailed discussion).

The most common measures of Big Five personality traits are variations on Costa and McCrae's NEO measure (Costa and McCrae, 1992). There are, however at least two other scales that measure traits that are very similar to the Big Five, and in our MLMA we include studies that use these two measures as well. These scales are typically employed by asking respondents to agree or disagree with a number of self-referential statements.

The Big Five traits were developed from a 5-factor analysis of these responses. These factors have subsequently been labelled Agreeableness, Conscientiousness, Extraversion, Neuroticism, and Openness to Experience. We include two scales which are similar to this canonical Big Five. The HEXACO measure (Lee and Ashton, 2004), is substantively very similar, but arises from a 6-factor solution, rather than a 5-factor one. The significant degree of similarity between these scales is underscored by the high correlation[1] between the Big Five factors with their relevant HEXACO factors (Lee and Ashton, 2004).

We also examine the Myers-Briggs Type Indicator (Myers et al., 1985), whose types of Extraversion/ Introversion and Sensing/Intuition correlate very strongly with the Big Five traits of Extraversion and Openness to Experience, and whose types of Thinking/Feeling and Judging/Perceiving show moderately strong correlations with Agreeableness and Conscientiousness respectively (McCrae and Costa, 1989). Given the evidence on the correspondence between the HEXACO scale, the Big Five traits and the Myers-Briggs Type Indicator, we feel comfortable including them all together in a single analysis, at least initially. The ability to do this is a function of the flexibility of MLMA in that it allows us to combine studies using each of the three scales, while also controlling for any effects that may arise from the measures themselves. Combining disparately measured concepts in this way would be much more problematic in a conventional meta-analysis because it would be difficult to control for these additional factors in an efficient way, especially with a small number of studies.

## Multilevel Meta-analysis

Though common in psychology and epidemiology, meta-analyses are quite rare in Political Science (Morton and Williams, 2010). Those that have been conducted and published in leading journals (e.g., Doucouliagos and Ulubaşoğlu, 2008; Lau et al., 1999) employ

[1] Lee and Ashton (2004) demonstrate correlations ranging between .68 (Openness) and .86 (Extraversion), with the other 3 factors falling between. Honesty, the 6th HEXACO factor, is not examined here due to insufficient coverage across our sample of studies.

the conventional approach which relies on statistical techniques to standardize and then synthesize published effect sizes.

MLMA is a more flexible approach to meta-analysis that employs multilevel regression techniques to individual-level data from multiple studies. The reliance on individual observations rather than published effects as a unit of analysis allows us to synthesize results across any number of studies using any of the following types of data:

1. individual characteristics (e.g. attitudes or personality traits);
2. study-level variables (e.g. common investigator, common methodology, etc.)[2];
3. aggregated data (i.e. study-level summary statistics such as treatment effects and their variability); and
4. a combination of (1) and (2).

MLMA, is a powerful technique that is capable of, among other things, examining differences between studies, conducting subgroup analyses and uncovering heterogeneous treatment effects (Thompson and Higgins, 2002). What ties together the items on our list above is their reliance on modeling the data's underlying multilevel structure. The multilevel structure allows us to treat the studies themselves as distinct sources of random error. Given the wide range of MLMA's potential uses, we cannot apply them all in a single study. In our application to personality and prosociality, we focus on (1), (2) and (4).

Multilevel modeling conceptualizes individual data points—called "level 1" units—as being nested within larger groups—called "level 2" units. Often these level 2 groups are countries, states or some other well defined grouping, in MLMA the level 2 units are the studies themselves. The level 1 units are the individual participants or subjects in a study. Depending on the application, the level 2 units themselves may be nested in larger

[2]This technique is sometimes called "meta-regression." For a more thorough discussion of meta-regression, see Thompson et al. (2001).

groups—level 3 units—and so on. While we only examine two-level multilevel structures in the present study, MLMA could, *mutatis mutandis*, be conducted with 3 or more levels.

Here we focus on the promise that MLMA holds in terms of its capacity to synthesize individual level data from many studies, while simultaneously accounting for study-level variance. Multilevel models can account for such variation in two distinct but complementary ways: through "random intercept" parameters, which model the intercept of each study as a random draw from a given distribution, and through "random slopes" parameters, which model one or more slope parameters in a similar way. With MLMA it is also possible to model interactions between individual and study level variables ("cross-level" interactions). In our empirical application we only consider random intercepts and random slopes models, but cross-level interactions can be incorporated into MLMA in a manner identical to that in multilevel models more generally. For a discussion of such issues see Gelman and Hill (2007) and Hox (2010).

To employ a MLMA that can model individual data, study level variables and a combination of both (cases 1, 2 and 4 above), we first specify a basic multi-level model with varying intercepts:

$$y_i \sim \mathrm{N}\left(\alpha_{j[i]} + \boldsymbol{\beta}\boldsymbol{x}_i, \sigma_y^2\right), \qquad \text{for } i \in \{1, \cdots, N\}, \tag{1}$$

$$\alpha_j \sim \mathrm{N}\left(\gamma + \boldsymbol{\delta}\boldsymbol{z}_j, \sigma_\alpha^2\right), \qquad \text{for } j \in \{1, \cdots, J\}, \tag{2}$$

where $y$ denotes the outcome of interest for observation $i$ (nested in study $j$), $\boldsymbol{x}_i$ is a vector of individual-level variables, and $\boldsymbol{z}_j$ is a vector of study-level covariates. The remainder are parameters to be estimated in the model. In the context of MLMA, the parameter vector $\boldsymbol{\beta}$ in equation (1) denotes the effects of individual-level predictors on the outcome of interest for individual observations in each study. The parameter $\alpha_j$ denotes the varying intercepts which are modeled as a function of study-level characteristics in equation (2).

The model specification described above assumes that the effects of individual-level characteristics, such as specific treatments, are constant across studies. We can relax this assumption by allowing the coefficients in $\beta$ to vary by study as well:

$$y_i \sim \mathrm{N}\left(\alpha_{j[i]} + \boldsymbol{\beta}_{j[i]}\boldsymbol{x}_i, \sigma_y^2\right), \qquad \text{for } i \in \{1, \cdots, N\}, \tag{3}$$

$$\alpha_j \sim \mathrm{N}\left(\gamma + \boldsymbol{\delta}\boldsymbol{z}_j, \sigma_\alpha^2\right), \qquad \text{for } j \in \{1, \cdots, J\}, \tag{4}$$

$$\boldsymbol{\beta}_j \sim \mathrm{N}\left(\boldsymbol{\mu}, \Sigma_\beta\right), \qquad \text{for } j \in \{1, \cdots, J\}, \tag{5}$$

Using such a specification, we allow treatment effects to vary by study but follow a common distribution with mean vector $\boldsymbol{\mu}$. Note that in equation (5), $\Sigma_\beta$ is diagonal covariance matrix with a unique variance parameter for each element in $\boldsymbol{\beta}$.

We recommend the MLMA approach to scholars who plan meta-analyses that involve looking at published summary statistics, such as effect sizes. MLMA using summary data, like all multilevel approaches, is more efficient than its fixed- or random-effects counterparts. Here, we focus on the promise that MLMA holds for leveraging individual-level data nested within studies. An important advantage of MLMA is that it can leverage the statistical power of all the individual data points in the constituent studies and therefore make efficient estimates of level 1 and level 2 effects with only a small number of studies.

Still, despite its efficiency, too few studies may limit the inferences MLMA can yield with respect to study-level factors. Nonetheless, there is no consensus on the minimum number of level 2 units required to properly estimate their coefficients in a multilevel framework. Rules of thumb vary from 8 at the low end to more than 100 at the high end (Stegmueller, 2013). With standard frequentist techniques, having a small number of studies in MLMA could result in biased estimates of coefficients and incorrect standard errors at the study-level. To avoid this problem, we employ Bayesian techniques that have been shown in Stegmueller (2013) to yield largely unbiased point estimates and more reasonable credible intervals (compared to frequentist confidence intervals) even when the number of level 2 units is as low as 5. For most applications of MLMA, we would recommend that researchers use the Bayesian approach, and we would insist on it if one's goal is to estimate the effects of study-level factors. MLMA is particularly useful in literatures in which diverse research designs are used, and in cases where there are thought to be important individual level variables that are subject to significant individual variation.

## Hypotheses

We hold clear hypotheses for two of the Big Five traits. The trait of *Openness to Experience* (sometimes referred to as "intellect") is aptly named, and reflects a general orientation towards learning and experiencing new things. This trait has also been found to relate to self-reported social Openness and willing to trust others, and is negatively related to prejudice (Flynn, 2005). We would therefore expect that individuals higher in Openness to Experience will exhibit more prosocial behavior. The results of previous literature in this area in no case show a significant negative relation between Openness and prosocial tendencies, but there are many more null findings than significantly positive ones. It is therefore appropriate to question whether these Openness findings are part of a pattern, or merely statistical anomalies. We anticipate, based on the literature on Openness, that a meta-analysis will detect a significant positive relation between Openness and prosocial behavior in experimental games.

The trait of *Agreeableness* carries with it a sense of likeability, but also care for others, and sometimes a tendency toward compliance. In short, people high in this trait are "nice" (McCrae and Costa, 1995). This, too, would sensibly generate more prosocial goals and behavior. Again, prior findings are not so clear cut. While no study we have examined detects a significantly negative relation between Agreeableness and prosocial game behavior, only four detect a significant positive one. Again, based on the conceptual underpinnings of the trait, we anticipate a significant positive relation between Agreeableness and prosocial tendencies in experimental games.

For the remaining traits, we have no clear theoretically-driven hypotheses. *Extraversion* is perhaps the most well-known Big Five trait. Some characterize it primarily as an orientation toward enjoying social interactions (Ashton and Lee, 2007; Denissen and Penke, 2008) whereas others emphasize assertiveness and activity-seeking, in addition to social warmth (McCrae and Costa, 1995). It is not clear that Extraversion has direct implications for prosocial behavior in games, in that the sociable aspect of Extraversion might lead one to expect greater prosocial behavior, but the tendency among extroverts to assert their own views and interests might suggest the reverse pattern. Indeed, the papers in our sample have contradictory findings with respect to Extraversion. Those who are *Conscientious* demonstrate a tendency to strive and plan. They are diligent and dependable (Ashton and Lee, 2007; Denissen and Penke, 2008), as well as competent, orderly, and self-disciplined (McCrae and Costa, 1995). It is unclear that this has any implications for prosociality, in that we would expect those who are conscientious to diligently and competently follow a strategy, but the end goal of that strategy (the common good or personal gain) would be best determined by other traits and factors. Indeed, only one paper in our sample shows a significant impact of Conscientiousness upon prosocial tendencies. Those who are high in the final Big Five trait, *Neuroticism* (sometimes called "emotionality"; McCrae and Costa (1995)), are anxious and sensitive, particularly when confronted with threats (Ashton and Lee, 2007; Denissen and Penke, 2008), but it is unclear how Neuroticism might relate to prosocial behavior or experimental games. Again, we find mixed results in the studies we identify examining Neuroticism and prosocial tendencies.

For the study-level variables, we lack sufficient prior evidence to form any theoretically-driven hypotheses. It is more intuitive to think that monetary payments might induce less prosocial behavior when compared to non-incentivized studies in which prosocial behavior does not imply monetary costs. Indeed in the only study we are aware of that directly tests the effects of incentives on prosociality, Lönnqvist et al. (2011) find evidence that participants are less generous in incentivized prisoner's dilemma games rather than hypothetical ones. Hence, we take an exploratory approach to the expected role of payments on prosocial behavior while we simply use the other study-level variables to better control for study-level variance.

## MLMA: Data and Analysis

Evidence on the relationship between personality and prosocial behavior is decidedly mixed. Even when limiting our focus to the Big Five type measures, there is a tremendous amount of heterogeneity in results across studies. Table 1 lists each of the studies we use in our MLMA along with the significant findings (as reported in the study's manuscript) and their direction for each study. A number of studies do show that Openness and Agreeableness are positively associated with prosocial behavior, though others show no such relationship. No two studies, however, have the same set of findings with respect to all factors: each factor is significant in at least one study, but never in all of them, or even a majority of them. In some studies most of the Big Five traits predict behavior (e.g., Brocklebank et al., 2011; Ben-Ner and Kramer, 2011), whereas in others, no trait does (e.g., Kurzban and Houser, 2001). In some case, depending on the study, a given trait can be both positively and negatively correlated with prosociality. For example, Koole et al. (2001) and Brocklebank et al. (2011) find that Extraversion and Neuroticism are *negatively* and significantly associated with prosociality. Whereas in other studies, Extraversion (Swope et al., 2008) and Neuroticism (Hirsh and Peterson, 2009) are found to have a *positive* and significant association. Sometimes, even negative and positive correlations appear in different studies in the same paper (e.g., Extraversion in Ben-Ner and Kramer, 2011). This variety of mixed and contradictory findings regarding how personality traits predict prosocial tendencies is precisely the situation in which MLMA is most useful. Meta-analysis can clarify these mixed findings and tell us whether sporadic significant findings are likely to be a result of Type I or Type II errors, or part of a systematic pattern of results. Below, table 1 summarizes the findings from the studies in our sample.

## Data

Here, we bring together individual data from a diverse set of studies from Economics and Psychology. We conduct a MLMA to assess the relationship between personality traits and prosocial behavior in experimental games. In order to identify relevant papers for our examination of the relation between personality traits and behavioral game responses, we searched the literature extensively. Several databases were searched using all pairwise combinations between keywords in the following two sets of keywords: (1) "personality traits", "Big Five", "HEXACO", "Social Value Orientations" and (2) "economic games", "cooperation", "public good experiments, "Dictator Games," "Ultimatum Game" and "Prisoner's Dilemma." Most articles were found through the EBSCOhost search engine. In a more targeted fashion, we made use of the listserv of the Economic Science Association to solicit studies (both published and unpublished) directly from scholars.

### Criteria for Inclusion in the Study

We conducted the first round of study selection based on the following broad criteria:

- The study uses a standard personality measure (e.g., Big Five, HEXACO or MBTI).
- The study includes a canonical experimental game form to gauge subjects' behavioral responses and their relationship to personality traits.

Our search process yielded 53 papers that met our selection criteria, connecting some personality measure to outcomes in experimental games. Due to our focus on Big Five personality traits and the limited number of studies employing less well known personality traits, we excluded 26 papers with other types of personality measures. We also excluded seven papers because they contained non-canonical versions of the classic games, which would have compromised comparability across studies. We contacted the authors of these 20 remaining studies, requesting that they share their data with us. This process led to a total of 12 primary data sets that met our criteria. Our search also yielded three additional studies which contain the requisite measure of prosocial behavior but used a type of personality measurement other than a Big Five type personality trait. As a result these could not be included in analyses in which we estimate the individual effects of personality traits. Nonetheless, the flexibility of MLMA allows us to use these studies in a separate model in which we exclude individual covariates and instead focus on study-level factors such as incentivization, pushing the total number of studies up to 15.

Table 1 summarizes the 12 studies which are included in our analysis of personality variables, and a total of 15 studies in the analysis of incentivization differences. This multidisciplinary set of studies displays considerable diversity in terms of research design and experimental game type used.

## Coding of the Variables

Because the studies in our sample use different game forms and parameterizations, we must make our dependent variable—prosociality—commensurable across studies. To do so, we determine the upper and lower bounds of prosocial behavior which were available to participants, and then re-scaled the values from 0 to 1. As discussed above, we operationalize prosocial behavior as the extent to which behavior does not coincide with the equilibrium prediction of defection in cooperative games or retaining all rewards for oneself in distributional games. For example, in a dictator game with an endowment of \$10, if a participant allocates \$3 to their counterpart, then our measure of prosociality would be 0.3. For studies which involved a prisoner's dilemma, "cooperate" was coded as 1 and "defect" was coded as 0. In total, our studies include five distinct games: prisoner's dilemma (PD), public goods game (PGG), common pool resource dilemma (CPRD), ultimatum game (UG) and dictator game (DG) and trust game (TG). These are summarized

Table 1: Studies Included in Meta-Analysis

| Study | Game | Personality Measure | Significant Predictors | Incentivized | Sample Size |
|---|---|---|---|---|---|
| Ben-Ner and Kramer (2011) | DG | Big Five | $(+)N; (-/+)E; (-)C$ | no | 198 |
| Ben-Ner et al. (2004) | DG | Big Five | $(+)A; (+)N$ | yes | 320 |
| Brocklebank et al. (2011) | DG | Big Five | $(-)N; (-)E; (+)O$ | yes | 67 |
| Hilbig and Zettler (2009) | DG | HEXACO | No test | no | 134 |
| Hilbig et al. (2012a) | DG | HEXACO | $(+)O$ | no | 424 |
| Hilbig et al. (2012b) | PGG | HEXACO | $(+)O; (+)A$ | no | 531 |
| Hirsh and Peterson (2009) | PD | Big Five | $(+)N$ | no | 52 |
| Koole et al. (2001) | CPRD | Big Five | $(-)E; (+)A$ | yes | 71 |
| Kurzban and Houser (2001) | PGG | Big Five | None | yes | 91 |
| Pothos et al. (2011) | PD | Big Five | $(+)A$ | no | 113 |
| Schmitt et al. (2004) | UG | MBTI | None | yes | 120 |
| Swope et al. (2008) | DG, TG, PD | MBTI | $(+)E$ | yes | 134 |
| Artinger et al. (2014) | DG | N/A | N/A | yes | 116 |
| Fischbacher et al. (2001) | PGG | N/A | N/A | yes | 44 |
| Gunnthorsdottir et al. (2002) | TG | N/A | N/A | yes | 67 |

Notes: CPRD = common pool resource dilemma; DG = dictator game; PD = prisoner's dilemma; PGG = public goods game; TG = trust game; and UG = ultimatum game.
N = "Neuroticism"; E = "Extraversion"; O = "Openness"; A = "Agreeableness"; C = "Conscientiousness";
(−) = statistically significant, negative association; (+) = statistically significant, positive association

Table 2: Game Types and Normalization of Outcome Measures

| Game Type | Game | Minimum (0) | Maximum (1) |
|---|---|---|---|
| Cooperative | Prisoner's Dilemma | "defect" | "cooperate" |
| | Public Goods Game | zero contribution | full contribution |
| | Common Pool Resource | full extraction | zero extraction |
| Distributive | Ultimatum Game (proposer) | retain all | give all |
| | Dictator Game | retain all | give all |
| | Trust Game (trustee) | retain all | give all |

in Table 2, along with the normalization rule for each type. In both UG and TG, there are two distinct roles. In order to simplify our analysis of these games, we analyze only the "distributive" roles in these two games: in UG, the role of the proposer—who proposes a split of a fixed sum between herself and the responder— and in TG the role of the trustee—who receives a transfer from the trustor, and then must decide how to split this fixed sum between them. We define "distributional" games as those involving the allocation of a pot of money that is fixed at the time of the decision—e.g. DG, UG (proposer) and TG (trustee)—and "cooperative" as those that involve a choice between cooperation and defection (PD, CPRD), or contribution and non-contribution (PGG). Strategically, these cooperative games are equivalent, with defection/non-contribution/full-extraction the unique equilibrium in the one-shot context. In these cases, we code cooperation and full contribution as 1 and defection and zero contribution as 0. There are of course alternative methods of scaling these outcome measures, but we feel that this approach is the most straightforward and generally better preserves the original study's scaling. Further details about the specific normalization procedure that was used for each study can be found in the appendix table A.1.

Proscociality, like personality, is now thought to be a stable and domain general individual trait (Peysakhovich et al., 2014), justifying the pooling of these different games in our analysis. In addition to Peysakhovich et al. (2014), who find behavioral stability across four games (DG, PGG, and both roles in TG); Yamagishi et al. (2013) also find that prosocial behavioral tendencies are consistent across PD, TG, and DG. To the extent that there is additional variance as a function of the different types of games, it will be incorporated in the multilevel structure of the MLMA estimations. We also include a dummy variable indicating whether the game type is cooperative or distributional in nature.

Finally, a note on our Payment variable. Psychologists rely primarily on participant pools composed of students who are given course credit in exchange for their participation in the experiment. On the other hand, the norms of experimental Economics require that subjects be compensated based on an incentivized payment scheme designed to induce preferences over experimental alternatives (Smith, 1976). The sample of studies included in this analysis spans both the psychological and Economics literature. A study is coded as offering incentivized payments if it offers monetary payments *and* the amount of these payments depends on the choices of the participants. Otherwise, studies are coded as unincentivized. All 15 studies in our sample were conducted with student populations.

## Multilevel Models and Results

Since we have access to very few common covariates in the datasets, the models we present here have to be parsimonious. As such, in the context of MLMA, there is a trade-off between complex model specifications and the total number of studies which can be included in the analysis. However, the multilevel nature of MLMA allows us to capture any remaining study-level effects as sources of variation in the random intercepts and random slopes, which reduces the chance that parsimony in variable selection results in underspecification.

As a first step, we specify three varying intercept models, as described in equation 2. Model 1 estimates the effects of personality on prosocial behavior across the 12 studies for which we have prosociality *and* personality data. Two of our studies (Schmitt et al., 2004; Swope et al., 2008) employ Myers-Briggs Type Indicator (MBTI). The four dimensions of MBTI have been shown to be well-correlated with four of the five Big Five factors, but, there is no MBTI factor that is analogous to Neuroticism so we exclude that factor in Model 1. Model 1 includes the individual-level variables of Extraversion, Openness, Agreeableness, and Conscientiousness, as well as the following study-level variables: a dummy for whether the study was incentivized (*Payment*), a dummy indicating the use of MBTI (*MBTI*), dummies for experimental game type—"cooperative"[3]—and dummy variables for common authorship of studies (*Hilbig* and *Ben-Ner*). Additionally, to the extent that HEXACO differs from the traditional measurement of the Big Five, our Hilbig author dummy would pick up this difference as well, as all three Hilbig et al. studies, and no others, employ the HEXACO measure. MLMA easily allows us to determine whether study-level variables that capture crucial conceptual distinctions have any significant effect on our outcomes.

If we include MBTI-based studies we cannot investigate the effect of Neuroticism. So, to estimate the effect of Neuroticism, and to test the robustness of our results, Model 2 includes only the 10 studies that employ the Big Five/HEXACO measurements of personality. In Model 2, we include the five personality dimensions as well as study-level variables relating to shared authorship and game type. Since the MBTI studies are dropped from this sample so is the variable in the specification of Model 2.

Finally, in Model 3, we include prosociality and incentivization data from an additional 3 studies that did not have personality data. This allows us to further examine the potential effect of incentivization in a larger set of studies.

Following the notation laid out above in equations 1 and 2, our multilevel specifications are as follows, for individuals $i \in \{1, \cdots, N\}$ within studies $j \in \{1, \cdots, J\}$ :

[3]We conduct all of the anlaysis with specifications that are identical to those below but excluding the cooperative dummy variable. The results, available in the figure A.4 in the appendix, are virtually identical to our main results reported here.

Level 1:[4] $y_i \sim \mathrm{N}(\alpha_{j[i]} + \beta_1 * Openness_i + \beta_2 * Agreeableness_i + \beta_3 * Extraversion_i$

$$+ \beta_4 * Conscientiousness_i + \beta_5 * Neuroticism_i, \sigma_y^2) \quad (6)$$

Level 2:[5] $\alpha_{j[i]} \sim \mathrm{N}(\gamma + \delta_1 * Payment_j + \delta_2 * Hilbig_j + \delta_3 * BenNer_j + \delta_4 * MBTI_j$

$$+ \delta_5 * Cooperative_j, \sigma_\alpha^2) \quad (7)$$

Throughout the models, we specify the following weakly informative prior distributions for our parameters of interest:

$$\alpha_j \sim N(0, 1) \quad (8)$$

$$\boldsymbol{\beta} \sim N(0, 1) \quad (9)$$

$$\gamma_j \sim N(0, 1) \quad (10)$$

$$\boldsymbol{\delta}_j \sim N(0, 1) \quad (11)$$

Note that the prior distributions are vague in relation to the scaling of the dependent and independent variables included in the model. The priors for the remaining variance parameters were specified as uniform. Alternatively, employing Cauchy(0,5) or non-informative uniform prior distributions for all parameters does not affect the substantive results (see for example Gelman et al., 2006, for a discussion of prior choice on variance parameters in multilevel models). All models were estimated via Hamiltonian Monte Carlo implemented in STAN (Stan, 2016). Individual chain lengths were set at 10,000 or 20,000 in order to assure convergence. Trace plots of the chains as well as $\hat{R}$ statistics for each model (not shown) indicate that the chains mixed well (c.f. Gelman et al., 2014).

Table 3 displays the results from our three models, consisting of the posterior mean

[4] Neuroticism is not included in Model 1. No personality variables are included in Model 3.
[5] MBTI is not included in Model 2.

Table 3: Prosociality, Personality and Incentivization: Multilevel Regression Results

| **Variable** | Model 1: Big Five & MBTI Studies | | Model 2: Big Five Studies | | Model 3: Incentivization | |
|---|---|---|---|---|---|---|
| **Subject-level variables** | | | | | | |
| | Mean | 95% Credible Interval | Mean | 95% Credible Interval | Mean | 95% Credible Interval |
| Extraversion | -0.02 | [-0.10, 0.06] | -0.06 | [-0.15, 0.03] | – | – |
| Openness | 0.14 | [0.05, 0.22] | 0.23 | [0.13, 0.33] | – | – |
| Agreeableness | 0.12 | [0.04, 0.21] | 0.16 | [0.06, 0.26] | – | – |
| Conscientiousness | -0.05 | [-0.14, 0.03] | -0.07 | [-0.17, 0.02] | – | – |
| Neuroticism | – | – | -0.05 | [-0.14, 0.03] | – | – |
| **Study-level variables** | | | | | | |
| | Mean | 95% Credible Interval | Mean | 95% Credible Interval | Mean | 95% Credible Interval |
| (Global) Intercept | 0.18 | [-0.18, 0.57] | 0.15 | [-0.22, 0.53] | 0.39 | [0.20, 0.56] |
| Payment | 0.17 | [-0.14, 0.48] | 0.17 | [-0.15, 0.46] | 0.05 | [-0.14, 0.24] |
| Author: Hilbig | 0.21 | [-0.18, 0.57] | 0.23 | [-0.16, 0.60] | – | – |
| Author: Ben-Ner | -0.10 | [-0.54, 0.32] | -0.09 | [-0.49, 0.34] | – | – |
| MBTI Personality Measure | 0.13 | [-0.27, 0.51] | – | – | – | – |
| Cooperative | 0.02 | [-0.30, 0.33] | 0.04 | [-0.27, 0.34] | 0.01 | [-0.20, 0.21] |
| # Observations | | 2,235 | | 1,981 | | 2,482 |
| # Studies | | 12 | | 10 | | 15 |

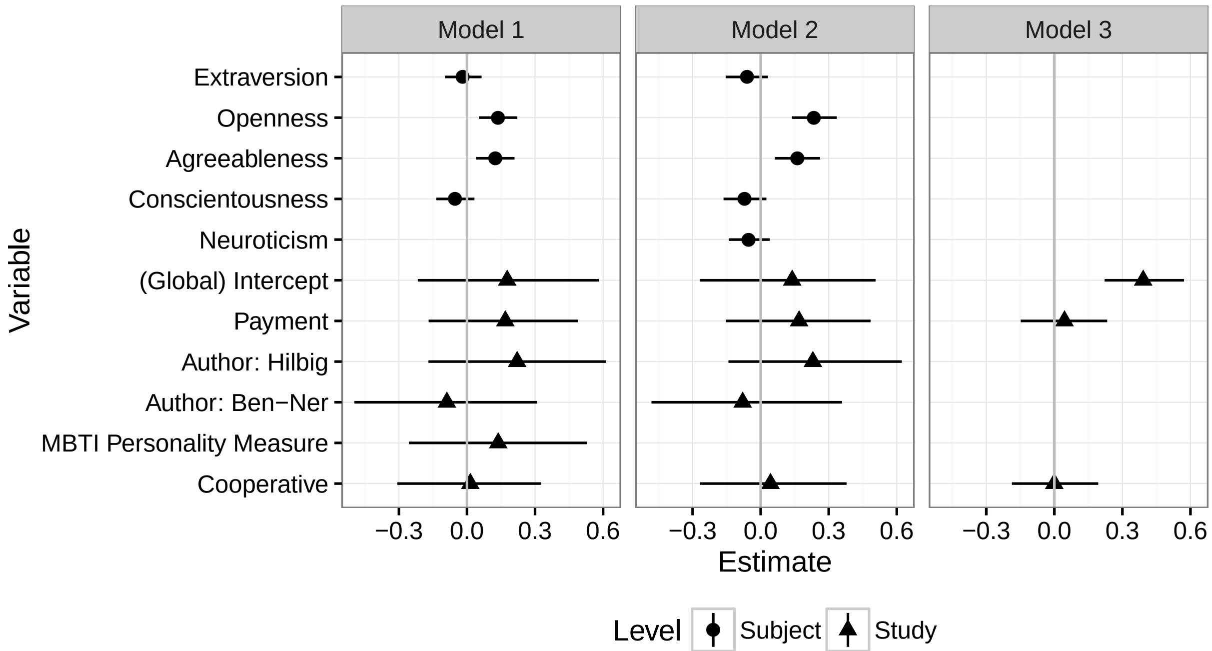


Figure 1: **Multilevel Meta-Analysis Results: Predicting Prosocial Behavior**. For each model and each parameter, the dot represents the posterior mean and the bar represents the 95% Credible Interval

and its 95% credible interval.[6] The credible interval describes the range of parameter values that encompasses 95% of the posterior distribution's probability mass. As such, there is a 95% probability that a given parameter lies in the respective credible interval. Figure 1 graphically displays this same information with each dot representing the mean prediction of the parameter value and the bar representing the 95% credible interval.

At the individual level, higher levels of Agreeableness and Openness increased prosociality in both Models 1 and 2. Though the magnitude of the effect is relatively small, the results for these two traits are quite robust. Irrespective of the specification of the model, these two factors are consistently associated with increased prosociality. For the remaining three traits, the effects are smaller and there is much larger uncertainty about their direction. Given the inconsistency of the findings when looking at each of the studies in

[6]For comparison's sake, the results from a frequentist MLMA estimation are included in the appendix table A.5.

our sample in isolation, it is remarkable that the results with respect to Agreeableness and Openness are so robust. These results demonstrate the power of MLMA in that it can extract aggregate relationships from a large number of studies that would not be apparent when looking at any of the studies on their own.

Moving on to the study-level variables, we do not find that any study-level factors systematically affect prosociality. This may simply be because of a lack of statistical power at the study-level. Additional analyses (not shown) controlling for each type of game (summarized in table 2) yield the same conclusion of no study-level effects, but such a model leaves us with very few degrees of freedom at the study-level. Thus, from our analysis we can conclude that incentivization, at least in the set of studies we have, does not play a significant role in prosocial behavior.

In order to ensure that the effects are not driven solely by a single study, we text whether the results are robust to individual studies being removed from the analysis. Using a jack-knife method, we find that the one-by-one exclusion of each of the 12 studies included in Model 1 does not change the estimates substantially. These results are included in the appendix, figure A.3. Both Agreeableness and Openness remain significant in each of the 12 estimations. As such, it does not appear to be the case that the results are driven by a single study included in the meta-analysis. A more systematic way to address this question, however, is to directly investigate the variation in individual-level effects across the studies. To do so, we specify a model which adds a random slopes component to the model outlined in equations (3) through (5). The model takes the following form:

$$\begin{aligned}
\text{Level 1:}\quad & y_i \sim \mathrm{N}(\alpha_{j[i]} + \beta_{j[i],1} * Openness_i + \beta_{j[i],2} * Agreeableness_i + \beta_{j[i],3} * Extraversion_i \\
& \quad + \beta_{j[i],4} * Conscientiousness_i, \sigma_y^2) && (12)\\
\text{Level 2:}\quad & \alpha_j \sim \mathrm{N}(\gamma + \delta_1 * Payment_j + \delta_2 * Hilbig_j + \delta_3 * BenNer_j + \delta_4 * MBTI_j \\
& \quad + \delta_5 * Cooperative_j, \sigma_\alpha^2) && (13)\\
& \boldsymbol{\beta}_j \sim \mathrm{N}(\boldsymbol{\mu}, \Sigma_\beta), && (14)
\end{aligned}$$

where prior specifications are equivalent to the previous models. The results are presented in Figure 2. It displays the individual-level effects for each of the 12 studies included in the analysis, the aggregate effects of individual-level variables across all studies, as well as study-level effects.

We can see that there is some substantial variation in individual-level effects between studies. While the posterior mean for the coefficients of Agreeableness and Openness is positive in most of the studies, the effects appear to be substantially smaller in some of the experiments than in the others—that is to say there is cross-study variation in the strength of the effect of these personality factors on prosociality. Still, the aggregated individual-level effects at the bottom of the figure (in the "Overall" panel) show the same pattern as the previous analysis. As such, even though there is some variation in effects between experiments, we recover a substantial positive influence of Agreeableness and Openness on prosocial behavior when combining the individual studies in a common framework.

# Discussion

Using MLMA—an underutilized tool in Political Science—we have investigated the effect of personality traits on prosociality. Despite the mixed or even contradictory results found in the literature, we find strong evidence that the Big Five factors of Agreeableness

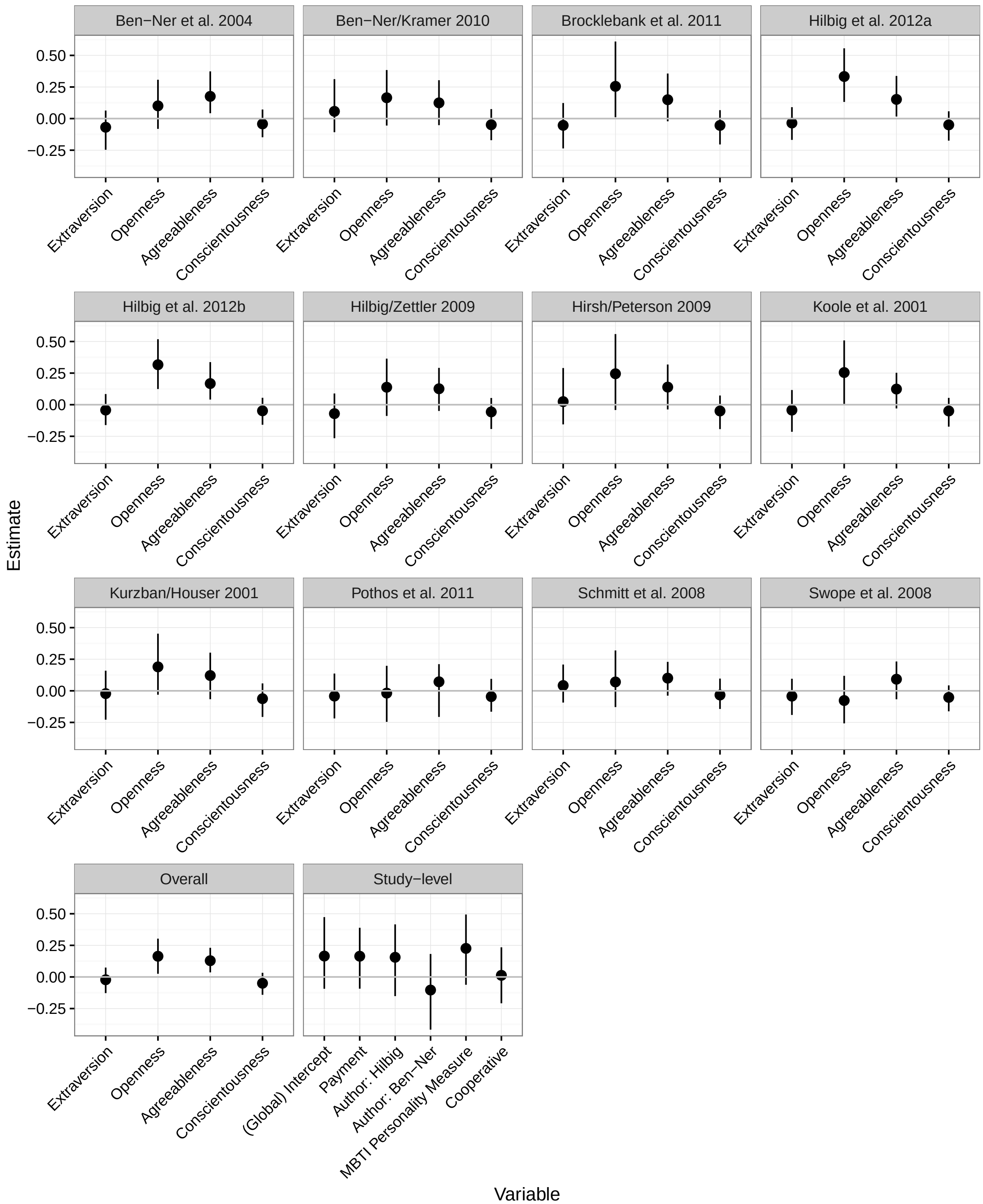


Figure 2: **Multilevel Meta-Analysis Results: Predicting Prosocial Behavior (varying slopes)**. For each model and each parameter, the dot represents the posterior mean and the bar represents the 95% Credible Interval

and Openness are consistent and significant predictors of increased prosocial behavior. Moreover, these results are robust across a number of distinct combinations of studies and model specifications. We do not, however, find evidence that, all else equal, monetary incentivization has any effect on prosocial behavior. These results are important because they clarify the inconsistent evidence found in the literature, and therefore provide a useful guide for further research on the effects of personality and incentivization on prosocial behavior. It is precisely in such a context that meta-analysis can be useful.

Given the multidisciplinary nature of experimental political science, MLMA represents a promising analytical method for the researcher's toolbox since it allows for a systematic and efficient analysis of study-level factors that may affect outcomes of interest. In addition to the issue of incentivization we explore here, these factors can also relate to the nature of the sample (i.e. representative versus convenience sample), the data collection process (i.e. online versus laboratory experiments), as well as differences in experimental stimuli (i.e. audio, visuals, text, etc.). Much like we did to model the differences between MBTI and the Big-5, the effects of other differences in scales or instruments can be estimated using MLMA. Generally, whenever experimental protocols are not completely identical across studies, multi-level meta analysis allows researchers to account for these variations and estimate their potential effects on the dependent variable. From an experimental perspective, this statistical approach does not just prevent confounding factors from clouding the effect of the treatment but it also opens up new research avenues that specifically examine the impact of these study-level variations. Moreover, with the Bayesian approach we both employ and recommend, a relatively small number of available studies does not preclude an unbiased estimate of study-level effects making MLMA even more appealing to experimentalists who might not have access to high number of studies. Moreover, because one can also model study level characteristics using only summary treatment effects from each study, a type of MLMA is also possible without access to individual level data.

The applicability of MLMA, however, is by no means limited to experimental studies. MLMA is a general and flexible framework that in addition to the applications we have demonstrated here subsumes conventional meta-analysis (footnote 8) and can also be utilized to combine analyses that rely on different observational data sources. In case of the latter, studies that, for example, vary in their data collection mode (i.e. phone, web, face-to-face), sample characteristics (e.g. geographic regions), interviewer characteristics (e.g. race), as well as timing of the data collection (e.g. pre- and post- election season) can feasibly be pooled together and analyzed in the MLMA framework.

However, meta-analysis is not without its limitations. A general concern with meta-analyses is the presence of publication bias. It is difficult to know for sure whether such bias (or other types of sample selection bias) are affecting our results, but the inclusion of some studies with opposing findings and several with null effects suggest that such bias is less a concern in our application than it may be in others. An additional type of selection bias is also possible with the type of MLMA we use here, because it requires researchers to share their datasets, including individual level data. In contrast traditional meta-analysis can proceed merely by extracting effect sizes from published studies and working papers. Therefore, MLMA results might be of limited generalizability due to potential selection biases that can occur if data is not missing at random. One potential way to address this concern is to compare the results from studies the researcher has obtained data for to the results from studies for which data was unavailable; if they differ systematically, there is reason to be concerned about the MLMA results. In this manuscript, the eight studies we identified which met our criteria, but for which we failed to acquire data, to a large extent mirrored the results from the twelve studies we included in our analysis of personality characteristics. Of the eight studies, one and seven respectively found a positive and significant effect of Openness and Agreeableness, whereas for the studies we included in our analysis, these numbers are three and four out of nine respectively. Just as in the data we analyzed, the studies that we did obtain are characterized by mixed results for the other three factors.

In addition to this relatively effortless check on the data, researchers can also actively contribute to remedying the problem of selection bias by supporting and participating in various initiatives that promote data sharing and transparency aims that every discipline should strive towards but which are—in the context of this manuscript—particularly worth mentioning as they directly increase MLMA's utility. Moreover, if the use of MLMA becomes widespread it may itself encourage increased data-sharing, creating a virtuous feedback loop.

# References


Theodor W Adorno, Else Frenkel-Brunswik, Daniel J Levinson, and R Nevitt Sanford. The authoritarian personality. 1950.

John R Alford and John R Hibbing. Personal, interpersonal, and political temperaments. *The Annals of the American Academy of Political and Social Science*, 614(1):196–212, 2007.

John R Alford, Carolyn L Funk, and John R Hibbing. Are political orientations genetically transmitted? *American political science review*, 99(02):153–167, 2005.

Bob Altemeyer. *Enemies of freedom: Understanding right-wing authoritarianism.* Jossey-Bass, 1988.

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.

Michael C Ashton and Kibeom Lee. Empirical, theoretical, and practical advantages of the hexaco model of personality structure. *Personality and Social Psychology Review*, 11 (2):150–166, 2007.

Harry Beilin and Nancy Eisenberg. *The development of prosocial behavior*. Academic Press, 2013.

René Bekkers. Participation in voluntary associations: Relations with resources, personality, and political values. *Political Psychology*, 26(3):439–454, 2005.

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.

Paul T Costa and Robert R McCrae. Normal personality assessment in clinical practice: the neo personality inventory. *Psychological assessment*, 4(1):5, 1992.

Jaap JA Denissen and Lars Penke. Motivational individual reaction norms underlying the five-factor model of personality: First steps towards a theory-based conceptual framework. *Journal of Research in Personality*, 42(5):1285–1302, 2008.

Kevin Denny and Orla Doyle. Political interest, cognitive ability and personality: Determinants of voter turnout in britain. *British Journal of Political Science*, 38(02):291–310, 2008.

Hristos Doucouliagos and Mehmet Ali Ulubaşoğlu. Democracy and economic growth: a meta-analysis. *American Journal of Political Science*, 52(1):61–83, 2008.

James N Druckman, Donald P Green, James H Kuklinski, and Arthur Lupia. The growth and development of experimental research in political science. *American Political Science Review*, 100(4):627, 2006.

Aaron S Edlin, Andrew Gelman, and Noah Kaplan. Voting as a rational choice: Why and how people vote to improve the well-being of others. *Rationality and society*, 1, 2007.

Urs Fischbacher, Simon Gächter, and Ernst Fehr. Are people conditionally cooperative? evidence from a public goods experiment. *Economics Letters*, 71(3):397–404, 2001.

Francis J Flynn. Having an open mind: the impact of openness to experience on interracial attitudes and impression formation. *Journal of Personality and Social Psychology*, 88(5): 816, 2005.

James H Fowler. Altruism and turnout. *Journal of Politics*, 68(3):674–683, 2006.

James H Fowler and Christopher T Dawes. Two genes predict voter turnout. *The Journal of Politics*, 70(03):579–594, 2008.

James H Fowler and Cindy D Kam. Beyond the self: Social identity, altruism, and political participation. *Journal of Politics*, 69(3):813–827, 2007.

Andrew Gelman and Jennifer Hill. *Data analysis using regression and hierarchical/multilevel models*. Cambridge University Press: Cambridge, UK, 2007.

Andrew Gelman, John B Carlin, Hal S Stern, David B. Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. New York: CRC Press, 3rd edition, 2014.

Andrew Gelman et al. Prior distributions for variance parameters in hierarchical models. *Bayesian analysis*, 1(3):515–534, 2006.

Alan S Gerber, Gregory A Huber, David Doherty, Conor M Dowling, and Shang E Ha. Personality and political attitudes: Relationships across issue domains and political contexts. *American Political Science Review*, 104(01):111–133, 2010.

Harvey Goldstein, Min Yang, Rumana Omar, Rebecca Turner, and Simon Thompson. Meta-analysis using multilevel models with an application to the study of class size effects. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 49(3):399–412, 2000.

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 personalityvia cooperativenessto 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.

Joop Hox. *Multilevel analysis: Techniques and applications*. Routledge, 2010.

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.

Harold D Lasswell. *Psychopathology and politics*. University of Chicago Press, 1930.

Richard R Lau, Lee Sigelman, Caroline Heldman, and Paul Babbitt. The effects of negative political advertisements: A meta-analytic assessment. *American Political Science Review*, pages 851–875, 1999.

Kibeom Lee and Michael C Ashton. Psychometric properties of the hexaco personality inventory. *Multivariate Behavioral Research*, 39(2):329–358, 2004.

Peter John Loewen. Affinity, antipathy, and political participation: How our concern for others makes us vote. *Canadian Journal of Political Science*, 43(3):661–687, 2010.

Jan-Erik Lönnqvist, Markku Verkasalo, and Gari Walkowitz. It pays to pay–big five personality influences on co-operative behaviour in an incentivized and hypothetical prisoner's dilemma game. *Personality and Individual Differences*, 50(2):300–304, 2011.

Robert R McCrae and Paul T Costa. Reinterpreting the myers-briggs type indicator from the perspective of the five-factor model of personality. *Journal of personality*, 57(1):17–40, 1989.

Robert R McCrae and Paul T Costa. Trait explanations in personality psychology. *European Journal of Personality*, 9(4):231–252, 1995.

Jeffery J Mondak. *Personality and the foundations of political behavior*. Cambridge University Press, 2010.

Rebecca B Morton and Kenneth C Williams. *Experimental political science and the study of causality: From nature to the lab*. Cambridge University Press, 2010.

Isabel Briggs Myers, Mary H McCaulley, and Robert Most. *Manual: A guide to the development and use of the Myers-Briggs Type Indicator*. Consulting Psychologists Press Palo Alto, CA, 1985.

Alexander Peysakhovich, Martin A Nowak, and David G Rand. Humans display a cooperative phenotypethat is domain general and temporally stable. *Nature communications*, 5, 2014.

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.

John A Sautter, Levente Littvay, and Brennen Bearnes. A dual-edged sword: Empathy and collective action in the prisoner's dilemma. *The ANNALS of the American Academy of Political and Social Science*, 614(1):154–171, 2007.

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.

Vernon L Smith. Experimental economics: Induced value theory. *The American Economic Review*, 66(2):274–279, 1976.

Development Team Stan. Rstan: the r interface to stan, version 2.9.0. http://mc-stan.org, 2016.

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 and Julian Higgins. How should meta-regression analyses be undertaken and interpreted? *Statistics in medicine*, 21(11):1559–1573, 2002.

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.

Rebecca M Turner, Rumana Z Omar, Min Yang, Harvey Goldstein, Simon G Thompson, et al. A multilevel model framework for meta-analysis of clinical trials with binary outcomes. *Statistics in medicine*, 19(24):3417–3432, 2000.

Toshio Yamagishi, Nobuhiro Mifune, Yang Li, Mizuho Shinada, Hirofumi Hashimoto, Yutaka Horita, Arisa Miura, Keigo Inukai, Shigehito Tanida, Toko Kiyonari, et al. Is behavioral pro-sociality game-specific? pro-social preference and expectations of pro-sociality. *Organizational Behavior and Human Decision Processes*, 120(2):260–271, 2013.