Research Article

Psychological Science
2017, Vol. 28(12) 1821–1832

DOI: 10.1177/0956797617723726
www.psychologicalscience.org/PS
SAGE

Probing Birth-Order Effects on Narrow Traits Using Specification-Curve Analysis

Julia M. Rohrer[1,2,3], Boris Egloff[4], and Stefan C. Schmukle[3]
[1]International Max Planck Research School on the Life Course (LIFE), Max Planck Institute for Human Development, Berlin, Germany; [2]German Institute for Economic Research, Berlin, Germany; [3]Department of Psychology, University of Leipzig; and [4]Department of Psychology, Johannes Gutenberg University of Mainz

Abstract
The idea that birth-order position has a lasting impact on personality has been discussed for the past 100 years. Recent large-scale studies have indicated that birth-order effects on the Big Five personality traits are negligible. In the current study, we examined a variety of more narrow personality traits in a large representative sample (n = 6,500–10,500 in between-family analyses; n = 900–1,200 in within-family analyses). We used specification-curve analysis to assess evidence for birth-order effects across a range of models implementing defensible yet arbitrary analytical decisions (e.g., whether to control for age effects or to exclude participants on the basis of sibling spacing). Although specification-curve analysis clearly confirmed the previously reported birth-order effect on intellect, we found no meaningful effects on life satisfaction, locus of control, interpersonal trust, reciprocity, risk taking, patience, impulsivity, or political orientation. The lack of meaningful birth-order effects on self-reports of personality was not limited to broad traits but also held for more narrowly defined characteristics.

Keywords
birth-order effects, siblings, trust, risk taking, reciprocity, political attitudes, specification-curve analysis, open materials

Received 4/28/17; Revision accepted 7/11/17

Recent analyses of large-scale samples have failed to detect meaningful effects of birth order on the Big Five personality traits (Damian & Roberts, 2015; Rohrer, Egloff, & Schmukle, 2015). These results confirmed the findings of a 45-year-old meta-analysis: Schooler (1972) concluded that there was no reliable evidence for birth-order effects on personality. However, hypotheses motivated by theories about the relationship between birth order and personality were not originally formulated in terms of the Big Five. For example, Sulloway's (1997) influential framework of the family niche theory was partly motivated by the observation that firstborn scientists were likely to embrace conservative scientific paradigms—supposedly because firstborns are more likely to identify with their parents and are thus more likely to embrace conservative values. His theory also predicted that those born later would take more risks (Sulloway, 1997) because they needed to explore to find their role in the family and because their lower life expectancy reduced the costs of risk taking (Sulloway & Zweigenhaft, 2010). Political orientation and risk taking, however, are not directly represented in the Big Five framework.

Following this line of reasoning, a large number of studies have addressed birth-order effects on more narrow traits such as locus of control (Hughes, 2005), trust and reciprocity (Courtiol, Raymond, & Faurie, 2009), life satisfaction (Shao, Yao, Li, & Huang, 2013), and risk taking (Sulloway & Zweigenhaft, 2010). Despite the richness of these studies, their results failed to provide conclusive evidence for birth-order effects on these (and other) narrow traits. For example, regarding locus of control, although some studies have reported that

Corresponding Author:
Julia M. Rohrer, Department of Psychology, University of Leipzig, Neumarkt 9-19, 04109 Leipzig, Germany
E-mail: julia.rohrer@uni-leipzig.de

firstborns have a more internal locus of control—supposedly because they foster control by assuming responsibility for their younger siblings (e.g., Falbo, 1981)—other studies have suggested that firstborns have a more external locus of control, purportedly because of increased parental attention (e.g., Walter & Ziegler, 1980), and still other studies have not found a difference between firstborn and later-born children (e.g., Newhouse, 1974).

It is interesting that results have often been more specific and less parsimonious when taking a closer look. For example, Walter and Ziegler (1980) found that firstborn and last-born children of families with three or more siblings had a more external locus of control compared with the middle children. In addition, Hughes (2005) reported a complex interaction in the absence of a main effect of birth-order position: Having a majority of siblings of the same sex was associated with a more external locus of control only for firstborns. However, this study's author did not account theoretically for this pattern.

This altogether incoherent pattern of results (which an anonymous reviewer trenchantly but probably correctly called "a complete mess"), along with specific justifications for each analytical strategy and elaborate explanations for each specific finding, was identified by Harris who criticized the "divide-and-conquer" strategy of birth-order research (Harris, 1999, p. 348). She observed that studies had frequently reported birth-order effects for specific subgroups (e.g., females, middle-class individuals, respondents from small families) and pointed out that replications of such interaction effects are crucial to ensure their validity. Furthermore, she described various other strategies found in studies investigating birth-order effects that can result in significant findings in the absence of real effects, such as administering a large number of personality tests or splitting outcome measures into various subscales.

From a contemporary and broader point of view, Harris's (1999) thoughts about the birth-order literature fit nicely with concerns regarding the reproducibility of research in general and psychological research in particular (Open Science Collaboration, 2015): It has been suggested that a large number of findings can be false positives when there is greater flexibility in designs and outcomes (e.g., Ioannidis, 2005). In addition, Simmons, Nelson, and Simonsohn (2011) demonstrated that when one has a large number of so-called researcher degrees of freedom, anything can be presented as significant. Applied to the case of birth order, these methodological issues lead to the suspicion that at least some of the previously reported effects have been false-positive findings. Thus, whether earlier findings of supposedly meaningful birth-order effects can be replicated remains an open question.

When considered in combination, recent large-scale studies that mostly yielded null results (Damian & Roberts, 2015; Rohrer et al., 2015) and questions about the robustness of earlier studies suggest that attempting to hunt down the effects of birth order on personality might not be particularly promising. However, scientists continue to research this topic today (see recent publications such as Black, Grönqvist, & Öckert, 2017; Ergüner-Tekinalp & Terzi, 2016; Lehmann, Nuevo-Chiquero, & Vidal-Fernandez, 2016; Salmon, Cuthbertson, & Figueredo, 2016). In light of the vast, unclear, and contradictory literature on birth-order effects, a critical, robust assessment of these effects on personality using state-of-the-art methods might help scientists focus their research attempts.

In this study, we investigated the effects of birth order on a range of potentially interesting narrow personality variables from the German Socio-Economic Panel (SOEP). The SOEP is a large-scale, nationally representative panel study that allows for both between- and within-family analyses of birth-order effects. Given the surprising number of different analytic approaches that can be found in the birth-order literature, we chose specification-curve analyses (Simonsohn, Simmons, & Nelson, 2015) to assess the robustness of findings across a large range of different model specifications. Thus, our goal was to evaluate the evidence for birth-order effects across the large number of researcher degrees of freedom inherent to traditional approaches to birth-order research.

Method

Data and respondents

Data came from the SOEP, an ongoing study of private households in Germany and their members (Wagner, Frick, & Schupp, 2007). The SOEP was launched in 1984 and has been refreshed multiple times since then to ensure that it remains representative of the German population.

Sample sizes varied across the analyses because of missing values for certain dependent variables. The smallest sample for analyses involved locus of control (n = 6,585 for the between-family analyses, and n = 925 for the within-family analyses); the largest sample involved life satisfaction (n = 10,628 for the between-family analyses, and n = 1,245 for the within-family analyses). Furthermore, certain specifications led to the exclusion of respondents (e.g., because of the number of siblings or the existence of step-, half-, or adoptive siblings within the same household). On average, participants were 51.17 years old (SD = 17.75) in 2013, and 53.93% were women.

Birth-order position

In 2013, respondents answered a number of questions about their siblings, providing information about the siblings' birth dates, the type of sibling relationship (full, half, step, or adoptive), and whether or not they had spent the first 15 years of life together (and, if not, how many years they had lived together). We dropped children with no siblings and twins from further analysis. In addition, to ensure that individuals from patchwork families with potentially ambiguous birth-order positions were not included, we excluded all individuals who reported living with a sibling for a time period of less than the first 15 years. Individuals who reported having more than 10 siblings had to be excluded by default because the questionnaire allowed respondents to list only 10 siblings (for a maximum sibship size of 11), which made the information insufficient to determine respondents' birth-order position. We used two definitions of birth-order position: A social one that counted all siblings who reportedly grew up together, regardless of whether they were full, half-, step-, or adoptive siblings; and a narrower one that considered only full siblings, which meant that we dropped all individuals who grew up with a half-, step-, or adoptive sibling. Furthermore, we ran analyses that differentiated between either (a) all possible birth-order positions (first, second, third, etc.) or (b) only firstborn and laterborn children.

Personality measures

Personality measures were chosen on the basis of the items included in the SOEP questionnaires from 2010 to 2014. This range of years was chosen because (a) later waves of the data were not available at the time of the conception of this study and (b) earlier waves would have resulted in substantially smaller sample sizes because respondents could be included in the analyses only if they also had taken the sibling questionnaire in 2013.

Multiple measures showed age trends. Because birth-order position can be confounded with age—especially in within-family analyses, in which every firstborn is by definition older than the children born later—we calculated age-controlled personality scores for all measures that were included. To do so without imposing a specific age trajectory (without assuming, e.g., a linear relationship), we used locally weighted regression to derive smoothed values that took into account scores of individuals of the same age and of younger and older individuals, and the weight of the observations decreased with increasing age differences. We then computed residuals by subtracting the smoothed score from the individual score (see Rohrer et al., 2015). Analyses were run twice: once using these age-controlled scores, and a second time using raw scores.

Locus of control. In 2010, locus of control was measured with 10 items answered on a 7-point scale (1 = *does not apply at all*, 7 = *applies fully*): for example, "How my life goes depends on myself," and "What one achieves in life mostly depends on fate or luck" (the latter item was reverse-scored). We recoded items if necessary so that higher scores indicated an internal locus of control, whereas lower scores indicated little belief in one's own control. Including all 10 items yielded a reliability coefficient (Cronbach's α) of .61. Three items ("If one engages socially or politically, one can change the social circumstances," "Success is the result of hard work," and "More important than all efforts are the skills one has") had item-total correlations below .08. We nonetheless calculated scores using all items because this was the original scale, but we also ran analyses using a 7-item version of the scale with a higher reliability of .70 (see Specht, Egloff, & Schmukle, 2013). Note that the locus-of-control scale included in the SOEP was found to be sensitive to the effects of age, gender, and education (Specht et al., 2013).

Reciprocity. In 2010, positive and negative reciprocity were assessed with three items each on a 7-point scale (1 = *does not apply at all*, 7 = *applies fully*): for example, "If someone does me a favor, I am willing to return it" (positive reciprocity) and "If somebody insults me, I will insult him likewise" (negative reciprocity). Both scales had acceptable reliabilities given their brevity (positive reciprocity: α = .61; negative reciprocity: α = .83). Reciprocity as measured in the SOEP has been linked to labor market behavior and other real-world outcomes in previous studies (e.g., Dohmen, Falk, Huffman, & Sunde, 2009).

Life satisfaction. A single-item measure of life satisfaction has been included in each wave of the SOEP: "How satisfied are you with your life, all things considered?" Participants answered on an 11-point scale (0 = *completely dissatisfied*, 10 = *completely satisfied*). Although this item reportedly has good psychometric properties (Lucas & Donnellan, 2012), we decided also to average across multiple years to arrive at a more reliable measure, which resulted in three different versions of the outcome variable life satisfaction: (a) the single item from 2013, which maximized both sample size and comparability because it was assessed in the same year as birth-order position; (b) the average across all years from 2010 to 2014 in which the respondent answered the item, which maximized the sample size and increased reliability but did not preserve the comparability of the scores across respondents; and (c) the average of the years 2012, 2013, and 2014 for individuals who answered in all 3 years, which led to a slightly smaller sample size but ensured comparability of the measure across respondents and simultaneously increased reliability.

Interpersonal trust. In 2013, interpersonal trust was assessed with three items rated on a 4-point scale (1 = *strongly agree*, 4 = *strongly disagree*): "People can generally be trusted," "Nowadays you can't rely on anyone," and "If you are dealing with strangers, it is better to be careful before trusting them" (the latter two items were reverse-coded). The scale showed acceptable reliability when we took into account its brevity (α = .62). Naef and Schupp (2009) demonstrated its good psychometric properties and validity.

Risk taking in different domains. In 2014, respondents answered six items asking for their willingness to take risks "while driving," "in financial matters," "during leisure and sport," "in your occupation," "with your health," and regarding "faith in other people," all answered on an 11-point scale (0 = *risk averse*, 10 = *risk prone*). A scale that included all six items showed good reliability (α = .83). However, because three of the items did not apply to all respondents ("driving," "financial matters," "occupation"), the complete scale could be computed for only about 80% of the sample. Therefore, we also used a risk score based on only three items that potentially applied to all respondents' life circumstances ("leisure and sport," "health," "faith in other people"; α = .68). The domain-specific risk items were found to be correlated with a range of corresponding behaviors such as investment in stocks, active sports, self-employment, and smoking (Dohmen et al., 2011).

Single-item measures of risk taking, patience, and impulsivity. In 2013, respondents responded to three single-item measures of risk taking, patience, and impulsivity using an 11-point scale (0 = *risk averse/very impatient/not at all impulsive*, 10 = *risk prone/very patient/very impulsive*). The risk-taking measure was validated in Dohmen et al. (2011); the single-item measures of patience and impulsivity were validated with an incentive-compatible intertemporal choice experiment for impatience (Vischer et al., 2013).

Political orientation. In 2014, respondents were administered a single item regarding their political view, which they rated on an 11-point scale (0 = *far left*, 10 = *far right*). The original coding was preserved so that higher scores indicated that respondents reported a political orientation further to the right. Left-right scales have been widely used in public-opinion research and have proven to be valuable for a wide range of research questions (for a brief overview, see Kroh, 2007); the choice of the specific SOEP scale was informed by a multitrait-multimethod investigation (Kroh, 2007).

Intellect. In 2013, respondents reported whether they considered themselves to be "eager for knowledge" on a 7-point scale (1 = *does not apply at all*, 7 = *applies completely*). This item was part of the openness scale of a Big Five inventory. Whereas the other three openness items on the scale were related to the imaginative/creative subdimension of openness, this single item provided a proxy measure of self-reported intellect. In an earlier study, it showed a small but significant birth-order effect (first-born children scored higher than later-born children: $d \approx -0.1$; see Rohrer et al., 2015). We thus included this item to assess whether the results of the specification-curve analysis converged with the results from our previous study, in which we tested only a selected number of specifications in separate analyses. In addition, the reanalysis of this item could be considered a test of whether specification-curve analysis allows the detection of birth-order effects using single-item measures.[1]

Specification-curve analyses

We ran between- and within-family analyses in nonoverlapping samples. Between-family analyses consisted of simple linear models in which personality traits were predicted by dummy-coded birth position while controlling for the number of siblings within respondents' sibships (included as a factor variable). Within-family analyses included dummy-coded birth position and dummy variables indicating the specific family, effectively controlling for similarity within families by introducing a sibship-specific intercept and thus estimating within-family effects. Statistically controlling for sibship size was neither necessary nor possible in these analyses because the family-specific intercept already captured the variance in outcomes associated with sibship size.

For each of the 11 outcome variables (locus of control, positive reciprocity, negative reciprocity, life satisfaction, interpersonal trust, risk taking in different domains, global risk taking, patience, impulsivity, political orientation, and intellect), we ran a specification-curve analysis according to the procedure outlined by Simonsohn et al. (2015). Model specifications included all combinations of the following:

- Different ways to calculate the outcome variable (in the case of locus of control, life satisfaction, and risk taking in different domains)

- Raw scores or age-adjusted *T* scores
- Within-family or between-family analyses
- The social definition of birth-order position or the more restrictive one limited to full siblings
- Differentiation of each birth-order position within a sibship (e.g., first, second, third) or differentiation only of firstborn from later positions
- Inclusion of all sibships, only those in which sibling spacing did not exceed 5 years, or only those in which sibling spacing exceeded 1.5 years but did not exceed 5 years between any two siblings (see Healey & Ellis, 2007)
- Exclusion of any gender effects, inclusion of the main effect of gender, or inclusion or both the main effect of gender and the interaction of birth-order position and gender
- Analysis of the complete sample (i.e., individuals who grew up in sibships with up to a total of 11 children), analysis of only individuals from sibships with 2 to 4 children (this category represented the majority of the sample), or separate analyses for sibships of 2, 3, and 4 children

This resulted in at least 720 analyses for most outcome variables. However, there were 1,440 specifications for locus of control and risk taking in different domains because we used two different versions of these scales. There were 2,160 specifications for life satisfaction because we used three different versions of this scale. The selection of these features was based on the features used in previous literature on birth-order effects; thus, we assumed that the resulting specifications were justified insofar that a study using one of them likely had a reasonable chance to be accepted by a peer-reviewed journal of good quality in the past, given that the findings were deemed interesting or novel by reviewers.

We used the estimated main effects of birth-order position in which firstborn children were differentiated from later-born children as a potentially intuitive effect-size estimate for a descriptive illustration of the results. The models including the interaction of gender and birth order were excluded from the illustration of effect sizes because they resulted in two distinct estimates of the effect of birth order.

Following the method of Simonsohn et al. (2015), we then applied a permutation technique to allow us to test how inconsistent the results were with the null hypothesis of no effect, considering all specifications jointly. We created 500 data sets by shuffling the independent variable birth-order position and applying certain constraints: (a) In the between-family sample, we shuffled between individuals from equal sibship sizes to preserve the effect of sibship size (which was not the focus of this study) and to avoid nonsensical combinations of sibship size and birth-order position (e.g., the fourth-born child among two siblings), and (b) in the within-family sample, we shuffled within sibships. The null hypothesis of no birth-order effects was by definition true in these shuffled data sets because birth-order position was assigned randomly, allowing us to investigate the distribution of specification curves under the null hypothesis.

Various test statistics can be derived from these specification curves under the null hypothesis. To be able to take into account analyses that would not result in one simple effect size (e.g., analyses distinguishing between all possible birth-order positions, analyses modeling the interaction between respondents' gender and birth-order position), we used the distribution of *p* values as the test statistic. More precisely, for each shuffled data set, we calculated the percentage of specifications in which the effect of birth-order position[2] reached the conventional significance threshold ($p < .05$). Using the distribution of this percentage across the 500 samples, we obtained a picture of what we would observe under the null hypothesis of no birth-order effect. The comparison of the observed percentage of significant values in our data with this approximation of the distribution under the null hypothesis allowed us to assess whether we could reject the null hypothesis of no birth-order effect. Specifically, the number of shuffled samples that had at least as many significant specifications as the unshuffled data divided by 500 gave us the *p* value of the permutation test, which reflected the probability of observing this many or even more significant specifications under the assumption of no birth-order effect.

Results

Table 1 shows the number of models in each specification-curve analysis that yielded significant results, as well as the *p* values from the permutation tests. For example, the analysis of positive reciprocity included a total of 720 different model specifications. Of these, 10.4% yielded *p* values less than .05. The median difference between firstborn and later-born children across these specifications was 0.017 *SD*, which indicates that later-born children scored negligibly higher on positive reciprocity. Among the 500 shuffled samples in which there was no birth-order effect by design, 77 samples had 10.4% or more specifications that yielded *p* values less than .05. Thus, the permutation test resulted in a *p* value of .154 (i.e., 77 divided by 500), which indicated that we should not reject the null hypothesis of no birth-order effect for positive reciprocity.

Table 1. Results for the Specification-Curve Analyses

	Description of specification curve (for original sample)			Permutation test with 500 shuffled samples	
Outcome	Number of specifications	Median difference between later-born and firstborn children	Significant ($p < .05$) specifications (%)	Number of shuffled samples with more significant specifications than for the original sample	p value of permutation test
Positive reciprocity	720	0.017	10.4	77	.154
Negative reciprocity	720	0.039	3.5	235	.470
Life satisfaction	2,160	−0.021	9.1	104	.208
Locus of control	1,440	−0.029	5.3	216	.432
Interpersonal trust	720	−0.055	16.3	24	.048
Risk taking	1,440	−0.007	12.7	39	.078
Single-item risk	720	0.000	1.4	384	.768
Single-item patience	720	−0.002	0.0	500	≥ .998
Single-item impulsivity	720	0.043	13.1	41	.082
Political orientation	720	0.019	2.6	254	.508
Intellect	720	−0.130	63.6	0	≤ .002

Note: For between-family analyses, n = 6,585–10,628; for within-family analyses (pooled across sibship sizes), n = 925–1,245. The median effects presented in this table are based on a subset of the specifications in which birth-order position was coded to distinguish between only firstborn and later-born children, and the interaction between birth-order position and gender was not included. These effects are expressed in standard-deviation units based on the respective outcomes in the complete sample; negative effects indicate that later-born children scored lower than firstborn children.

Figure 1 visually represents the result by displaying the estimated effect sizes of the difference between firstborn and later-born children in positive reciprocity for the different specifications (i.e., for all analyses in which birth order was coded in a binary fashion and the interaction between birth order and gender was not included). Effects are ordered by size; thus, we can see that, across specifications, the effects varied from −0.4 *SD* to slightly above +0.3 *SD* of the positive-reciprocity scale across the complete sample. The lines indicating statistically significant effects are longer, and it is easy to see that the majority of the specifications did not result in a significant (i.e., $p < .05$) birth-order effect. Furthermore, we can see that certain variations in the analyses never resulted in a significant effect, such as the analyses that excluded sibships with age gaps exceeding 5 years but not age gaps below 1.5 years, the analyses of sibships of three or four, and almost all between-family analyses.

We were also not able to reject the null hypothesis of no birth-order effect for the following outcome measures (see Table 1): negative reciprocity, life satisfaction, locus of control, risk taking in different domains, and the single-item measures of risk taking, patience, impulsivity, and political orientation.

By contrast, the null hypothesis was unambiguously rejected for the single-item measure of self-reported intellect (see Table 1 and Fig. 2). Of the 720 model specifications, 63.6% resulted in p values less than .05. None of the shuffled samples resulted in such a high number of significant specifications. However, the effect was rather small in most specifications (between −0.1 and −0.2 *SD*, *Mdn* = −0.130 *SD*) except for some larger values (approximately −0.45 *SD*) that emerged only in within-family analyses of sibships of three and that may have been exaggerated by random fluctuations. Significant effects emerged in both within- and between-family analyses. However, they did not emerge in sibships of four, perhaps because of the comparably small sample size of this specification (i.e., there were only 1,264 respondents from sibships of four in the between-family analysis of self-reported intellect compared with 5,311 respondents from sibships of two and 2,969 respondents from sibships of three). Overall, the specification-curve analysis arrived at the same conclusion as our previous analyses of the very same data: a small (*Mdn* d = −0.13) but significant decline in self-reported intellect from firstborn to later-born children (Rohrer et al., 2015).

One of the outcome measures—interpersonal trust—resulted in less clear results (see Table 1 and Fig. 3). Of the 720 model specifications, 117 (16.3%) p values fell below .05, with a median effect size of −0.055 *SD*. Only 4.8% of the randomly shuffled samples resulted in that many (or more) significant specifications. A closer look at the results revealed that all statistically

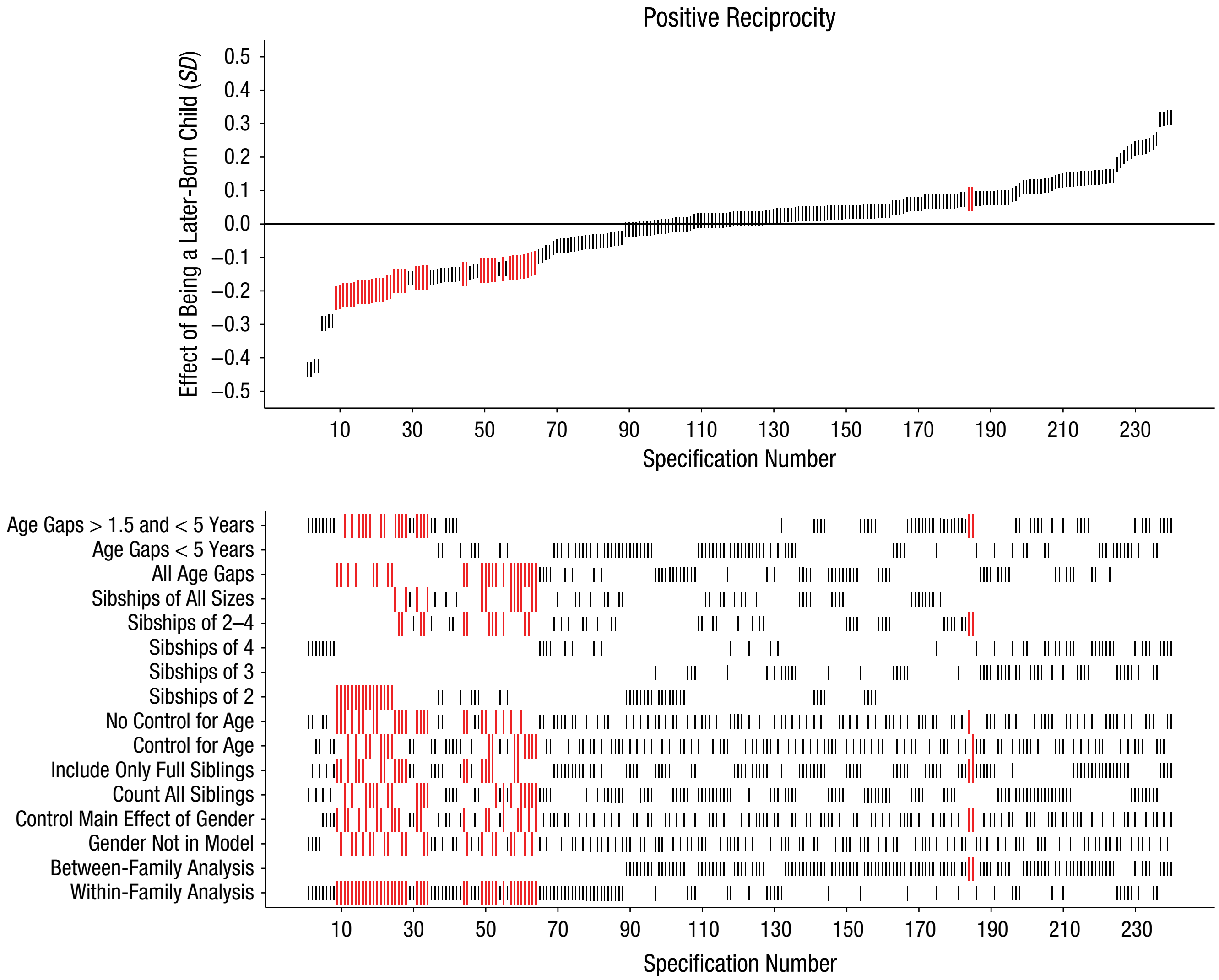

Fig. 1. Results of the specification-curve analysis of birth-order effects on self-reported positive reciprocity. The upper graph displays the estimated difference between firstborn and later-born children from each specification; the specifications are ordered by size of the effect. The lower graph shows the details of these specifications. Specifications that resulted in a significant effect of birth order ($p < .05$) are indicated by longer lines. This figure includes analyses in which birth order was treated as a dichotomous variable (firstborn vs. later-born children) and analyses that did not model the interaction between birth-order position and gender.

significant specifications were between-family analyses, whereas all within-family analyses resulted in no significant birth-order effect (all *p*s > .10). Furthermore, considering the sibships included, none of the analyses for sibships of two was significant even though the statistical power was largest for these analyses because sibships of two are the most frequent.

Discussion

In this study, we examined birth-order effects on a variety of narrow personality traits in a large representative sample. Taken together, our analyses indicated that there were no statistically significant birth-order effects across various model specifications on locus of control, negative and positive reciprocity, life satisfaction, interpersonal trust, risk taking, patience, impulsivity, and political orientation. By contrast, our analyses showed that the small effect of birth order on self-reported intellect, which had already been reported for the present sample (Rohrer et al., 2015), was robust across a wide range of possible specifications, which demonstrates that specification-curve analysis is sensitive enough to detect small effects, even on a single-item measure.

Results were somewhat ambiguous for interpersonal trust because the *p* value was just below the conventional significance threshold ($p < .05$). Note that, all

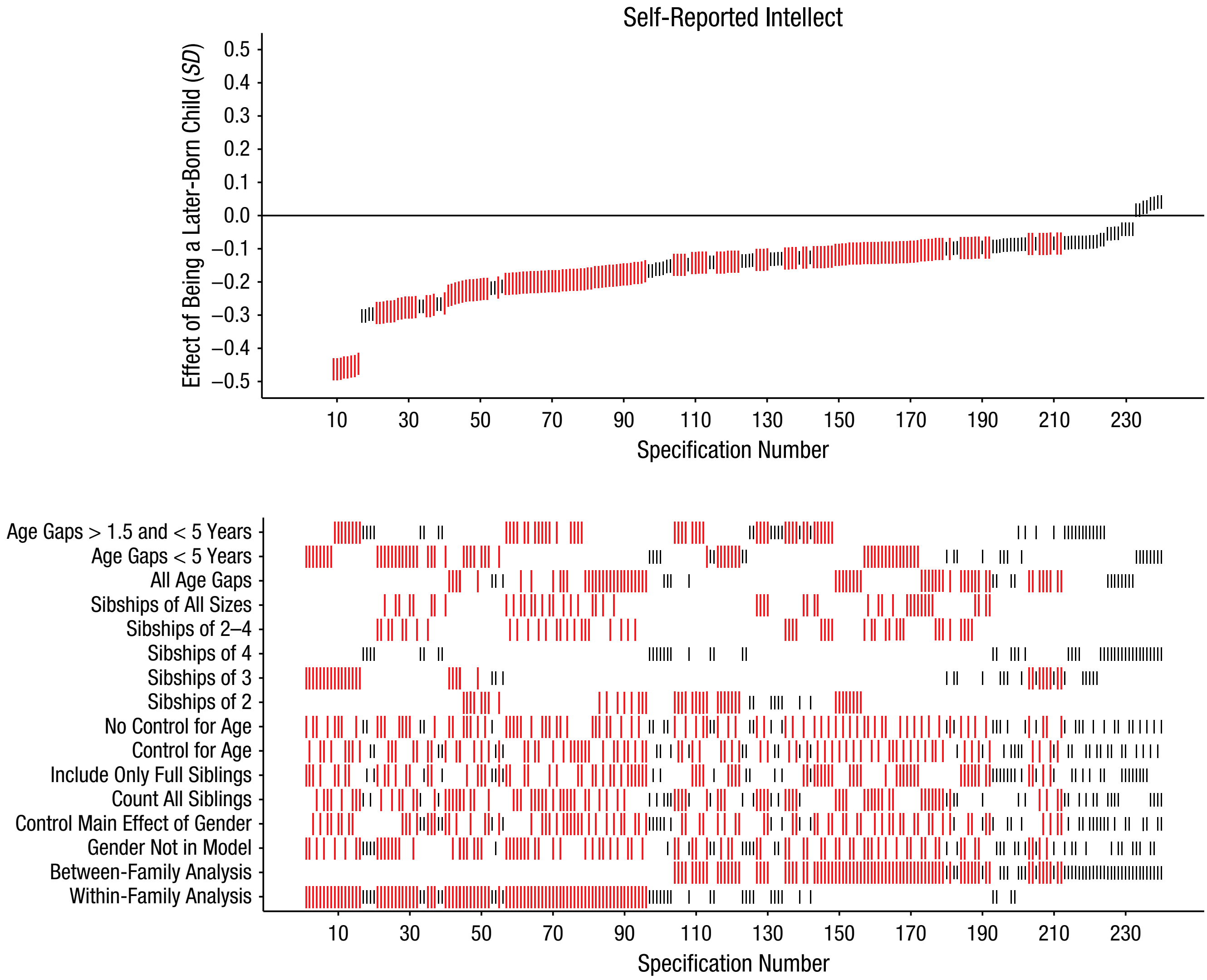

Fig. 2. Results of the specification-curve analysis of birth-order effects on self-reported intellect (eagerness for knowledge). The upper graph displays the estimated difference between firstborn and later-born children from each specification; the specifications are ordered by size of the effect. The lower graph shows the details about these specifications. Specifications that resulted in a significant effect of birth order ($p < .05$) are indicated by longer lines. This figure includes analyses in which birth order was treated as a dichotomous variable (firstborn vs. later-born children) and analyses that did not model the interaction between birth-order position and gender.

together, we tested 11 different outcome variables in separate analyses, including one that we were confident would confirm the effect we had found in a previous study. Assuming no birth-order effects for the other 10 outcomes, there would have been a 37.0% chance of obtaining at least one false-positive finding, which is why we were reluctant to assign much meaning to this effect. However, if one were inclined to take this statistically significant effect at face value, one should consider that it was (a) driven only by between-family analyses, (b) not found in sibships of two, and (c) very small (firstborn children's scores were 0.055 *SD* units higher). In addition, the effect ran contrary to a previous finding of lower levels of trust in firstborn children (Courtiol et al., 2009). Thus, the effect of birth order on interpersonal trust in our study seems neither to be convincing nor to be of considerable magnitude. Thus, we must conclude that the previously reported lack of an effect of birth order on personality (Damian & Roberts, 2015; Rohrer et al., 2015) is not simply the result of the use of very broad personality constructs, such as the Big Five, but also holds true for more narrowly defined personality traits.

We feel it is important to point out that we could have written a very different research article, centered around a single statistically significant and methodologically justified analysis. Our Results section might have read like this: "Birth-order position had a small

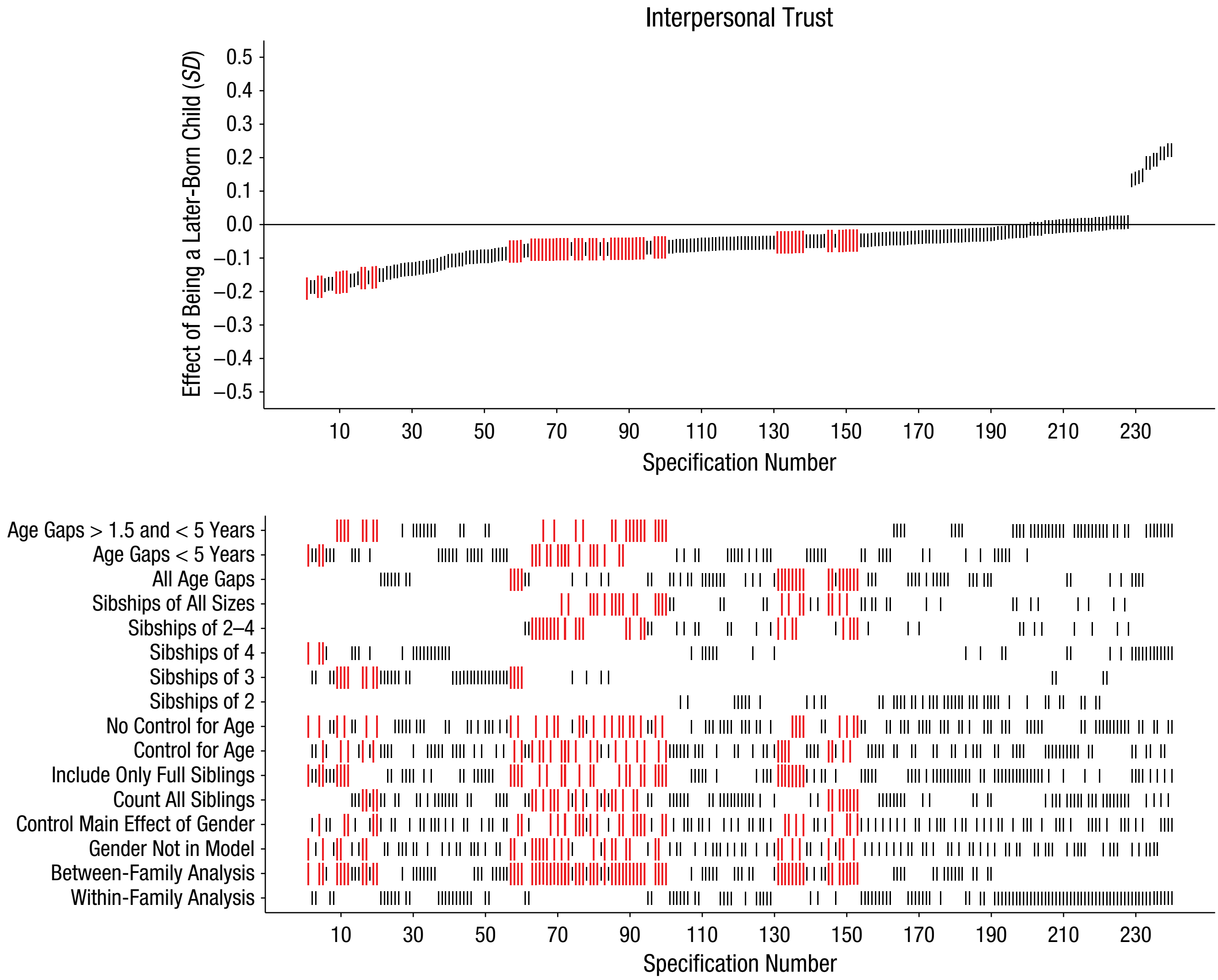

Fig. 3. Results of the specification-curve analysis of birth-order effects on self-reported interpersonal trust. The upper graph displays the estimated difference between firstborn and later-born children from each specification; the specifications are ordered by size of the effect. The lower graph shows the details about these specifications. Specifications that resulted in a significant effect of birth order ($p < .05$) are indicated by longer lines. This figure includes analyses in which birth order was treated as a dichotomous variable (firstborn vs. later-born children) and analyses that did not model the interaction between birth-order position and gender.

but significant effect on positive reciprocity: Firstborn children were more likely to pay back favors and to make an effort to help those who helped them, $p < .05$." The following specifications would lead to this result: a within-family analysis, controlling for the main effect of gender, counting only full siblings, controlling for age, including sibships of two to four people, including sibships regardless of age gaps, and a dichotomous coding of birth order (first vs. later). The analytical decisions that led to this result are easily justified post hoc: (a) Analyzing the within-family sample makes sense because this design controls for family background characteristics; (b) including all sibships, regardless of the age gaps between siblings, makes sense because it maximizes the sample size and thus the statistical power; (c) controlling for the main effect of gender is reasonable because gender might be associated with reciprocity, and so forth. In addition, we might have come up with a seemingly convincing substantive explanation for this effect and presented it as a theoretically deduced prediction in our introduction: "Because firstborn children are more likely to identify with parents who try to enforce norms of positive reciprocity among their offspring, we expect them to have internalized these norms and thus to score higher on positive reciprocity."

Conversely, we could have chosen another analysis (specification: between-family analysis, controlling for main effect of gender, counting only full siblings, controlling for age, including sibships of two to four people, limiting age gaps between consecutive siblings to greater than 1.5 years and less than 5 years, dichotomous coding of birth order) and reported: "Birth-order position had a small but significant effect on positive reciprocity: Firstborn children were less likely to pay back favors and to make an effort to help those who helped them, $p < .05$." Again, analytical decisions are easily justified post hoc: (a) Analyzing the between-family sample makes sense because it results in a much larger sample size; (b) excluding sibships with very narrow or very large age gaps makes sense because they are "atypical" and thus might not follow the typical birth-order dynamics, and so forth. In this case, our prediction in the introduction might have looked like this: "Because later-born children crucially depend on social cooperation to defend their vulnerable position against the physically superior firstborn children, we expect them to score higher on positive reciprocity."

It should be obvious that these two "defensible" analyses cannot simultaneously reflect a systematic effect of birth order on positive reciprocity, because the conclusions are diametrically opposed. Instead—as indicated by the specification-curve analysis shown in Figure 1—the data do not provide much evidence of any birth-order effect on this outcome variable, $p = .154$, and any single statistically significant analysis might be a fluke. However, the behaviors that would have led us to publish either of the two significant analyses—analyzing multiple outcome measures but reporting only those that "work," presenting exploratory findings as predicted—seem to be widespread in psychology (John, Loewenstein, & Prelec, 2012).

We believe that such a study—confidently overstating the actual evidence for a birth-order effect and framing it in a confirmatory manner—would do a disservice to a field that has already been flooded with a large number of unclear, incoherent, and even contradictory findings. Researchers might have taken our results at face value and invested their resources into studies to follow up on our *p*-hacked finding. To prevent such a waste of resources, and to ensure that psychology can accumulate insights about human behavior, researchers should rely on complete and honest reporting of the actual research process.

Finally, it is important to acknowledge that our study was limited to self-report measures. It has been argued that self-reports are not suitable for detecting birth-order effects. Sulloway (1999), for example, claimed that socially desirable responding is stronger in firstborn children, potentially canceling out existing birth-order effects. A number of studies investigating birth-order effects have instead focused on other outcomes, such as behavior in economic games (Courtiol et al., 2009; Salmon et al., 2016) or participation in dangerous sports (Sulloway & Zweigenhaft, 2010), and indeed these researchers succeeded in discovering statistically significant birth-order effects, although this observation might be less informative in the presence of publication bias (Fanelli, 2012). We acknowledge that it might be worthwhile to investigate birth-order effects on alternative outcome measures such as reports from other people or behavioral measures. However, in the face of the large number of researcher degrees of freedom observed in previous studies on birth-order effects, and given that behavioral measures might be associated with an even larger number of decisions to be made by the researcher (see, e.g., Elson, 2016; Elson, Mohseni, Breuer, Scharkow, & Quandt, 2014), we strongly recommend that such investigations should (a) use a large sample size to ensure adequate power, given the small effect sizes expected; (b) be either preregistered in detail or use specification-curve analysis; and (c) more generally follow state-of-the-art recommendations for replicable research (Asendorpf et al., 2013; Munafò et al., 2017).

Action Editor

Brent W. Roberts served as action editor for this article.

Author Contributions

All the authors developed the study concept. J. M. Rohrer performed the data analysis under the supervision of S. C. Schmukle. J. M. Rohrer drafted the manuscript, and all the authors provided critical revisions. All the authors approved the final version of the manuscript for submission.

Acknowledgments

We thank Uri Simonsohn, Joseph P. Simmons, and Leif D. Nelson, who shared their analysis scripts with us. The Socio-Economic Panel (SOEP) data were made available by the German Institute for Economic Research (DIW).

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

Open Practices

All materials have been made publicly available via the Open Science Framework and can be accessed at https://osf.io/4rtv2/. The complete Open Practices Disclosure for this article can be found at http://journals.sagepub.com/doi/suppl/10.1177/0956797617723726. This article has received the badge for Open Materials. More information about the Open Practices badges can be found at https://www.psychologicalscience.org/publications/badges.

Notes

1. We reanalyzed all Big Five personality traits using specification-curve analysis, and the results are available at https://osf.io/4rtv2/.

2. In analyses including the interaction between gender and birth-order position, the p value of interest resulted from the joint test of the main effect of birth order and its interaction with gender against zero.

References

Asendorpf, J. B., Conner, M., De Fruyt, F., De Houwer, J., Denissen, J. J. A., Fiedler, K., . . . Wicherts, J. M. (2013). Recommendations for increasing replicability in psychology. *European Journal of Personality*, *27*, 108–119. doi:10.1002/per.1919

Black, S. E., Grönqvist, E., & Öckert, B. (2017). *Born to lead? The effect of birth order on non-cognitive abilities* (IZA Discussion Paper No. 10560). Retrieved from Institute for Labor Economics Web site: http://ftp.iza.org/dp10560.pdf

Courtiol, A., Raymond, M., & Faurie, C. (2009). Birth order affects behaviour in the investment game: Firstborns are less trustful and reciprocate less. *Animal Behaviour*, *78*, 1405–1411. doi:10.1016/j.anbehav.2009.09.016

Damian, R. I., & Roberts, B. W. (2015). The associations of birth order with personality and intelligence in a representative sample of U.S. high school students. *Journal of Research in Personality*, *58*, 96–105. doi:10.1016/j.jrp.2015.05.005

Dohmen, T., Falk, A., Huffman, D., & Sunde, U. (2009). Homo reciprocans: Survey evidence on behavioural outcomes. *The Economic Journal*, *119*, 592–612. doi:10.1111/j.1468-0297.2008.02242.x

Dohmen, T., Falk, A., Huffman, D., Sunde, U., Schupp, J., & Wagner, G. G. (2011). Individual risk attitudes: Measurement, determinants and behavioral consequences. *Journal of the European Economic Association*, *9*, 522–550. doi:10.1111/j.1542-4774.2011.01015.x

Elson, M. (2016). Competitive reaction time task. Retrieved from http://crtt.flexiblemeasures.com/

Elson, M., Mohseni, M. R., Breuer, J., Scharkow, M., & Quandt, T. (2014). Press CRTT to measure aggressive behavior: The unstandardized use of the competitive reaction time task in aggression research. *Psychological Assessment*, *26*, 419–432. doi:10.1037/a0035569

Ergüner-Tekinalp, B., & Terzi, Ş. (2016). Coping, social interest, and psychological birth order as predictors of resilience in Turkey. *Applied Research in Quality of Life*, *11*, 509–524. doi:10.1007/s11482-014-9378-3

Falbo, T. (1981). Relationships between birth category, achievement, and interpersonal orientation. *Journal of Personality and Social Psychology*, *41*, 121–131. doi:10.1037/0022-3514.41.1.121

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*, 891–904. doi:10.1007/s11192-011-0494-7

Harris, J. R. (1999). *The nurture assumption: Why children turn out the way they do*. New York, NY: Free Press.

Healey, M. D., & Ellis, B. J. (2007). Birth order, conscientiousness, and openness to experience: Tests of the family-niche model of personality using a within-family methodology. *Evolution & Human Behavior*, *28*, 55–59. doi:10.1016/j.evolhumbehav.2006.05.003

Hughes, B. M. (2005). Birth order and locus of control revisited: Sex of siblings as a moderating factor. *Psychological Reports*, *97*, 419–422.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine*, *2*(8), Article e124. doi:10.1371/journal.pmed.0020124

John, L. K., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, *23*, 524–532.

Kroh, M. (2007). Measuring left–right political orientation: The choice of response format. *Public Opinion Quarterly*, *71*, 204–220.

Lehmann, J. Y. K., Nuevo-Chiquero, A., & Vidal-Fernandez, M. (2016). The early origins of birth order differences in children's outcomes and parental behavior. *Journal of Human Resources*. Advance online publication. doi:10.3368/jhr.53.1.0816-8177

Lucas, R. E., & Donnellan, M. B. (2012). Estimating the reliability of single-item life satisfaction measures: Results from four national panel studies. *Social Indicators Research*, *105*, 323–331. doi:10.1007/s11205-011-9783-z

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie, N., . . . Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behaviour, 1*, Article 0021. doi:10.1038/s41562-016-0021

Naef, M., & Schupp, J. (2009). *Measuring trust: Experiments and surveys in contrast and combination* (IZA Discussion Paper No. 4087). Retrieved from Institute for Labor Economics Web site: http://ftp.iza.org/dp4087.pdf

Newhouse, R. C. (1974). Locus of control and birth order in school children. *Journal of Clinical Psychology*, *30*, 364–365.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*, Article aac4716. doi:10.1126/science.aac4716

Rohrer, J. M., Egloff, B., & Schmukle, S. C. (2015). Examining the effects of birth order on personality. *Proceedings of the National Academy of Sciences, USA*, *112*, 14224–14229. doi:10.1073/pnas.1506451112

Salmon, C., Cuthbertson, A. M., & Figueredo, A. J. (2016). The relationship between birth order and prosociality: An evolutionary perspective. *Personality and Individual Differences*, *96*, 18–22. doi:10.1016/j.paid.2016.02.066

Schooler, C. (1972). Birth order effects: Not here, not now! *Psychological Bulletin*, *78*, 161–175. doi:10.1037/h0033026

Shao, Y., Yao, X., Li, Y., & Huang, Z. (2013). Personality and life satisfaction in China: The birth order effect under the influence of national policy. *Personality and Individual Differences*, *54*, 536–541. doi:10.1016/j.paid.29012.11.001

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359–1366. doi:10.1177/0956797611417632

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Specification curve: Descriptive and inferential statistics on all reasonable specifications. Retrieved from Social Science Research Network Web site: http://doi.org/10.2139/ssrn.2694998

Specht, J., Egloff, B., & Schmukle, S. C. (2013). Everything under control? The effects of age, gender, and education on trajectories of perceived control in a nationally representative German sample. *Developmental Psychology*, *49*, 353–364. doi:10.1037/a0028243

Sulloway, F. J. (1997). *Born to rebel: Birth order, family dynamics, and creative lives*. New York, NY: Vintage Books.

Sulloway, F. J. (1999). Birth order. In M. A. Runco & S. Pritzker (Eds.), *Encyclopedia of creativity* (Vol. 1, pp. 189–202). San Diego, CA: Academic Press.

Sulloway, F. J., & Zweigenhaft, R. L. (2010). Birth order and risk taking in athletics: A meta-analysis and study of Major League Baseball. *Personality and Social Psychology Review*, *14*, 402–416. doi:10.1177/1088868310361241

Vischer, T., Dohmen, T., Falk, A., Huffman, D., Schupp, J., Sunde, U., & Wagner, G. G. (2013). Validating an ultra-short survey measure of patience. *Economics Letters*, *120*, 142–145.

Wagner, G. G., Frick, J. R., & Schupp, J. (2007). The German Socio-Economic Panel Study (SOEP)—scope, evolution and enhancements. *Journal of Applied Social Science Studies*, *127*, 139–169.

Walter, D. A., & Ziegler, C. A. (1980). The effects of birth order on locus of control. *Bulletin of the Psychonomic Society*, *15*, 293–294.