

**SPECIAL SECTION ARTICLE**

# Genetic differential susceptibility on trial: Meta-analytic support from randomized controlled experiments

MARINUS H. VAN IJZENDOORN AND MARIAN J. BAKERMANS-KRANENBURG

*Leiden University*

**Abstract**

The most stringent test of differential susceptibility theory is provided by randomized control trials examining the moderating role of genetic markers of differential susceptibility in experimental manipulations of the environment (Gene  $\times$  Experimental Environment interactions), being at least 10 times more powerful than correlational Gene  $\times$  Environment interaction studies. We identified 22 experiments involving 3,257 participants with various developmental outcomes (e.g., externalizing problems, internalizing behaviors, and cognitive development). Effect sizes contrasting experimental versus control group were computed both for subjects with the polymorphism considered indicative of heightened susceptibility (e.g., the dopamine receptor D4 gene seven-repeat allele and the serotonin transporter polymorphic region short allele) and others expected to be low in susceptibility (e.g., the dopamine receptor D4 gene four-repeat allele and the serotonin transporter polymorphic region short allele). Clear-cut experimental support for genetic differential susceptibility emerged: the combined effect size of the interventions for the susceptible genotypes amounted to  $r = .33$  (95% confidence interval = 0.23, 0.42;  $p < .01$ ) versus a nonsignificant  $r = .08$  (95% confidence interval =  $-0.02, 0.17$ ;  $p = .12$ ) for the hypothesized nonsusceptible genotypes. Macrotrials showed more evidence of genetic differential susceptibility than microtrials, and differential susceptibility was more clearly observed in trials with externalizing and cognitive outcomes than with internalizing problems. This meta-analysis shows proof of principle for genetic differential susceptibility and indicates that it is time to explore its mechanisms and limits. The concept of differential susceptibility alters the idea of constitutional “risk” factors (reactive temperament and risk genotypes), and points to intervention efficacy hidden in Gene  $\times$  Environment interactions.

Differential susceptibility theory predicts that some individuals are more susceptible to environmental influences than others, not only for the worse but also for the better (Belsky, Bakermans-Kranenburg, & van IJzendoorn, 2007). Genetic differential susceptibility suggests that specific genotypes previously considered “risk genes” are markers of susceptibility (Bakermans-Kranenburg, van IJzendoorn, Pijlman, Mesman, & Juffer, 2008; Belsky et al., 2009). This implies a bold set of hypotheses fundamentally different from the conventional idea of diathesis stress that only in adverse circumstances do vulnerable individuals show deviant developmental pathways compared to their more resilient peers (Belsky & Pluess, 2013). Differential susceptibility proposes that the so-called vulnerable carriers of risk genes will outperform their peers in supportive environments, demonstrating their susceptibility instead of their vulnerability.

A bold theory requires audacious tests and thorough empirical scrutiny. Randomized control trials (RCTs) are the most stringent experimental test in the study of human development. The quantitative combination and analysis of the available experimental evidence with meta-analytic methods goes beyond the sometimes limited power of individual experiments to provide a more definite proof or refutation of differential susceptibility. In this paper, we aim to test differential susceptibility theory by examining the RCTs on genetic differential susceptibility conducted to date (i.e., March 2014), including the studies reported in the Special Section of *Development and Psychopathology* (January 2015) on genetically moderated interventions.

## Correlational Gene $\times$ Environment (G $\times$ E) Susceptibility Studies

Genetic differential susceptibility studies constitute a subclass of G  $\times$  E studies, with the additional directed hypothesis of a specific type of crossover interaction to be found in environments with a sufficiently broad range from positive to negative characteristics, and developmental or behavioral assessments that not only cover negative consequences but also positive outcomes (for reviews, see Belsky & Pluess, 2009, 2013). Until recently, most G  $\times$  E studies testing differential susceptibility have used correlational designs. The first G  $\times$  E study testing

---

The authors were supported by awards from the Netherlands Organization for Scientific Research (SPINOZA prize to M.H.v.IJ. and VICI grant to M.J.B.-K.); the Gravitation programme of the Dutch Ministry of Education, Culture, and Science; and the Netherlands Organization for Scientific Research (NWO Grant 024.001.003). The authors declare no conflicts of interest.

Address correspondence and reprint requests to: Marinus H. van IJzendoorn, Graduate School of Education and Child Studies, Centre for Child and Family Studies, Leiden University, Pieter de la Court Gebouw, Wassenaarseweg 52, Leiden 2333 AK, The Netherlands; E-mail: [vanijzen@fsw.leidenuniv.nl](mailto:vanijzen@fsw.leidenuniv.nl).

the moderating effect of the genotype in a for better and for worse manner showed that children carrying the dopamine receptor D4 gene (*DRD4*) seven-repeat allele displayed the most externalizing behavior at 39 months when their mothers were observed to be insensitive during home observations at 10 months of age but the least externalizing behavior when their mothers were highly sensitive (Bakermans-Kranenburg & van IJzendoorn, 2006). Five years later, the number of studies focusing on dopamine-system related genotypes had accumulated sufficiently to meta-analyze the data ( $k = 15$ ,  $N = 1,232$ ). The meta-analysis confirmed that genotypes that in adverse contexts put children at risk for behavior problems made them also benefit more from support (Bakermans-Kranenburg & van IJzendoorn, 2011).

A larger number of correlational  $G \times E$  studies have been conducted using serotonin-system related genotypes, in particular the serotonin transporter linked polymorphic repeat (*5-HTTLPR*), as a marker of differential susceptibility ( $k = 77$ ,  $N = 9,361$ ; van IJzendoorn, Belsky, & Bakermans-Kranenburg, 2012). In the total study set, including studies with non-Caucasian and mixed ethnicities, children with short serotonin transporter (*5-HTT*) alleles were more negatively affected by adverse contexts than carriers of two long alleles with regard to negative outcomes, but they did not benefit significantly more from positive environments. This result seemed to support the diathesis–stress model that short alleles should be considered “risk” alleles only making individuals more vulnerable to environmental adversity but not more open to supportive contexts (van IJzendoorn et al., 2012). Because ethnicity was a significant moderator of the effect sizes, the analyses were repeated for studies with Caucasian participants ( $k = 52$ ,  $N = 6,626$ ). In this set, clear evidence for differential susceptibility emerged. Carriers of short alleles appeared to be more sensitive to negative as well as positive environmental influences than individuals homozygous for the long allele. Ethnicity might thus be an important moderator of genetic differential susceptibility.

### **$G \times E$ Experiments**

In addition to the more common correlational studies, a growing body of work concerns experimental  $G \times E$  studies. Genetic differential susceptibility *experiments* are experimental  $G \times E$  studies testing the “bright side” of the moderating role of genotypes that have been shown to be related to vulnerability to negative conditions (Bakermans-Kranenburg & van IJzendoorn, 2015). From the perspective of diathesis–stress theory, “vulnerable” or “risk” genotypes are associated with bad outcomes when exposed to stressful or unfavorable environments. Differential susceptibility theory predicts that carriers of the very same genotypes conferring risk in negative circumstances profit most from interventions aimed at changing the (rearing) environment for the better.

The most decisive  $G \times E$  experiments ( $G \times \text{Experimental } E$  [ $G \times eE$ ]) are RCTs fulfilling the requirement of randomized assignment of participants to control group and intervention,

that is, to a putative improvement of the environment. Genetic variation can be a fixed factor but randomized environment is a necessary condition of experimental  $G \times E$  (Bakermans-Kranenburg & van IJzendoorn, 2015). Experimental  $G \times E$  studies provide causal, not just observational and correlational, evidence that existing interventions vary in their efficacy as a function of the genetic makeup of the individuals exposed to such interventions. This is a necessary next step in the differential susceptibility research program to overcome the limitations of correlational research on differential susceptibility (van IJzendoorn et al., 2011).

Differential susceptibility experiments have at least three distinct advantages compared to correlational genetic differential susceptibility studies. First, genes and the environment are uncorrelated. In RCTs, the environment is manipulated in standard ways, and randomization breaks the potential gene–environment correlation ( $rGE$ ). Genetic factors influencing an individual’s exposure to particular environments could make those environments themselves heritable (Jaffee & Price, 2007). Correlations between genotype and environment cannot play a contaminating role in experimental interventions changing the environment, however, because genes may only moderate the effectiveness of the intervention. Of course, in correlational  $G \times E$  studies, the  $rGE$  is often set aside by ascertaining that the genetic marker is not correlated with the indicator of the environment, but this is not a definite proof of the absence of  $rGE$  because some unmeasured genetic component might be responsible for the environment. In behavioral genetic studies, environmental measures themselves have been shown to be partly heritable (Jaffee & Price, 2007), and in particular, self-report assessments of the environment might be liable to heritable response biases (Eaves et al., 1999). Only random assignment creates true independence of (change in) the environment and genetic makeup (van IJzendoorn et al., 2011).

Second,  $G \times E$  experiments avoid or decrease the risk of unequal measurement errors in the  $G \times E$  equation, that is, the varying error components in the interaction equation of genetics and environment. If genetic assessments are done in a careful way but broad or “quick and dirty” measures are used for the environment (e.g., self-reported retrospective childhood experiences), the error components are smaller for genes than for the environment, creating risks for Type 1 and Type 2 errors.  $G \times E$  findings are critically dependent on accurate assessments of both the genotype and the environment (McGuffin, Alsbabban, & Uher, 2011). A huge strength of experiments with a well-defined, standardized manipulation of a specific dimension of the environment is the reduction of measurement error in the environment. Of course, ineffective interventions not resulting in measurable change of the environment do not contribute to a reduction of measurement error in the environment. Assessing the changes in the environment might be important to check the impact of the manipulation and to examine dose–response relations between environmental change and outcome in the experimental condition (for an example, see Bakermans-Kranenburg et al., 2008). Even when the ex-

perimental intervention is effective in creating a different environment for the experimental group compared to the control subjects, it is still crucial to assess the outcome of the manipulation in a sensitive and reliable way.

Third, randomized  $G \times E$  experiments provide substantially increased statistical power compared to correlational  $G \times E$  studies. Experimental studies create more variance in the product term because interventions make experimental participants maximally different from controls. Correlational studies generally show truncated distributions at the extremes and many observations toward the center of the distributions. This is because of unavoidable selective recruitment and attrition especially in the eccentric parts of the distribution. As a result, the power of experimental  $G \times E$  studies may be more than 10 times larger than that of correlational studies. In simulations with two factors (e.g., treatment and genotype) McClelland and Judd (1993) demonstrate that, independent of the effect size of the moderator, 1,300 subjects are needed to achieve the same power in a correlational study as in an experiment with 100 subjects. They show that this does not depend on the effect size of the moderator (McClelland & Judd, 1993).

### Why Meta-Analysis?

Some  $G \times E$  experiments seem to support genetic moderation of manipulation of the environment for the better, whereas others only show a main effect of the intervention. In narrative reviews of the literature, it is rather easy to fall into the trap of counting significant and nonsignificant  $G \times E$  interactions and of letting the majority tip the balance. The narrative counting strategy has several shortcomings and risks. The most important one is that it heavily relies on null hypothesis significance testing (NHST), which during the past few years has once more come in for severe criticism from statistically sophisticated researchers (e.g., Cumming, 2014). Significance is largely a function of sample size, and NHST does not sufficiently take effect sizes and their confidence intervals into account. “Do not trust any  $p$ -value” is one of the Cumming (2014) guidelines for new statistics.

Furthermore, several major journals now recommend less reliance on NHST in order to lower the risk of (unintentional) data manipulation triggered by dependence on the magic  $p < .05$  (e.g., *Psychological Science*). For these reasons, the new statistical approaches proposed by Cumming (2014) and others avoid all significance testing and rely on effect estimation and meta-analysis to document the replicability of any finding. Cumming (2014) and many others emphasize using small- or large-scale meta-analysis to help avoid dependence on single studies, however impressive they might seem, and build a cumulative discipline based on replicated results. A quantitative or meta-analysis might generate insight in a replicable and responsible way because it allows for an estimate of the overall trend in the data and for an examination of study characteristics that may moderate the effect sizes. Thus, we decided to conduct a meta-analysis to examine whether the randomized  $G \times E$  experiments reported in the current Special

Section as well as in previous publications elsewhere support or refute differential susceptibility theory.

### Types of $G \times E$ Experiments

Three types of  $G \times E$  interventions might be differentiated, depending on the level of manipulation of the environment (Bakermans-Kranenburg & van IJzendoorn, 2015). First, *nanotrials* examine the immediate neural or behavioral responses to a small range of positive and negative stimuli, to minor manipulations of stress levels, or to subtle priming (e.g., Fox, Zoukou, Ridgewell, & Garner, 2011; Sasaki et al., 2013). Nanotrials are meant to provide insight into mechanisms of change through a small window but with a sharp focus on part of the cascade of changes to be expected in broader trials.

Second, *microtrials* address a somewhat broader component of the environment but maintain a clear focus and modest manipulation of the environment. Consider in this regard evaluations of computerized early literacy instruction with and without personalized feedback (Kegel, Bus, & van IJzendoorn, 2011; Plak, Kegel, & Bus, 2015 [this issue]) and the manipulation of social acceptance, rejection, stress, or retaliation to test its effects on aggression (Gallardo-Pujol, Andres-Pueyo, & Maydeu-Olivares, 2013; McDermott, Tingley, Cowden, Frazzetto, & Johnson, 2009; Verona, Joiner, Johnson, & Bender, 2006). Microtrials might be conducted to test proof-of-principle of whether a specific ingredient of broader and more ecological valid trials might contribute to its efficacy (see also Andersson et al., 2013; Soderqvist et al., 2012).

Third, *macrotrials*, or field trials, are broad educational, parent training, or social programs that aim at changing general life circumstances for target groups including proximal and more distal components of their environment (Albert et al., 2015 [this issue]; Bakermans-Kranenburg et al., 2008; Beach, Brody, Lei, & Philibert, 2010; Bockting, Mocking, Lok, Koeter, & Schene, 2013; Brett et al., 2015 [this issue]; Brody, Chen, & Beach, 2013; Brody, Chen, Beach, Kogan, et al. 2014; Brody et al., 2006; Brody, Yu, & Beach, 2015 [this issue]; Cicchetti, Rogosch, and Toth, 2011; Cleveland et al., 2015 [this issue]; Kohen et al., 2011; Van den Hoofdakker et al., 2012). This is the type of trial that might be implemented on a large scale and is closest to the daily realities of parents, practitioners, or policymakers. In the current study, we examined whether genetic differential susceptibility can be found across the three types of trials, by contrasting the combined effect sizes of susceptible and nonsusceptible genotypes within macro-, micro-, and nanotrials.

### Additional $G \times eE$ Moderators

Beyond type of trial, additional potential moderators of the effect sizes tested in the meta-analysis need to be considered. First, we tested whether experimental outcome was similar in  $G \times E$  studies with mostly Caucasian versus mostly non-Caucasian participants. Ethnicity may play an important

role in gene–environment interaction effects, as also evident from the divergent effects in studies with mostly (>80%) Caucasian participants versus studies with more mixed ethnicities involving the *5-HTT* gene (van IJzendoorn et al., 2012; see also Propper, Willoughby, Halpern, Capone, & Cox, 2007; Williams et al., 2003; but see Vijayendran et al., 2012).

Second, a number of intervention studies aimed at reducing aggressive and externalizing behavior, or at decreasing youth's alcohol or drug abuse, whereas other interventions aimed at decreasing internalizing symptomatology. Furthermore, some trials focused on promoting cognitive development. In the meta-analysis, we tested whether within these sets of intervention studies, the differential susceptibility effect (with larger combined effect sizes for the susceptible genotypes) can be found, to a larger or smaller extent.

Third, it is of interest to know whether specific candidate genes show larger differential susceptibility effects than others. It may be the case that some candidate genes make more of a difference for specific interventions or behavioral outcomes (e.g., monoamine oxidase A for aggression, *5-HTTLPR* for depression); but there may also be a general trend for dopamine-related genes to show the largest differential susceptibility effect due to the engagement of the dopaminergic system in general attention, motivation, and reward mechanisms (Robbins & Everitt, 1999).

## Hypotheses

The first prediction is that the combination of the intervention effects across the  $G \times eE$  RCTs for the carriers of the putative susceptible genotypes is significantly larger than the combined effect size shown for genotypes assumed to be associated with less susceptibility to environmental influences. Second, we expect that the more controlled experiments, that is, nanotrials, will show stronger differential susceptibility effects than the broader interventions, because the latter type of interventions might leave more room for error variance in the environmental component of the  $G \times E$  equation. Third, we expect to find stronger intervention differences between more susceptible and less susceptible individuals in trials with predominantly Caucasian participants, because the usual genetic suspects of differential susceptibility mainly emerged from  $G \times E$  studies on subjects of Caucasian ethnicity. Based on two previous meta-analyses on genetic differential susceptibility (Bakermans-Kranenburg & van IJzendoorn, 2011; van IJzendoorn et al., 2012), we predict stronger differential susceptibility for dopamine-system genes than for other genotypes. We do not have specific expectations for differences in effect sizes between studies aiming at decreasing externalizing behaviors versus those aiming at internalizing problems or cognitive delays.

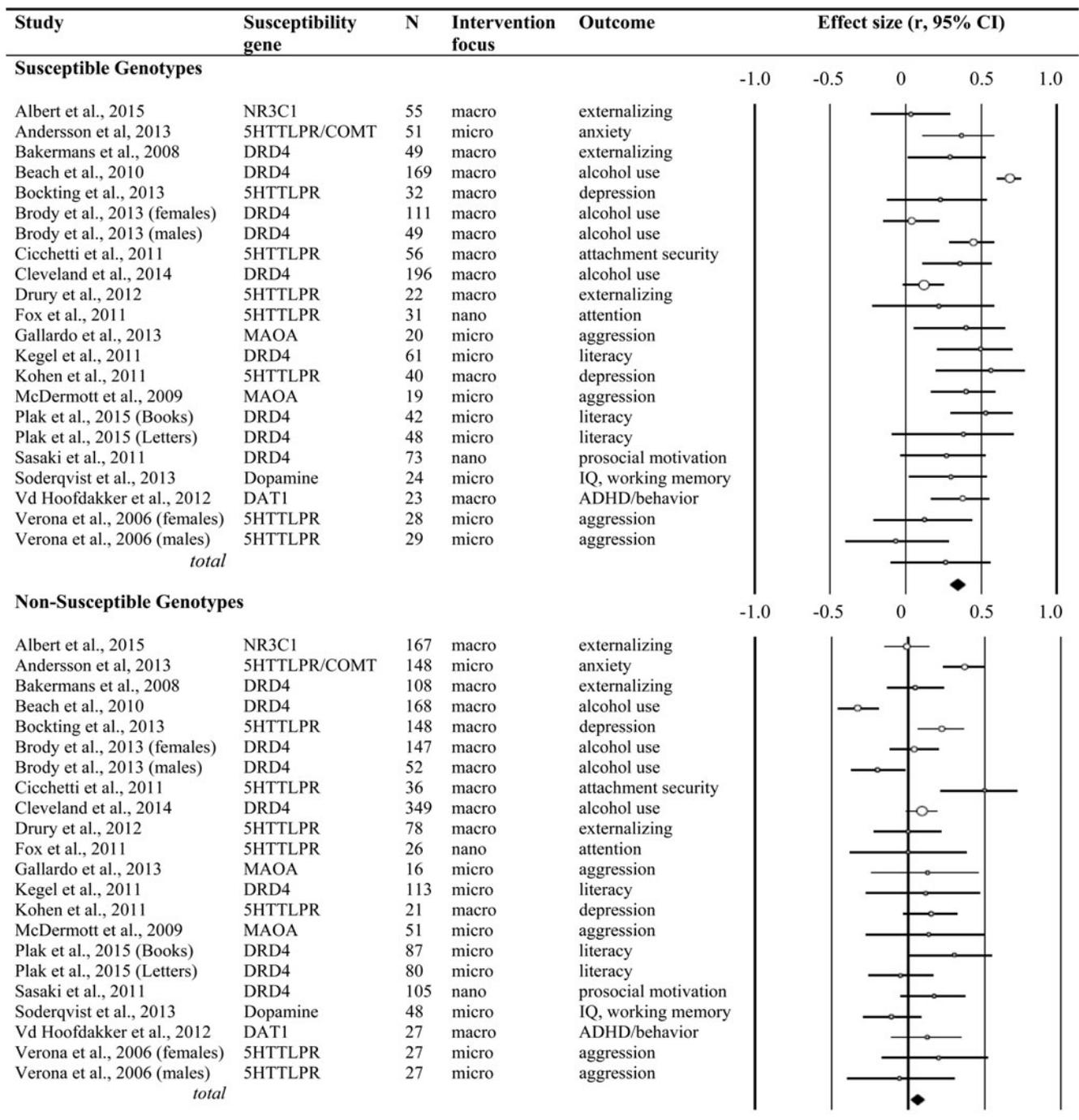
## Method

We systematically searched the databases Web of Science and MEDLINE, with the key words experiment\*, genetic\*, environment\*, intervention\*, random\* trial, differential suscepti-

bility, RCT, and  $G \times E$  in the title or abstract (the asterisks indicate that the search contained the word or word fragment). The search was restricted to RCTs with humans, and we excluded pharmacological treatment (such as treatment with drugs, alcohol infusion, or oxytocin inhalation) as the experimental intervention. We finished the search in electronic databases in March 2014, but we also included the randomized trials reported in this Special Section of *Development and Psychopathology* on experimental genetic differential susceptibility in humans. The selected studies included polymorphisms in the serotonin and dopamine system genes, glucocorticoid receptor gene *NR3C1*, and monoamine oxidase A (see Figure 1).

We identified 22 experiments involving 3,257 participants. The Comprehensive Meta-Analysis program was used to transform the results of the individual studies into the common metric of correlations and to combine effect sizes (Borenstein, Hedges, Higgins, & Rothstein, 2009; Borenstein, Rothstein, & Cohen, 2005). For every study, we computed the effect size of the experimental manipulation (i.e., experimental vs. control group) for subjects with the polymorphism considered indicative of heightened susceptibility (e.g., *DRD4* seven-repeat allele, short variant of the *5-HTT* gene) and for those with the genotype expected to convey low susceptibility (e.g., *DRD4* four-repeat allele, long variant of the *5-HTT* gene). In case of multiple outcome measures, effect sizes were combined within the study before adding the resulting effect size to the final meta-analyses. In case of multiple time points with the same outcome assessments, we selected the time point closest to the end of the intervention, to enhance comparability of effect sizes across studies.

Heterogeneity across sets of outcomes was assessed using the  $Q$ -statistic. Because most of our data sets were heterogeneous in their effect sizes and because random effects models are somewhat more conservative than fixed effects parameters in such cases, combined effect sizes and confidence intervals from random effects models are presented. We tested the influence of genotypes as susceptibility markers on the variation in combined effect sizes with the  $Q_{\text{contrast}}$  statistic in a random effects model. The  $Q_{\text{contrast}}$  statistic is based on the logic of analysis of variance, with the total variance  $Q_{\text{total}}$  partitioned into  $Q_{\text{between}}$  and  $Q_{\text{within}}$ ;  $Q_{\text{total}}$  is the variance with any grouping factors ignored; and  $Q_{\text{within}}$  for each group refers to the variances in the specific subsets of outcomes. The  $Q_{\text{within}}$  for the susceptible group thus reflects the variance in intervention effects across all studies for the subgroup with the assumed susceptible genotypes, and the  $Q_{\text{within}}$  for the nonsusceptible group reflects the variance in intervention effects across all studies for the subgroup with the nonsusceptible genotypes. Here,  $Q_{\text{between}} = Q_{\text{total}} - Q_{\text{within}}$  and is tested for significance using the chi-square distribution (Borenstein et al., 2009). A significant  $Q_{\text{contrast}}$  value indicates that the difference in effect size between susceptible genotypes and nonsusceptible genotypes is significant. We tested the significance of  $Q_{\text{contrast}}$  for the total set of RCTs and for specific subgroups of studies.



**Figure 1.** Randomized controlled genetic differential susceptibility studies; characteristics and effect sizes for the susceptible and nonsusceptible genotypes.

Following the overall contrast, we tested whether the contrast was similar in studies with mostly Caucasian participants. Next, we examined whether genetic differential susceptibility could be found across the three types of trials, by contrasting the combined effect sizes of susceptible and non-susceptible genotypes within macro-, micro-, and nanotrials. A number of intervention studies sought to reduce or prevent the development of aggressive and externalizing behavior, or

hoped to decrease youth’s alcohol consumption. Another set of studies aimed at decreasing internalizing symptoms. Yet another set of studies focused at improving cognitive development. We tested whether within these sets of intervention studies the differential susceptibility effect (with larger combined effect sizes for the susceptible genotypes) would emerge. Finally, we explored whether specific candidate genes (i.e., dopamine-related genotypes) showed larger dif-

ferential susceptibility effects than others (e.g., serotonin-related genotypes). In some cases, more than one genotype was tested for differential susceptibility, in which case the primary marker was selected (e.g., *DRD4*; Cleveland et al., 2015 [this issue]). Moderators were coded by two coders (inter-coder reliability  $\kappa > 0.80$ ); discrepancies were discussed, and consensus codes were used.

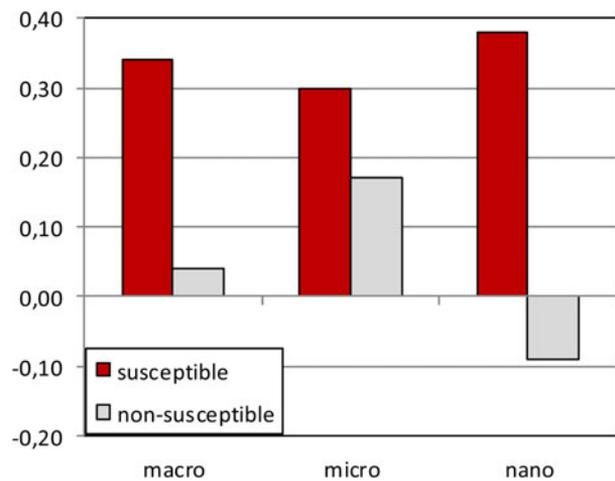
Finally, we computed the effect sizes for the difference in intervention effect between the susceptible and nonsusceptible genotypes *within* each of the RCTs. Positive effect sizes point to a larger effect for the susceptible genotypes than for the nonsusceptible genotypes. The overall combined effect size then reflects the susceptibility effect across studies.

In addition, we tested whether the distribution of these effect sizes showed any publication bias (Borenstein, 2005) favoring the publication of studies with larger differential susceptibility effects in smaller samples with the “trim and fill” method (Duval & Tweedie, 2000a, 2000b). Using this method, a funnel plot is constructed of each study’s effect size against the sample size or the standard error (usually plotted as  $1/SE$ , or precision). It is expected that this plot has the shape of a funnel, because studies with smaller sample sizes and larger standard errors have increasingly large variation in estimates of their effect size as random variation becomes increasingly influential, whereas studies with larger sample sizes have smaller variation in effect sizes (Duval & Tweedie, 2000b; Sutton, Duval, Tweedie, Abrams, & Jones, 2000). The plots should be shaped like a funnel if no data censoring is present. We used Egger’s test for detecting funnel plot asymmetry. Because smaller nonsignificant studies are less likely to be published (the “file-drawer” problem; Mullen, 1989), studies in the bottom left-hand corner of the plot are often omitted (Sutton et al., 2000). The studies considered to be symmetrically unmatched can then be trimmed; that is, their missing counterparts can be imputed or “filled” as mirror images of the trimmed outcomes, allowing for the computation of an adjusted overall effect size and confidence interval (Gilbody, Song, Eastwood, & Sutton, 2000; Sutton et al., 2000).

## Results

### Overall effect sizes

We identified 22 studies, including  $N = 3,257$  participants, 1,228 of whom were carriers of vulnerability or rather susceptibility genes. The combined effect size of the intervention effects in this susceptible group yielded a Pearson  $r = .33$  (95% confidence interval [CI] = 0.23, 0.42;  $p < .01$ ); see Table 1. The nonsusceptible group consisted of 2,029 cases, and the combined size of the intervention effects in this group was not significant,  $r = .08$  (95% CI =  $-0.02, 0.17$ ;  $p = .12$ ). Formal test of the contrast between the two combined effect sizes yielded a  $Q(1) = 13.47$ ,  $p < .01$ , showing much stronger effects for the susceptible group.



**Figure 2.** (Color online) Combined effect sizes ( $r$ ) for the susceptible and nonsusceptible genotypes in macrotrials, microtrials, and nanotrials.

### Ethnicity

In the 14 studies with more than 80% Caucasian participants ( $n = 689$  susceptible,  $n = 1,371$  nonsusceptible), we found basically the same results, with significantly larger intervention effects for the susceptible genotypes,  $r = .26$  (95% CI = 0.17, 0.34;  $p < .01$ ) than for the nonsusceptible genotypes,  $r = .12$  (95% CI = 0.05, 0.19;  $p < .01$ ). These two sets of studies were both homogeneous, and the contrast was significant,  $Q(1) = 5.63$ ,  $p = .02$ . The 8 studies with less than 80% Caucasian participants were too heterogeneous in terms of ethnicity to be combined in separate analyses.

### Intervention focus

Eleven studies were macrotrials; they were significantly more effective in carriers of susceptible genotypes ( $r = .34$ , 95% CI = 0.19, 0.47;  $p < .01$ ) than in carriers of nonsusceptible genotypes ( $r = .04$ , 95% CI =  $-0.11, 0.19$ ;  $p = .60$ ); contrast  $Q(1) = 7.76$ ,  $p < .01$  (see Figure 2). The difference among the nine microtrials was not significant (susceptible groups  $r = .30$ , 95% CI = 0.18, 0.41;  $p < .01$ , nonsusceptible groups  $r = .17$ , 95% CI = 0.07, 0.27;  $p < .01$ ); contrast  $Q(1) = 2.83$ ,  $p = .09$ . The two nanotrials showed larger effects in susceptible genotype groups ( $r = .38$ , 95% CI = 0.20, 0.54,  $p < .01$ ) than in nonsusceptible genotypes ( $r = -.09$ , 95% CI =  $-0.26, 0.09$ ,  $p = .32$ ), though the contrast between the two groups could not be tested due to the small number of studies.

### Behavioral outcome

Twelve interventions targeted *externalizing* behaviors (including alcohol abuse). Carriers of susceptible genotypes were significantly more affected,  $Q(1) = 8.19$ ,  $p < .01$ , by the interventions ( $r = .31$ , 95% CI = 0.17, 0.44,  $p < .01$ ) than their nonsusceptible peers ( $r = .01$ , 95% CI =  $-0.13, 0.16$ ,  $p = .87$ ). The four interventions targeting *internalizing*

**Table 1.** Combined effect sizes for the susceptible and nonsusceptible genotypes in Gene  $\times$  Environment experiments

	<i>k</i>	Susceptible Genotype				Nonsusceptible Genotype				Contrast <sup>d</sup> <i>Q</i>
		<i>N</i>	<i>r</i>	95% CI	<i>Q</i>	<i>N</i>	<i>r</i>	95% CI	<i>Q</i>	
Total set	22	1,228	.33**	0.23–0.42	83.11**	2,029	.08	–0.02–0.17	73.20**	13.47**
Caucasian (0.80%)	14	689	.26**	0.17–0.34	18.87	1,371	.12**	0.05–0.19	19.67	5.63*
Level										
Fieldtrials	11	802	.34**	0.19–0.47	73.63**	1,301	.04	–0.11–0.19	44.80**	7.76**
Microtrials	9	322	.30**	0.18–0.41	8.46	597	.17**	0.07–0.27	13.42	2.83
Nanotrials	2	104	.38**	0.20–0.54	0.01	131	–.09	–0.26–0.09	0.23	—
Outcome										
Externalizing	12	770	.31**	0.17–0.44	77.43**	1,217	.01	–0.13–0.16	33.08**	8.19**
Internalizing	4	154	.40**	0.25–0.53	2.39	343	.26**	0.14–0.37	4.40	2.27
Cognitive	4	175	.30**	0.16–0.43	1.89	328	.10	–0.01–0.21	2.63	4.83*
Prosocial	1	73	.38**	0.16–0.56	—	105	–.11	–0.30–0.08	—	—
Attachment	1	56	.36**	0.11–0.57	—	36	.50**	0.21–0.71	—	—
Gene system <sup>b</sup>										
Dopamine	11	845	.35**	0.21–0.47	67.25**	1,284	–.00	–0.14–0.13	30.66**	12.80**
Serotonin	7	238	.30**	0.16–0.44	8.78	363	.16*	0.01–0.29	9.47	2.13
MAOA	2	39	.46**	0.22–0.65	0.21	67	.24*	0.01–0.44	0.57	—
NR3C1	1	55	.04	–0.29–0.36	—	167	–.01	–0.25–0.24	—	—

Note: *k*, Number of study outcomes; *N*, total sample size; *d*, effect size (Cohen *d*); 95% CI, 95% confidence interval around the point estimate of the effect size;  $Q_{\text{homogeneity}}$ , homogeneity statistic;  $Q_{\text{contrast}}$ , moderation statistic.

<sup>a</sup>Subgroups with  $k < 4$  excluded from contrast.

<sup>b</sup>Andersson et al. (2013) reported on combined dopamine/serotonin system genes.

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

behaviors were effective both in the susceptible genotypes ( $r = .40$ , 95% CI = 0.25, 0.53,  $p < .01$ ) and in the nonsusceptible genotypes ( $r = .26$ , 95% CI = 0.14, 0.37,  $p < .01$ ), and the contrast was not significant,  $Q(1) = 2.27$ ,  $p = .13$ . Interventions in the *cognitive* domain showed larger effects in susceptible genotypes ( $r = .30$ , 95% CI = 0.16, 0.43,  $p < .01$ ) than in nonsusceptible genotypes ( $r = .10$ , 95% CI = –0.01, .21,  $p = .06$ ), and the contrast was significant,  $Q(1) = 4.83$ ,  $p < .05$ .

### Susceptibility genes

Considering the genetic marker of susceptibility, dopamine-related genes were markers of susceptibility; the eleven studies with dopamine-related genotypes as moderator showed larger intervention effects in susceptible genotype groups ( $r = .35$ , 95% CI = 0.21, 0.47,  $p < .01$ ) than in nonsusceptible genotypes ( $r = -.00$ , 95% CI = –0.14, 0.13,  $p = .96$ ); the contrast was significant,  $Q(1) = 12.80$ ,  $p < .01$ . Seven studies with *5-HTTLPR* as moderator showed significant combined effects in the susceptible genotype group ( $r = .30$ , 95% CI = 0.16, 0.44,  $p < .01$ ) but also in the nonsusceptible genotype ( $r = .16$ , 95% CI = 0.01, 0.29,  $p = .04$ ); the contrast was not significant,  $Q(1) = 2.13$ ,  $p = .15$ .

### Within-study differential susceptibility effects

As a final step, we computed the difference between the Fisher *Z*-transformed effect sizes for the susceptible and nonsusceptible groups *within* each study. The combined effect

size was Fisher  $Z = 0.23$  (95% CI = 0.09, 0.37;  $p < .01$ ), showing a significant combined effect for the difference between susceptible and nonsusceptible genotypes. The funnel plot of these effect sizes did not show publication bias; thus trim-and-fill was not necessary, and the Egger's test was not significant ( $t = 0.54$ ,  $p = .30$ ).

### Discussion

Clear-cut experimental support emerged for genetic differential susceptibility. In our meta-analysis on 22 RCTs including 3,257 participants, 38% of whom were carriers of susceptibility genes, the combined effect size of the interventions for the carriers of the susceptible genotypes amounted to  $r = .33$ . This is a large effect even in terms of Cohen's (1988) conventional criteria. In contrast, the hypothesized nonsusceptible group did not appear to be affected by the interventions. The two earlier meta-analyses on genetic differential susceptibility, including mostly correlational studies instead of RCTs yielded remarkably convergent evidence (Bakermans-Kranenburg & van IJzendoorn, 2011; van IJzendoorn et al., 2012), thus demonstrating the generativity and emerging validity of genetic differential susceptibility theory.

From the perspective of science or intervention practice, it might be too early to consider the available evidence for genetic differential susceptibility sufficient. Replication is paramount in science (Cumming, 2014), and although the number of replicating studies is rapidly increasing, there is still some way to go, perhaps most important in terms of further understanding the biological and psychological mechanisms of dif-

ferential change. Sound intervention practice also requires the most firm scientific foundation, and the number of populations and environments covered in the current set of studies is limited to small part of the Western, educated, industrialized, rich, and democratic societies (Henrich, Heine, & Norris, 2010).

Nevertheless, the finding of a robust genetic differential susceptibility effect is of great importance for at least two reasons. First, experimental  $G \times E$  studies exclude or control for several alternative interpretations that have plagued correlational  $G \times E$  studies, such as undetected  $rGE$  or lack of power and thus replicability. Second, they make clear that even in the absence of overall efficacy, interventions should not be discarded as ineffective and useless because they may have large impact on the substantial minority of more susceptible participants. Differential susceptibility theory allows for theory-guided examination of moderators because a small set of potential differential susceptibility markers have been proposed: biological sensitivity to context, temperamental reactivity, and genetic markers, in particular related to the dopaminergic and serotonergic systems (Ellis, Boyce, Belsky, Bakermans-Kranenburg, & van IJzendoorn, 2011).

### *Ethnicity*

Population stratification is a potential threat to the validity of genetic findings, which is why we repeated the meta-analysis in the largest and ethnically most homogeneous set of 14 studies with more than 80% Caucasian participants. Genetic differential susceptibility was clearly present in this set, but it was not possible to examine differential susceptibility in other ethnicities due to the small numbers of studies. Because the difference in effect sizes between the Caucasian carriers of susceptibility genotypes and the non-Caucasian carriers of susceptibility genotypes was small, we presume that the genetic differential susceptibility effect is not restricted to individuals with Caucasian ethnicity (see, e.g., Brody et al., 2015 [this issue], on African American families). Nevertheless, it is important to keep in mind that the biological functionality of genotypes have been shown to be different across ethnicities, and therefore, similar alleles might be responsible for contrasting levels of neurotransmitters, enzymes, or hormones regulating information processing and behavior. Besides using methods such as principal coordinates analysis to assess and control for population stratification (Cleveland et al., 2015 [this issue]), testing for the robustness of differential susceptibility evidence in ethnically homogeneous subgroups of the study remains recommended.

### *Focus and outcome*

It should be noted that genetic differential susceptibility was documented more firmly in macrotrials than in microtrials and nanotrials, despite the hypothesized large amount of error variance in the treatment manipulation in macrotrials. Moreover, it was more difficult to demonstrate differential suscep-

tibility for internalizing outcomes than for externalizing and cognitive outcomes. Of course, this pattern of results is partly due to the smaller number of studies, as in the cases of nanotrials and the focus on depression and anxiety outcome measures. With the current small set of studies, it is impossible to tease apart the moderating effects (or absence thereof) of focus and outcome because they might be intertwined, with macrotrials more often aiming at externalizing issues, and micro- and nanotrials more often focusing on cognition or anxiety. However, we should be aware of the possibility that with some types of interventions (microtrials) or specific outcomes (internalizing symptoms) the impact on carriers of the less susceptible genotypes might be too large to be exceeded by the impact on the susceptible participants. In a similar vein, some interventions might be so effective that every subject profits, especially when these subjects have been living in extremely bad environments (e.g., maltreating families; Cicchetti et al., 2011). The  $G \times E$  equation underlying genetic differential susceptibility illustrates that an extremely large (change of the)  $E$  component might drown out the influence of the  $G$  component.

### *Dopamine or serotonin?*

The majority of  $G \times E$  experiments targeted dopamine-related genes as the genetic markers of susceptibility, and they appeared to moderate intervention efficacy. This was not the case in the smaller set of studies focusing at the serotonin transporter gene (*5-HTTLPR*) as moderator. We found a similar trend in the two previous meta-analyses on mostly correlational  $G \times E$  studies. In the meta-analysis on dopamine-related genotypes, differential susceptibility was more pronounced (Bakermans-Kranenburg & van IJzendoorn, 2011) than in the meta-analysis on serotonin-related genotypes (van IJzendoorn et al., 2012), although in the latter case the set of Caucasian samples showed significant differential susceptibility. In the current meta-analysis on  $G \times E$  experiments, carriers of the putatively more susceptible *5-HTT* short alleles profited substantially from the interventions, but their counterparts with the long alleles also profited, albeit to a lesser extent. The difference was too small to be significant.

It is too early to draw conclusions from this trend because it is based on only 11 and 7 RCTs, but if confirmed in subsequent investigations, one of the alternative interpretations might reside in epigenetic change. In a study on traumatic childhood experiences and adult functioning, we found that increased methylation of the long *5-HTT* alleles made their carriers more vulnerable to symptoms of unresolved loss and posttraumatic stress (van IJzendoorn, Caspers, Bakermans-Kranenburg, Beach, & Philibert, 2010), and thus more similar to carriers of the short variants. Different genotypes might be differentially open to (de)methylation, and epigenetic changes influencing the expression of genetic markers of differential susceptibility might have to be taken into account when conducting  $G \times E$  trials (Bakermans-Kranenburg & van IJzendoorn, 2015; Meaney, 2010).

### Statistics

Various statistical approaches have been developed to examine the precise shape of  $G \times E$  interactions and to decide whether interactions fit diathesis–stress or differential susceptibility models. One of them uses the regions of significance (Kochanska, Kim, Barry, & Philibert, 2011), another the proportion of interaction or the proportion affected (Roisman et al., 2012), and yet another method was developed by Widaman et al. (2012). In this approach, the predictor is centered at the crossover point, and a confidence interval for the crossover point is estimated. When both the crossover point and its confidence interval fall within the range of observed predictor values, the interaction represents differential susceptibility; when both fall outside the range of observed predictor values, the interaction suggests diathesis stress. The strong version of differential susceptibility implies that those who are not susceptible are not at all affected by the environmental predictor; the weak version implies that some are less affected than others. The advantage of this approach is the formal testing of nested diathesis–stress and differential susceptibility models (Belsky, Pluess, & Widaman, 2013).

Although developed for correlational  $G \times E$  designs, we propose that it can be used with  $G \times E$  experiments as well. Using this method, we tested the results of our trial with video-feedback parenting training resulting in lower levels of daily cortisol production in toddlers with the *DRD4* seven-repeat allele (Bakermans-Kranenburg et al., 2008). The estimated crossover point and its confidence interval fell completely within the range of the dichotomous predictor (intervention or control group), convergent with differential susceptibility. Moreover, relaxing the constraint that the effect of the intervention for the nonsusceptible group was zero did not improve model fit significantly, and it could thus be concluded that the data supported the strong version of the differential susceptibility model (see Plak et al., 2015 [this issue]). This example demonstrates that  $G \times E$  trials may profit from the application of the Widaman et al. (2012) model-fitting approach.

### Misunderstandings

Some critical misunderstandings about  $G \times E$  trials need to be discussed. The first misunderstanding is the assumption that to find a replicable  $G \times E$  interaction, one would need a genetic main effect (e.g., Munafò Zammit, & Flint, in press; Risch et al., 2009). Rutter, Thapar, and Pickles (2009) listed three reasons for doubting the soundness of this assumption. First, a crossover interaction is not accompanied by a main effect if the crossover point is in the middle of the environmental continuum. Second, if  $G \times E$  is found in individuals without psychopathology, a main genetic effect on psychopathology might be absent. Third, statisticians have not finished debating the pros and cons of the assumption from a purely statistical point of view, and no consensus has been reached. Elsewhere we argued that the assumption is a logical implica-

tion of a diathesis–stress or vulnerability viewpoint, which becomes obsolete in a differential susceptibility perspective (Bakermans-Kranenburg & van IJzendoorn, 2015). Differential susceptibility theory implies that environmental effects are small or absent for one genotype, but present for the other genotype, for better and for worse. In that case, the two directions within one genotype cancel each other out and significant  $G \times E$  effects are found in the absence of a genetic main effect. Similarly, the overall efficacy of interventions (main effect) might go undetected because it is hidden in a  $G \times E$  interaction in a “for better and for worse” fashion.

A second misunderstanding is involved in the application of the so-called vantage sensitivity concept (Manuck & McCaffery, 2014; Pluess & Belsky, 2013; Sweitzer et al., 2012) to experimental  $G \times E$  studies. To characterize the “bright side” of differential susceptibility (Bakermans-Kranenburg & van IJzendoorn, 2011), the term *vantage sensitivity* (Manuck & McCaffery, 2014; Sweitzer et al., 2012) was introduced to suggest that some individuals profit more than others from supportive environments but adapt to negative environments in similar ways as their genotypic counterparts (Pluess & Belsky, 2013). When in a randomized trial at baseline carriers of susceptible and nonsusceptible genotypes show the same number of externalizing, internalizing, or cognitive problems, as one would hope in a sound RCT (for examples, see Kegel et al., 2011; Plak et al., 2015 [this issue]), a positive outcome of the intervention might be automatically interpreted as an example of vantage sensitivity. However, this is a non sequitur because in genetic differential susceptibility experiments the genetic moderators have been chosen *because* they were markers of differential susceptibility in correlational  $G \times E$  studies, and considered risk genes in the first place. Furthermore, the vantage interpretation of positive outcomes in a positive environment in a  $G \times E$  trial would suffer from the same problem as interpreting negative outcomes in a negative environment in terms of diathesis stress in a correlational study: both may be based on a truncated continuum of environments and through a narrow window are only able to see one side of the equation, dark or bright, but not both.

### Limitations

An obvious limitation of the current meta-analysis is the relatively small number of studies. This may reflect that RCTs require almost heroic investments in time and other resources, and that the rewards may be somewhat disappointing, for example, in terms of impact and number of publications. Nevertheless, more genetic RCTs are badly needed. The prevailing candidate gene approach in  $G \times E$  experiments may be complemented by approaches that use biologically functional genetic pathways instead of single genes to serve as markers of differential susceptibility.

Moreover, we need more experimental work with within-subject designs. A critical assumption of differential susceptibility is that the same individual would be more or less open

to environmental pressures, for better and for worse. No  $G \times E$  experiments have used designs in which the same individual is exposed both to supportive and to negative conditions, for obvious ethical reasons. However, micro- or nanotrials using mildly negative and positive manipulations, as in attention bias modification trials, may provide crucial evidence for this core differential susceptibility hypothesis (e.g., Fox et al., 2011; Hakamata et al., 2010). Furthermore, more sophisticated treatment of covariates may be necessary (Keller, 2014; for an example, see Cicchetti, Toth, & Handley, 2015 [this issue]).

A final limitation is the sparsely available evidence on biological and psychological mechanisms responsible for differential susceptibility. We need more insight into mediators that transfer intervention influences to outcomes. Genetically moderated intervention efficacy is always moderated mediation, and the search for pertinent mediators should be prioritized (for an example, see Brody et al., 2015 [this issue]).

### Implications

If the evidence for genetic differential susceptibility proves to be solid, it will have far-reaching consequences (see Bakermans-Kranenburg & van IJzendoorn, in press; for a more detailed discussion, see also Ellis et al., 2011). It may not only radically change our view of risk genotypes and related bad phenotypes (Belsky et al., 2007; Belsky & Pluess, 2009, 2013; Ellis et al., 2011) but also contribute to the solution of a tenacious problem, that is, the missing heritability in worldwide immensely large genome-wide association studies

failing to find substantial genetic main effects on complex traits and disorders (Plomin, 2013). Even genome-wide complex trait analysis (Yang, Lee, Goddard, & Visscher, 2011), a promising new tool to estimate heritability, still leaves a considerable gap with behavioral genetics estimates of the same phenotypes (Trzaskowski, Dale, & Plomin, 2013). Missing heritability might be caused by the neglect of environmental influences, or better:  $G \times E$  effects that are absorbed in the additive genetic part of the pie in many twin studies and completely neglected in genome-wide association studies or genome-wide complex trait analysis studies (van IJzendoorn et al., 2011). Experimental manipulation of E might be the best way to show that complex human behavior can only be understood as the outcome of  $G \times E$  interplay, in support of the famous dictum by Bronfenbrenner (1979) that main effects will be shown to reside in interactions.

We conclude that the available  $G \times E$  experiments on more than 3,000 subjects provide replicable support for genetic differential susceptibility across a number of intervention modalities and target populations. This is a rich harvest after less than a decade since the first  $G \times E$  experiment in the domain of human development demonstrating genetic differential susceptibility (Bakermans-Kranenburg et al., 2008). The combined RCTs have provided proof of principle that genetic differential susceptibility exists, and it is now time to explore its mechanisms and limits. The concept of differential susceptibility has already altered the way in which we interpret so-called weakly effective or ineffective (preventive) interventions, because it teaches us to look beyond intervention main effects into the hidden interactions.

### References

- Albert, W. D., Dodge, K. A., Crowley, D. M., Bates, J. E., Pettit, G. S., Lansford, J. E., et al. (2015). Developmental mediation of genetic variation in response to the Fast Track prevention program. *Development and Psychopathology*, 27, 81–95.
- Andersson, E., Ruck, C., Lavebratt, C., Hedman, E., Schalling, M., Lindfors, N., et al. (2013). Genetic polymorphisms in monoamine systems and outcome of cognitive behavior therapy for social anxiety disorder. *PLOS ONE*, 8, 1–7.
- Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2006). Gene-environment interaction of the dopamine D4 receptor (DRD4) and observed maternal insensitivity predicting externalizing behavior in preschoolers. *Developmental Psychobiology*, 48, 406–409.
- Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2011). Differential susceptibility to rearing environment depending on dopamine-related genes: New evidence and a meta-analysis. *Development and Psychopathology*, 23, 39–52.
- Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2015). The hidden efficacy of interventions: Gene  $\times$  Environment experiments from a differential susceptibility perspective. *Annual Review of Psychology*, 66, 11.1–11.29.
- Bakermans-Kranenburg, M. J., van IJzendoorn, M. H., Pijlman, F. T., Mesman, J., & Juffer, F. (2008). Experimental evidence for differential susceptibility: Dopamine D4 receptor polymorphism (DRD4 VNTR) moderates intervention effects on toddlers' externalizing behavior in a randomized controlled trial. *Developmental Psychology*, 44, 293–300.
- Beach, S. R. H., Brody, G. H., Lei, M. K., & Philibert, R. A. (2010). Differential susceptibility to parenting among African American youths: Testing the DRD4 hypothesis. *Journal of Family Psychology*, 24, 513–521.
- Belsky, J., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2007). For better and for worse: Differential susceptibility to environmental influences. *Current Directions in Psychological Science*, 16, 300–304.
- Belsky, J., Jonassaint, C., Pluess, M., Stanton, M., Brummet, B., & Williams, R. (2009). Vulnerability genes or plasticity genes? *Molecular Psychiatry*, 14, 746–754.
- Belsky, J., & Pluess, M. (2009). Beyond diathesis–stress: Differential susceptibility to environmental influences. *Psychological Bulletin*, 135, 885–908.
- Belsky, J., & Pluess, M. (2013). Beyond risk, resilience and dysregulation: Phenotypic plasticity and human development. *Development and Psychopathology*, 25, 1243–1261.
- Belsky, J., Pluess, M., & Widaman, K. F. (2013). Confirmatory and competitive evaluation of alternative gene–environment interaction hypotheses. *Journal of Child Psychology and Psychiatry*, 54, 1135–1143.
- Bockting, C. L., Mocking, R. J., Lok, A., Koeter, M. W., & Schene, A. H. (2013). Therapy genetics: The 5-HTTLPR as a biomarker for response to psychological therapy? *Molecular Psychiatry*, 18, 744–745.
- Borenstein, M. (2005). Software for publication bias. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 193–220). Chichester: Wiley.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester: Wiley.
- Borenstein, M., Rothstein, D., & Cohen, J. (2005). *Comprehensive meta-analysis: A computer program for research synthesis*. Englewood, NJ: Biostat.
- Brett, Z. H., Humphreys, K. L., Smyke, A. T., Gleason, M. M., Nelson, C. A., Zeanah, C. H., et al. (2015). Serotonin transporter linked polymorphic region (5-HTTLPR) genotype moderates the longitudinal impact of early caregiving and attachment on externalizing behavior. *Development and Psychopathology*, 27, 7–18.

- Brody, G. H., Chen, Y., & Beach, S. R. H. (2013). Differential susceptibility to prevention: GABAergic, dopaminergic, and multilocus effects. *Journal of Child Psychology and Psychiatry*, *54*, 863–871.
- Brody, G. H., Chen, Y., Beach, S. R. H., Kogan, S. M., Yu, T., DiClemente, R. J., et al. (2014). Differential sensitivity to prevention programming: A dopaminergic polymorphism-enhanced prevention effect on protective parenting and adolescent substance use. *Health Psychology*, *33*, 182–191.
- Brody, G. H., Murry, V. M., Kogan, S. M., Gerrard, M., Gibbons, F. X., Mologard, V., et al. (2006). The strong African American families program: A cluster randomized prevention trial of long-term effects and a mediational model. *Journal of Consulting and Clinical Psychology*, *74*, 356–366.
- Brody, G. H., Yu, T., & Beach, S. R. H. (2015). A differential susceptibility analysis reveals the “who and how” about adolescents’ responses to preventive interventions: Tests of first- and second-generation Gene  $\times$  Environment hypotheses. *Development and Psychopathology*, *27*, 37–49.
- Bronfenbrenner, U. (1979). *The ecology of human development*. Cambridge, MA: Harvard University Press.
- Cicchetti, D., Rogosh, F. A., & Toth, S. L. (2011). The effects of child maltreatment and polymorphisms of the serotonin transporter and dopamine receptor D4 genes on infant attachment and intervention efficacy. *Development and Psychopathology*, *23*, 357–372.
- Cicchetti, D., Toth, S. L., & Handley, E. D. (2015). Genetic moderation of interpersonal psychotherapy efficacy for low-income mothers with major depressive disorder: Implications for differential susceptibility. *Development and Psychopathology*, *27*, 19–35.
- Cleveland, H. H., Schlomer, G. L., Vandenberg, D. J., Feinberg, M., Greenberg, M., Spoth, R., et al. (2015). The conditioning of intervention effects on early adolescent alcohol use by maternal involvement and dopamine receptor D4 (*DRD4*) and serotonin transporter linked polymorphic region (*5-HTTLPR*) candidate genes. *Development and Psychopathology*, *27*, 51–67.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*, 7–29.
- Duval, S., & Tweedie, R. (2000a). A nonparametric “trim and fill” method for accounting for publication bias in meta-analysis. *Journal of the American Statistical Association*, *95*, 89–98.
- Duval, S., & Tweedie, R. (2000b). Rim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455–463.
- Eaves, L., Heath, A., Martin, N., Maes, H., Neale, M., Kendler, K., et al. (1999). Comparing the biological and cultural inheritance of personality and social attitudes in the Virginia 30,000 study of twins and their relatives. *Twin Research*, *2*, 62–80.
- Ellis, B. J., Boyce, W. T., Belsky, J., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2011). Differential susceptibility to the environment: A neurodevelopmental theory. *Development and Psychopathology*, *23*, 7–28.
- Fox, E., Zoukou, K., Ridgewell, A., & Garner, K. (2011). The serotonin transporter gene alters sensitivity to attention bias modification: Evidence for a plasticity gene. *Biological Psychiatry*, *70*, 1049–1054.
- Gallardo-Pujol, D., Andres-Pueyo, A., & Maydeu-Olivares, A. (2013). MAOA genotype, social exclusion and aggression: An experimental test of a gene–environment interaction. *Genes, Brain, and Behavior*, *12*, 140–145.
- Gilbody, S. M., Song, F., Eastwood, A. J., & Sutton, A. (2000). The causes, consequences and detection of publication bias in psychiatry. *Acta Psychiatrica Scandinavica*, *102*, 241–249.
- Hakamata, Y., Lissek, S., Bar-Haim, Y., Britton, J. C., Fox, N. A., Leibenluft, E., et al. (2010). Attention bias modification treatment: A meta-analysis toward the establishment of novel treatment for anxiety. *Biological Psychiatry*, *68*, 982–990.
- Henrich, J., Heine, S. J., & Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, *33*, 61–135. doi:10.1017/S0140525X0999152X
- Jaffee, S. R., & Price, T. S. (2007). Gene–environment correlations: A review of the evidence and implications for prevention of mental illness. *Molecular Psychiatry*, *12*, 432–442.
- Kegel, C. A. T., Bus, A. G., & van IJzendoorn, M. H. (2011). Differential susceptibility in early literacy instruction through computer games: The role of the dopamine D4 receptor gene (*DRD4*). *Mind, Brain, and Education*, *5*, 71–79.
- Keller, M. C. (2014). Gene  $\times$  Environment interaction studies have not properly controlled for potential confounders: The problem and the (simple) solution. *Biological Psychiatry*, *75*, 18–24.
- Kochanska, G., Kim, S., Barry, R. A., & Philibert, R. A. (2011). Children’s genotypes interact with maternal responsive care in predicting children’s competence: Diathesis–stress or differential susceptibility? *Development and Psychopathology*, *22*, 605–616.
- Kohen, R., Cain, K. C., Buzaitis, A., Johnson, V., Becker, K. J., Teri, L., et al. (2011). Response to psychosocial treatment in post-stroke depression is associated with serotonin transporter polymorphisms. *Stroke*, *42*, 2068–2070.
- Manuck, S. B., & McCaffery, J. M. (2014). Gene–environment interaction. *Annual Review of Psychology*, *65*, 41–70.
- McClelland, G. H., & Judd, C. M. (1993). Statistical difficulties of detecting interactions and moderator effects. *Psychological Bulletin*, *114*, 376–390.
- McDermott, R., Tingley, D., Cowden, J., Frazzetto, G., & Johnson, D. D. P. (2009). Monoamine oxidase A gene (*MAOA*) predicts behavioral aggression following provocation. *Proceedings of the National Academy of Sciences*, *106*, 2118–2123.
- McGuffin, P., Alshabban, S., & Uher, R. (2011). The truth about genetic variation in the serotonin transporter gene and response to stress and medication. *British Journal of Psychiatry*, *198*, 424–427.
- Meaney, M. J. (2010). Epigenetics and the biological definition of Gene  $\times$  Environment interactions. *Child Development*, *81*, 41–79.
- Mullen, B. (1989). *Advanced BASIC meta-analysis*. Hillsdale, NJ: Erlbaum.
- Munafò, M. R., Zammit, S., & Flint, J. (in press). Gene  $\times$  Environment interaction models: What impact should they have on clinical perceptions and practice? *Journal of Child Psychology and Psychiatry*.
- Plak, R., Kegel, C., & Bus, A. G. (2015). Genetic differential susceptibility in literacy delayed children: A randomized controlled trial on emergent literacy in kindergarten. *Development and Psychopathology*, *27*, 69–79.
- Plomin, R. (2013). Child development and molecular genetics: 14 years later. *Child Development*, *84*, 104–120.
- Pluess, M., & Belsky, J. (2013). Vantage sensitivity: Individual differences in response to positive experiences. *Psychological Bulletin*, *139*, 901–916.
- Propper, C., Willoughby, M., Halpern, C. T., Capone, M. A., & Cox, M. (2007). Parenting quality, *DRD4*, and the prediction of externalizing and internalizing behaviors in early childhood. *Developmental Psychology*, *49*, 619–632.
- Risch, N., Herrell, R., Lehner, T., Liang, K. Y., Eaves, L., Hoh, J., et al. (2009). Interaction between the serotonin transporter gene (*5-HTTLPR*), stressful life events, and risk of depression: A meta-analysis. *Journal of the American Medical Association*, *301*, 2462–2471.
- Robbins, T. W., & Everitt, B. J. (1999). Motivation and reward. In M. J. Zigmond (Ed.), *Fundamental neuroscience* (pp. 1246–1260). San Diego, CA: Academic Press.
- Roisman, G. I., Newman, D. A., Fraley, C., Haltigan, J. D., Groh, A. M., & Hayden, K. C. (2012). Distinguishing differential susceptibility from diathesis–stress: Recommendations for evaluating interaction effects. *Development and Psychopathology*, *24*, 389–409.
- Rutter, M., Thapar, A., & Pickles, A. (2009). Gene–environment interactions: Biologically valid pathway or artefact? *Archives of General Psychiatry*, *66*, 1287–1289.
- Sasaki, J. Y., Kim, H. S., Mojaverian, T., Kelley, L. D. S., Young Park, I., & Janušonis, S. (2013). Religion priming differentially increases prosocial behavior among variants of the dopamine D4 receptor (*DRD4*) gene. *Social Cognitive and Affective Neuroscience*, *8*, 209–215.
- Söderqvist, S., Nutley, S. B., Peyrard-Janvid, M., Matsson, H., Humphreys, K., Kere, J., et al. (2012). Dopamine, working memory, and training induced plasticity: Implications for developmental research. *Developmental Psychology*, *48*, 836–843.
- Sutton, A. J., Duval, S. J., Tweedie, R. L., Abrams, K. R., & Jones, D. R. (2000). Empirical assessment of effect of publication bias on meta-analyses. *British Medical Journal*, *320*, 1574–1577.
- Sweitzer, M. M., Halder, I., Flory, J. D., Craig, A. E., Gianaros, P. J., Ferrell, R. E., et al. (2012). Polymorphic variation in the dopamine D4 receptor predicts delay discounting as a function of childhood socioeconomic status: Evidence for differential susceptibility. *Social Cognitive and Affective Neuroscience*, *8*, 499–508.
- Trzaskowski, M., Dale, P. S., & Plomin, R. (2013). No genetic influence for childhood behavior problems from DNA analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*, 1048–1056.

- Van den Hoofdakker, B. J., Nauta, M. H., Dijck-Brouwer, D. A. J., Van der Veen-Mulders, L., Sytma, S., Emmelkamp, P. M., et al. (2012). Dopamine transporter gene moderates response to behavioral parent training in children with ADHD: A pilot study. *Developmental Psychology, 48*, 567–574.
- van der Doelen, R. H. A., Arnoldussen, I. A., Homberg, J. R., & Kozicz, T. (2015). Early life adversity and serotonin transporter gene variation affect DNA methylation of the corticotropin-releasing factor gene promoter region in the adult rat brain. *Development and Psychopathology, 27*, 123–135.
- van IJzendoorn, M. H., Bakermans-Kranenburg, M. J., Belsky, J., Beach, S., Brody, G., Dodge, K. A., et al. (2011). Gene-by-environment experiments: A new approach to finding the missing heritability. *Nature Reviews Genetics, 12*, 881.
- van IJzendoorn, M. H., Belsky, J., & Bakermans-Kranenburg, M. J. (2012). Serotonin transporter genotype 5-HTTLPR as a marker of differential susceptibility? A meta-analysis of child and adolescent gene-by-environment studies. *Translational Psychiatry, 2*, 1–6.
- van IJzendoorn, M. H., Caspers, K., Bakermans-Kranenburg, M. J., Beach, S. R. H., & Philibert, R. (2010). Methylation matters: Interaction between methylation density and 5-HTT genotype predicts unresolved loss or trauma. *Biological Psychiatry, 68*, 405–407.
- Verona, E., Joiner, T. E., Johnson, F., & Bender, T. W. (2006). Gender specific gene-environment interactions on laboratory-assessed aggression. *Biological Psychiatry, 71*, 33–41.
- Vijayendran, M., Cutrona, C., Beach, S. R. H., Brody, G. H., Russell, D., & Philibert, R. A. 2012. The relationship of the serotonin transporter (SLC6A4) extra long variant to gene expression in an African American sample. *American Journal of Medical Genetics, 159B*, 611–612.
- Widaman, K. F., Helm, J. L., Castro-Schilo, L., Pluess, M., Stallings, M. C., & Belsky, J. (2012). Distinguishing ordinal and disordinal interactions. *Psychological Methods, 17*, 615–622.
- Williams, R. B., Marchuk, D. A., Gadde, K. M., Barefoot, J. C., Grichnik, K., Helms, M. J., et al. (2003). Serotonin-related gene polymorphisms and central nervous system serotonin functions. *Neuropsychopharmacology, 28*, 533–541.
- Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A tool for genome-wide complex trait analysis. *American Journal of Human Genetics, 88*, 76–82.