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

Julia M. Rohrer[1,2], Boris Egloff[3], Stefan C. Schmukle[2]

[1]International Max Planck Research School on the Life Course (LIFE), Freie Universität Berlin, & German Institute for Economic Research

[2]University of Leipzig

[3]Johannes Gutenberg University Mainz

Psychological Science, in press.

Author Note

Additional online material and analysis scripts can be retrieved from https://osf.io/tf5dh/

Correspondence concerning this article should be addressed to Julia M. Rohrer, Department of Psychology, University of Leipzig, Neumarkt 9-19, 04109 Leipzig, Germany. E-mail: julia.rohrer@uni-leipzig.de

Abstract

The idea that birth-order position has a lasting impact on personality has been discussed for the last 100 years. Recent large-scale studies have indicated that birth-order effects on the Big Five are negligible. We examine 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 use Specification Curve Analysis (Simonsohn, Simmons, & Nelson, 2015) to assess evidence across a large number of defensible yet arbitrary analytical decisions (e.g., whether to control for age effects or to exclude participants on the basis of sibling spacing). Whereas 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 is not limited to broad traits but also holds for more narrowly defined characteristics.

156 words

Keywords: Birth-order effects; siblings; trust; risk-taking; reciprocity; political attitude; Specification Curve Analysis

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

Recent analyses of large-scale samples have failed to detect meaningful birth-order effects on the Big Five personality traits (Damian & Roberts, 2015; Rohrer, Egloff, & Schmukle, 2015), confirming the findings of a 40-year-old meta-analysis that concluded that there is no reliable evidence for birth-order effects on personality (Schooler, 1972). However, theoretically motivated hypotheses 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 first-born scientists were likely to embrace conservative scientific paradigms—supposedly because first-borns are more likely to identify with their parents and are thus more likely to embrace conservative values. His theory also predicted that later-borns should be more risk-taking (Sulloway, 1997) as they need to find their family role by means of exploration and because the costs of risk-taking are reduced by their lower life expectancy (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 have failed to provide conclusive evidence for birth-order effects on these (and other) narrow traits. To illustrate the state of affairs using the example of locus of control, whereas some studies have reported that first-borns have a more internal locus of control—supposedly because they foster control by assuming responsibility for their younger siblings (e.g., Falbo, 1981)—others have suggested that first-borns have a more external locus of control, purportedly because of increased parental attention (e.g., Walter & Ziegler, 1980), and still others have not found a difference between first- and later-borns (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 first- and last-borns of families with three or more siblings had a more external locus of control compared with the middle-born siblings. Also, 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 first-borns. However, this study's author did not theoretically account 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, Appendix 1). She observed that studies had frequently reported birth-order effects for specific subgroups (e.g., only females, only middle-class individuals, only 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' (1999) thoughts about the birth-order literature fit nicely with concerns regarding the reproducibility of (psychological) research in general (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 (cf. Ioannidis, 2005). Also, 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, it remains an open question whether earlier findings of supposedly meaningful birth-order effects can be replicated at all.

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

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 with sufficient power. Given the surprising number of different analytical approaches that can be found in the birth-order literature, we chose Specification Curve Analyses (Simonsohn et al., 2015) to assess the robustness of findings across a large range of different model specifications. Thus, our goal was to integrate across the large number of researcher degrees of freedom inherent to traditional approaches in birth-order research to evaluate the evidence for birth-order effects in our data.

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 since been refreshed multiple times to ensure representativity of the German population.

Sample sizes varied across the analyses due to missing values on certain dependent variables, with the smallest sample for analyses involving locus of control (*N* = 6,585 for the between-family analyses; *N* = 925 for the within-family analyses) and the largest sample for life satisfaction (*N* = 10,628 for the between-family analyses; *N* = 1,245 for the within-family analyses). Furthermore, certain specifications led to the exclusion of respondents (e.g., based on number of siblings, or existence of step/half/adoptive siblings within the same household). Please refer to the OSF project for a full breakdown of the sample size by analysis and sibship size. On average, participants were 51.17 years old in 2013 (*SD* = 17.75), and 53.93% were women.

Birth-Order Position

In 2013, respondents answered a number of questions about their siblings, providing information about their birthdate, type of sibling relationship (full, half, step, 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 only-children and twins from further analysis. Also, 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 (i.e., sibship size > 11) had to be excluded by default because the questionnaire allowed respondents to list only 10 siblings, leading to insufficient information to determine respondents' birth-order position within such large sibships. We employed two definitions of birth-order position: A social one that counts all siblings who reportedly grew up together, regardless of whether they were full, half, step, or adoptive siblings; and a narrower one that considers only full siblings, consequently dropping 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-born, second-born, third-born, and so on) or (b) only first- and later-borns.

Personality Measures

Personality measures were chosen on the basis of the items included in the SOEP questionnaires 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 because (b) earlier waves would have resulted in substantially smaller sample sizes because respondents could be included in the analyses only if they also took 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 first-born is per definition older than the second-born—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, with the weight of the observations decreasing with increasing age differences. We then computed residuals by subtracting the smoothed score from the individual score (cf. Rohrer et al., 2015). Analyses were run twice, once using these age-controlled scores, another 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*; e.g., “How my life goes depends on myself”; “What one achieves in life mostly depends on fate or luck” [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 of Cronbach’s $\alpha = .61$. Three items (“If one engages socially or politically, one can change the social circumstances”; “Success is the result of hard work”; “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 seven-item version of the scale with a higher reliability of $\alpha = .70$ (cf. Specht, Egloff, & Schmukle, 2013). Notice 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, Huggman, & 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?" 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 to additionally average across multiple years to arrive at a more reliable measure, resulting in three different versions of the outcome variable life satisfaction: (a) the single item from 2013, maximizing 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, maximizing the sample size and increasing reliability although not preserving 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, leading to a slightly smaller sample size but ensuring comparability of the measure across respondents while simultaneously increasing reliability.

Interpersonal trust. In 2013, interpersonal trust was assessed with three items ("People can generally be trusted"; "Nowadays you can't rely on anyone" [reverse-coded]; "If you are dealing with strangers, it is better to be careful before trusting them" [reverse-coded]) that were rated on a 4-point scale (1 = *strongly agree*, 4 = *strongly disagree*). The scale showed an 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 additionally used a risk score that was 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 answered three single items for the global assessment of risk-taking, patience, and impulsivity on an 11-point scale (0 = *risk-averse/very impatient/not at all impulsive*; 10 = *risk-prone/very patient/very impulsive*). These items were used as single-item measures of the respective constructs. 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, 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 (see Kroh, 2007, for a brief overview); 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 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-borns scoring higher than later-borns: $d \approx -0.1$; cf. 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 can be considered a test of whether single-item measures are capable of detecting birth-order effects by means of Specification Curve Analysis.[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 following the procedure outlined by Simonsohn et al. (2015). Model specifications included all combinations of the following variations:

– different scale variations of the outcome (relevant for locus of control, life satisfaction, risk-taking in different domains);

– either raw scores or age-adjusted T-scores;

[1] Refer to https://osf.io/tf5dh/ for a reanalysis of all Big Five personality traits using Specification Curve Analysis.

– within- or between-family analyses;

– either the social definition of birth-order position or the more restrictive one limited to full siblings;

– birth order was coded as a factor variable in which each position within a sibship was differentiated from the others (e.g., first-, second-, third-born) or else the common coding scheme that differentiated between only first- and later-borns was applied;

– all sibships, only those in which sibling spacing (age gaps between consecutive siblings) 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 (cf. Healey & Ellis, 2007);

– either gender was ignored, the main effect of gender was included, or else both the main effect of gender and the interaction of birth-order position and gender were included;

– and either the complete sample (including individuals who grew up in sibships with up to a total of 11 children) was included, only individuals from sibships with two to four children (representing the majority of the sample) were included, or analyses were run separately for sibships of two, three, or four children.

This resulted in at least 720 analyses for most outcome variables, 1,440 specifications for locus of control and risk-taking in different domains because we used two different versions of these scales, and 2,160 specifications for life satisfaction because we used three different versions of this scale.

The selection of these features was based on 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 first-borns were differentiated from later-borns 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 as they resulted in two distinct estimates of the effect of birth order.

Following 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, that is, (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 (i.e., fourth-born of two), and (b) in the within-family sample, we shuffled within sibships. The null hypothesis of no birth-order effects was per definition true in these shuffled data sets as 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 (i.e., 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 of $p < .05$. 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. Please refer to the OSF page (https://osf.io/tf5dh/) for all analysis scripts and additional figures that visually represent the results of all Specification Curve Analyses in detail.

Results

Table 1 shows the number of significant specifications per Specification Curve Analysis as well as the *p*-value from the permutation test. For example, the analysis of positive reciprocity included a total of 720 different model specifications. Of these, 10.4% yielded $p < .05$. The median difference between first- and later-borns across these specifications was 0.017 *SD*,

[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.

indicating that later-borns scored negligibly higher on positive reciprocity. Among the 500 shuffled samples in which there was no birth-order effect per design, 77 had 10.4% or more specifications that yielded $p < .05$. Thus, the permutation test resulted in a *p*-value of 77/500 = .154, which indicated that we should not reject the null hypothesis of no birth-order effect for positive reciprocity.

Figure 1 visually represents the result by displaying the estimated effect sizes of the difference between first- and later-borns 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. Statistically significant effects are highlighted in red, 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.

Table 1. *Results for the Specification Curve Analyses*

	Description of specification curve (for original sample)			Permutation test with 500 shuffled samples	
Outcome	No. of specifications	Median effect between first- and later-borns in *SD*	% of significant specifications ($p < .05$)	No. of shuffled samples with ≥ significant specifications	*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 domains	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. *N* between 6,585 and 10,628 in between-family analyses; 925 to 1,245 in within-family analyses (pooled across sibship sizes). Median effects are expressed in *SD* units based on the respective outcome in the complete sample; negative effects indicate that later-borns scored lower. Median effects are based on a subset of the specifications in which birth-order position was coded to distinguish between only first- and later-borns and did not contain the interaction between birth-order position and gender.

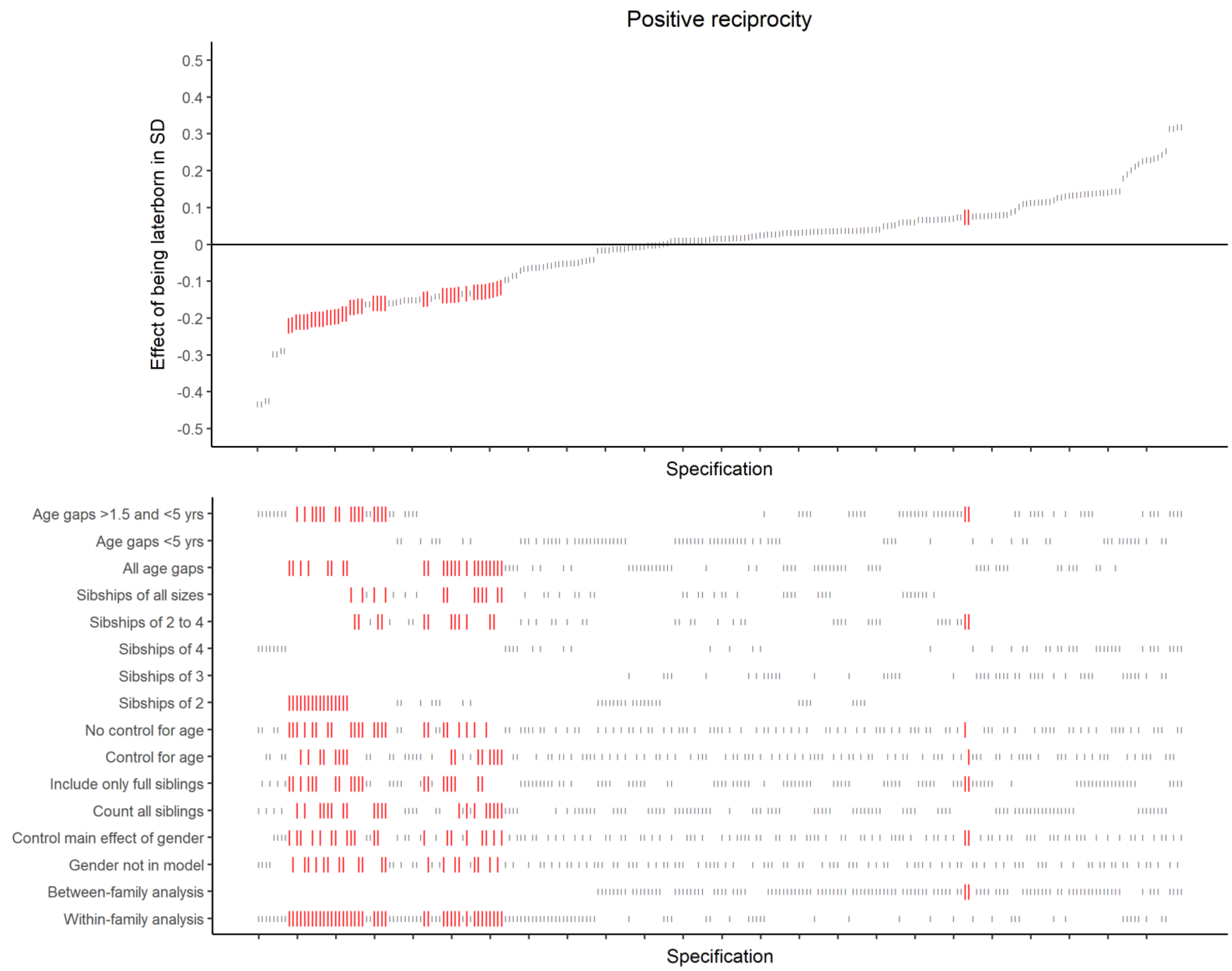

Figure 1. Results of the Specification Curve Analysis of birth-order effects on self-reported positive reciprocity. The upper part displays the estimated differences between first- and later-borns ordered by size; the lower part shows details about the various specifications. Specifications that resulted in a significant effect of birth order ($p < .05$) are highlighted. This figure includes analyses (a) in which birth order was treated as a dichotomous variable (first- vs. later-borns) and (b) that did not model the interaction between birth-order position and gender.

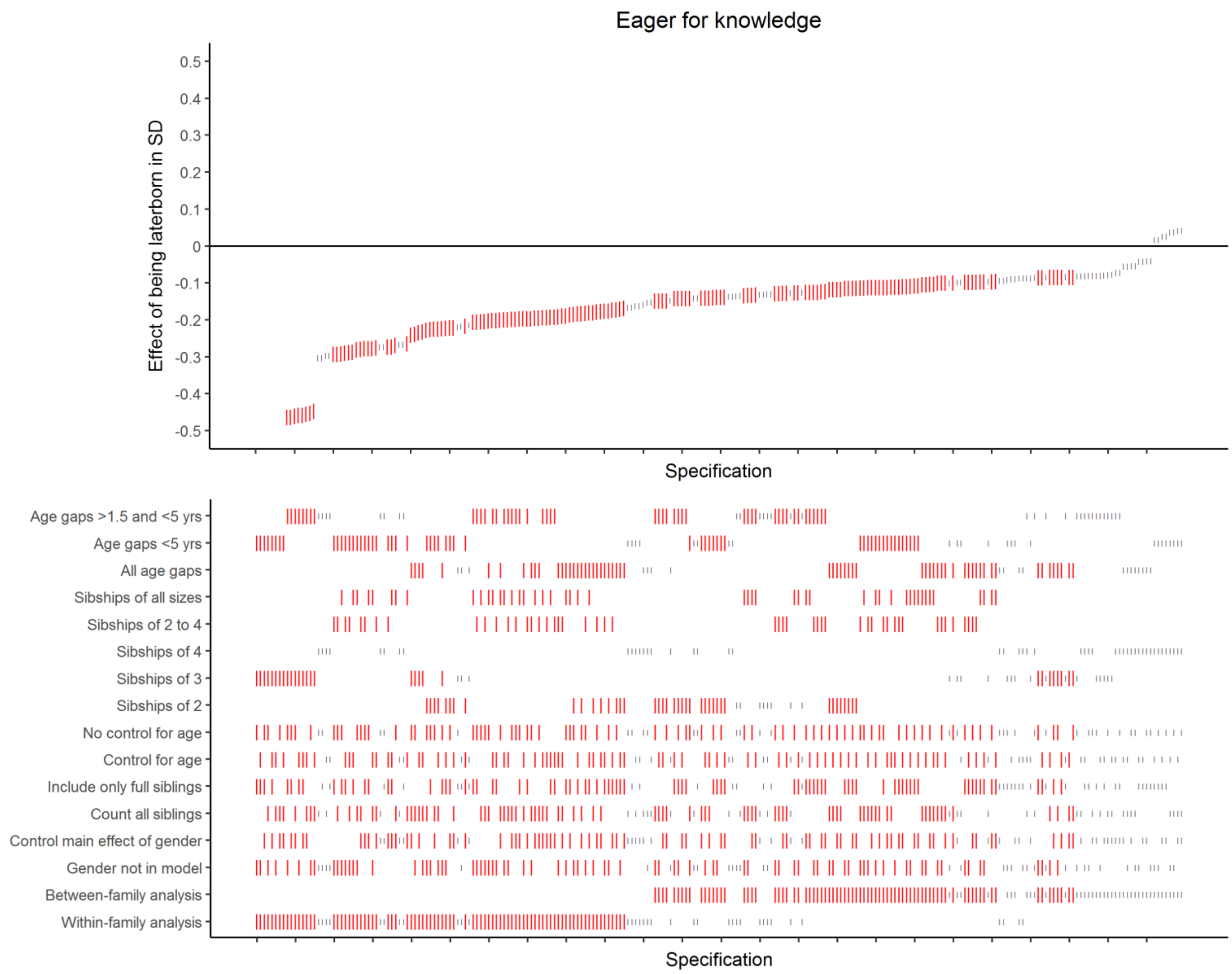

Figure 2. Results of the Specification Curve Analysis of birth-order effects on self-reported intellect. The upper part displays the estimated differences between first- and later-borns ordered by size, and the lower part shows details about the various specifications. Specifications that resulted in a significant effect of birth order ($p < .05$) are highlighted. This figure includes analyses (a) in which birth order was treated as a dichotomous variable (first- vs. later-borns) and (b) that did not model the interaction between birth-order position and gender.

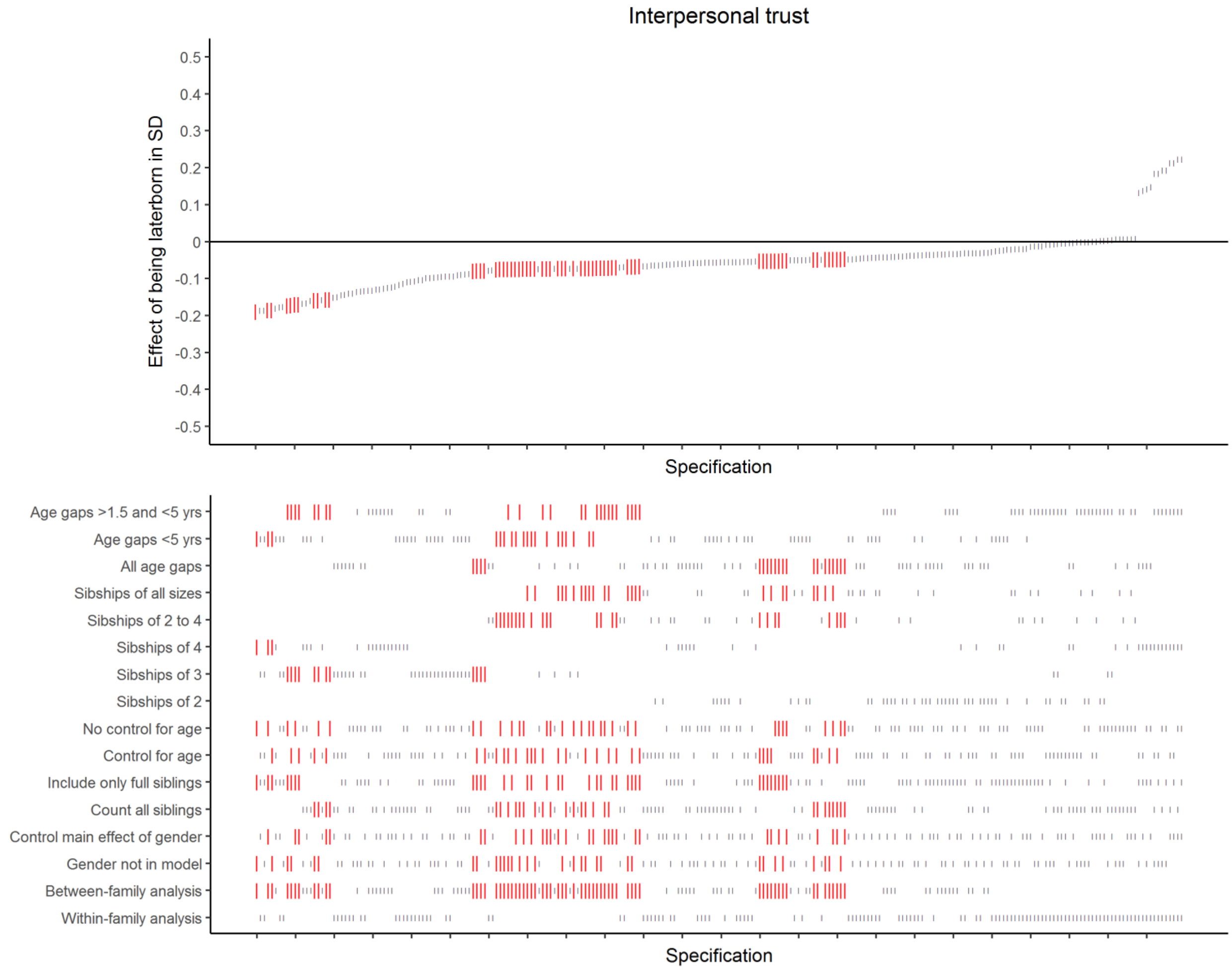

Figure 3. Results of the Specification Curve Analysis of birth-order effects on self-reported interpersonal trust. The upper part displays the estimated differences between first- and later-borns ordered by size, and the lower part shows details about the various specifications. Specifications that resulted in a significant effect of birth order ($p < .05$) are highlighted. This figure includes analyses (a) in which birth order was treated as a dichotomous variable (first- vs. later-borns) and (b) that did not model the interaction between birth-order position and gender.

Similar to the results for positive reciprocity, 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 single-item measures of risk, 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 Figure 2). Of the 720 specifications, 63.6% resulted in $p < .05$. 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 *SD* and -0.2 *SD*, *Mdn* = -0.130 *SD*) except for some larger values (ca. -0.45 *SD*) that emerged only in within-family analyses of sibships of three and 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 due to 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 vs. 5,311 respondents from sibships of two and 2,969 respondents from sibships of three). Overall, the Specification Curve Analysis arrived at exactly the same conclusion as our previous analyses of the very same data, that is, a small (*Mdn d* = -0.13) but significant decline in self-reported intellect from first-borns to later-borns (Rohrer et al., 2015).

One of the outcome measures—interpersonal trust—resulted in less clear results (see Table 1 and Figure 3). Of the 720 specifications, 117 (16.3%) fell below $p < .05$ 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 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 were significant at $p < .05$ 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 more 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, demonstrating that Specification Curve Analysis is sensitive enough to detect small effects even on a single-item measure.

Results were somewhat ambiguous for interpersonal trust as the *p*-value was just below the conventional significance threshold of $p < .05$. Note that altogether, we tested 11 different outcome variables in separate analyses, including one of which 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 (3) very small (with first-borns scoring 0.055 *SD* units higher). In addition, the effect ran contrary to a previous finding of lower levels of trust in first-borns (Courtiol et al., 2009). Thus, the effect of birth order on interpersonal trust in our study seems neither convincing nor of considerable magnitude.

We thus have to 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 but significant effect on positive reciprocity: First-borns 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, only counting full siblings, controlling for age, including 2-to-4-person sibships, 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: For example, analyzing the within-family sample makes sense as this design controls for family background characteristics; including all sibships regardless of the age gaps between siblings makes sense as it maximizes the sample size and thus the statistical power; controlling for the main effect of gender is reasonable as 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: “As first-borns 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, only counting full siblings, controlling for age, including 2-to-4-person sibships, limiting age gaps between consecutive siblings to >1.5 years and < 5 years, dichotomous coding of birth order) and reported: “Birth-order position had a small but significant effect on positive reciprocity: First-borns 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: Analyzing the between-family sample makes sense as it results in a much larger sample size; excluding sibships with very narrow or very large age gaps makes sense as these 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: “As later-borns crucially depend on social cooperation to defend their vulnerable position against the physically superior first-born, 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, as 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 only reporting 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 first-borns, 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, Cuthbertson, & Figueredo, 2016) or participation in dangerous sports (Sulloway & Zweigenhaft, 2010) and indeed 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 other-reports 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, we strongly recommend that such investigations should (a) utilize a large sample size to ensure adequate power given the small to-be-expected effect sizes, (b) be either preregistered in detail or employ Specification Curve Analysis, and (c) more generally follow state-of-the-art recommendations for replicable research (Asendorpf et al., 2013; Munafò et al., 2017).

Author Contributions

All 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 authors provided critical revisions. All 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 Economic Research (DIW).

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. https://doi.org/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]. Retrieved from 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*(6), 1405–1411. https://doi.org/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*(July), 96–105. https://doi.org/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*(536), 592-612. https://doi.org/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. https://doi.org/10.1111/j.1542-4774.2011.01015.x

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*(2), 509-524. https://doi.org/10.1007/s11482-014-9378-3

Falbo, T. (1981). Relationships between birth category, achievement, and interpersonal orientation. *Journal of Personality and Social Psychology*, *41*(1), 121–131. https://doi.org/10.1037/0022-3514.41.1.121

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries. *Scientometrics*, *90*(3), 891–904. https://doi.org/10.1007/s11192-011-0494-7

Harris, J. R. (1999). *The nurture assumption: Why children turn out the way they do*. Bloomsbury.

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 and Human Behavior*, *28*(1), 55–59. https://doi.org/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), 0696–0701. https://doi.org/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(5), 524-532.

Kroh, M. (2007). Measuring left–right political orientation: The choice of response format. *Public Opinion Quarterly*, 71(2), 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*(3), 323–331. https://doi.org/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 Behavior, 1, 0021*. https://doi.org/10.1038/s41562-016-0021

Naef, M., & Schupp, J. (2009). Measuring trust: Experiments and surveys in contrast and combination. [IZA Discussion Paper]. Retrieved from http://ftp.iza.org/dp4087.pdf

Newhouse, R. C. (1974). Locus of control and birth order in school children. *Journal of Clinical Psychology*, *30*(3), 364–365.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251). https://doi.org/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*, *112*(46), 14224–14229. https://doi.org/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. https://doi.org/10.1016/j.paid.2016.02.066

Schooler, C. (1972). Birth order effects: Not here, not now! *Psychological Bulletin*, *78*(3), 161–175. https://doi.org/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*(4), 536–541. https://doi.org/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*(11), 1359–1366. https://doi.org/10.1177/0956797611417632

Simonsohn, U., Simmons, J. P., & Nelson, L. D. (2015). Specification Curve: Descriptive and inferential statistics on all reasonable specifications. *SSRN Electronic Journal*, (November), 1–18. https://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*(2), 353–64. https://doi.org/10.1037/a0028243

Sulloway, F. J. (1997). *Born to rebel: Birth order, family dynamics, and creative lives*. 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*(4), 402–416. https://doi.org/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, 1*(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*(5), 293–294.