Journal of Experimental Child Psychology 154 (2017) 78–97

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Children's differential susceptibility to parenting: An experimental test of "for better and for worse"

Meike Slagt [a,*], Judith Semon Dubas [b], Marcel A.G. van Aken [b], Bruce J. Ellis [c], Maja Deković [a]

[a] Department of Clinical Child and Family Studies, Utrecht University, 3508 TC Utrecht, The Netherlands
[b] Department of Developmental Psychology, Utrecht University, 3508 TC Utrecht, The Netherlands
[c] Department of Psychology, University of Utah, Salt Lake City, UT 84112, USA

ARTICLE INFO

Article history:
Received 14 July 2016
Revised 7 October 2016
Accepted 9 October 2016

Keywords:
Diathesis–stress
Differential susceptibility
Experiment
Person-by-environment interactions
Temperament

ABSTRACT

Differential susceptibility theory proposes that a subset of individuals exist who display enhanced susceptibility to both negative (risk-promoting) and positive (development-enhancing) environments. This experiment represents the first attempt to directly test this assumption by exposing children in the experimental group to both negative and positive feedback using puppet role-plays. It thereby serves as an empirical test as well as a methodological primer for testing differential susceptibility. Dutch children ($N = 190$, 45.3% girls) between the ages of 4 and 6 years participated. We examined whether negative and positive feedback would differentially affect changes in positive and negative affect, in prosocial and antisocial intentions and behavior, depending on children's negative emotionality. Results show that on hearing negative feedback, children in the experimental group increased in negative affect and decreased in positive affect more strongly than children in the control group. On hearing positive feedback, children in the experimental group tended to increase in positive affect and decrease in prosocial behavior. However, changes in response to negative or positive feedback did not depend on children's negative emotionality. Moreover, using reliable change scores, we found support for a subset of "vulnerable" children but not for a subset of "susceptible" children. The findings offer suggestions to guide future differential susceptibility experiments.

* Corresponding author. Fax: +31 30 2534718.
E-mail address: m.i.slagt@uu.nl (M. Slagt).

http://dx.doi.org/10.1016/j.jecp.2016.10.004
0022-0965/

Introduction

Differential susceptibility theory suggests that children vary in their general susceptibility to environmental influences, with some being more strongly affected than others by both negative (risk-promoting) and positive (development-enhancing) experiences (Belsky, 1997b, 2005; Belsky, Bakermans-Kranenburg, & van IJzendoorn, 2007; Boyce & Ellis, 2005; Boyce et al., 1995; Ellis, Boyce, Belsky, Bakermans-Kranenburg, & van IJzendoorn, 2011). Thus, the very characteristics that make children disproportionately vulnerable to negative experiences might also make them disproportionately likely to benefit from positive experiences and vice versa (''for better *and* for worse''). Past research has not been able to directly test this key assumption, however, because participants in differential susceptibility studies have not been exposed, experimentally, to both negative and positive environmental conditions. The resulting lacuna presents a challenge to differential susceptibility theory because if the assumption of ''for better and for worse'' was falsified, a central assumption of the theoretical framework would be called into question. Using an experimental research design, the current study was designed to test this assumption by exposing the same individuals to both negative and positive social contexts. The study introduced a new and powerful way to test for differential susceptibility, thereby serving as an empirical test of differential susceptibility as well as a methodological primer.

The differential susceptibility model differs from the traditional diathesis–stress model (Zuckerman, 1999). Whereas the latter emphasizes the disproportionate vulnerability to negative environments of some individuals, the former highlights the disproportionate susceptibility to both the negative effects of harsh environments and the beneficial effects of supportive environments in the same individuals. Differences in susceptibility are hypothesized to have lasting impact on children, with susceptible children experiencing sustained developmental change (for better or for worse) based on the environment they encounter (Ellis et al., 2011). It is important to recognize that our experiment was designed, instead, to test short-term reactions to minor changes in the environment, that is, differential reactivity. Whether reactivity also implies developmental susceptibility, and vice versa, is an open question (Stamps, 2016). However, given the need to manipulate children's environment for better *and* for worse, a focus on reactivity instead of susceptibility was ethically preferable. The strength of experiments like these lies in providing a ''test of principle.''

To investigate which children are more or less susceptible, previous studies have tested genotypic variations, physiological reactivity, and temperament traits as indicators of differences in susceptibility. Following early research on differential susceptibility (Belsky, 1997a, 2005; Belsky, Hsieh, & Crnic, 1998), in this study we focused on temperament traits as markers of susceptibility. Correlational studies suggest that children higher on negative emotionality (defined as the tendency to be easily distressed; Rothbart, Ahadi, Hershey, & Fisher, 2001) are more susceptible to parenting and other environmental influences (for reviews, see Belsky & Pluess, 2009; Ellis et al., 2011; Pluess & Belsky, 2010; Slagt, Dubas, Deković, & van Aken, 2016). Compared with their counterparts lower on negative emotionality, these children showed more behavior problems and lower social and academic adjustment when parenting quality was low and showed fewer behavior problems and better adjustment when parenting quality was high (e.g., Roisman et al., 2012).

Experimental tests of differential susceptibility

Although correlational studies have provided support for differential susceptibility, experimental evidence for differential susceptibility remains limited (Ellis et al., 2011; van IJzendoorn & Bakermans-Kranenburg, 2012). Experimental tests of differential susceptibility studies have several advantages over correlational studies (Bakermans-Kranenburg & van IJzendoorn, 2015; van IJzendoorn & Bakermans-Kranenburg, 2012). First, in correlational studies, children's environment (E) and their score on the susceptibility marker (P) can be correlated; child characteristics can evoke parenting (evocative rPE), children with certain characteristics can seek out certain environments (active rPE), and characteristics shared by parents and their children can underlie associations between parenting and child development (passive rPE) (Rutter, 2006). Furthermore, environmental conditions can shape susceptibility factors (Boyce & Ellis, 2005; Del Giudice, Ellis, & Shirtcliff, 2011). In experimental studies, the environment is manipulated in standard ways and randomization of participants to conditions breaks down the possibility of rPE. Second, experimental studies prevent the oftentimes highly skewed distributions of environmental measures by manipulating the environment and having similar numbers of participants in each environmental condition. Third, manipulation of the environment creates standardized, clear, and targeted measures of environmental stimuli. Such measures decrease ''noise'' in the assessment of environmental stimuli and increase the power to detect interactions if present. In sum, experimental examination of differential susceptibility affords the most solid basis for causal inference.

To date, a handful of experimental studies—all macrotrials studying developmental susceptibility instead of reactivity—have shown that children high on negative emotionality might be more susceptible to broad parenting interventions. Two of these studies found that infants high on negative emotionality profited more from experimentally induced increases in supportive parenting than less negative infants, as evidenced by their increased attachment security (Cassidy, Woodhouse, Sherman, Stupica, & Lejuez, 2011; Klein Velderman, Bakermans-Kranenburg, Juffer, & Van IJzendoorn, 2006). Others showed that children high on negative emotionality profited disproportionately from experimentally induced increases in supportive parenting, as evidenced by their decreased internalizing and externalizing problems and increased cognitive functioning (Blair, 2002; Scott & O'Connor, 2012). Finally, a quasi-experiment showed that girls higher on sensory-processing sensitivity, compared with those low on this trait, benefitted more from an intervention aimed at reducing depression (Pluess & Boniwell, 2015).

The next step in differential susceptibility experiments

The studies discussed above testify to the progress being made in experimentally testing differential susceptibility. Nevertheless, experimental studies to date examined only positive changes in parenting and, for clear ethical reasons, did not look at the experimentally induced effects of negative changes in parenting. Two solutions have been suggested to test this (Ellis et al., 2011). First, animal models could be used to conduct experiments involving both positive and negative changes in the environment of the same subjects (see, e.g., Suomi, 1997). Second, experimentally induced changes in the microenvironment of the same individuals (i.e., minor stimuli in an individual's immediate surroundings), both for better and for worse, could be used. These ''nanotrials'' examine the immediate neural or behavioral responses to a small range of positive and negative stimuli (van IJzendoorn & Bakermans-Kranenburg, 2015, p. 153) and, strictly speaking, test for differential reactivity instead of susceptibility. For example, one study used an attention bias modification procedure to train one group of children to pay attention to negative pictures while training another group to pay attention to positive pictures (Fox, Zougkou, Ridgewell, & Garner, 2011). Children with the low-expression form of the 5-HTTLPR gene developed stronger attention bias toward both negative and positive affective pictures than children with the high-expression form of the gene. In another study, children were questioned by either a supportive or non-supportive interviewer (Quas, Bauer, & Boyce, 2004). Autonomic reactivity was associated with increased memory accuracy among children questioned in a supportive manner, but it was associated with decreased accuracy among children questioned in a non-supportive manner.

Following these studies, the current experiment used a ''nanotrial'' approach. To provide a stringent test of ''for better and for worse,'' we manipulated children's environment in two directions, simulating both positive and negative feedback from parents to their children using puppet role-play scenarios (Kamins & Dweck, 1999). Such experimentally induced changes in children's microenvironment have proven to be effective in changing children's emotions and behaviors at posttest (Kamins & Dweck, 1999; Zentall & Morris, 2010).

Apart from mostly focusing on exposure to positive environments, previous intervention studies used between-groups designs. That is, children in the experimental group were exposed to one type of environment only. It could be, however, that some children are more susceptible to positive parenting and others are more susceptible to negative parenting, which would oppose the idea that certain individuals are susceptible "for better *and* for worse." Within-person designs exposing the same children to both negative and positive changes in their microenvironments are needed to demonstrate that those who profit most from a positive change in the environment also react most to a negative change. To achieve this, we used a novel approach to testing differential susceptibility, combining a between-groups design (i.e., experimental group and control group) with a within-person design. Children in the experimental group received two manipulations, namely positive feedback and negative feedback, whereas children in the control group received no feedback.

In addition, previous intervention studies did not always assess child outcomes ranging from negative to positive. Predicting child outcomes with a restricted range, such as ranging from the presence to absence of behavior problems, enables a test of only half of the "for better and for worse" interaction (Belsky et al., 2007; Roisman et al., 2012). It does not reveal whether highly susceptible children, for example, also show the highest levels of social competence or positive affect under supportive circumstances. Therefore, we assessed child outcomes ranging from negative (negative affect and antisocial intentions and behaviors) to positive (positive affect and prosocial intentions and behaviors).

Furthermore, in none of the aforementioned experiments was random assignment to intervention and control groups stratified according to the susceptibility marker. In our study, children's temperament was measured during a screening phase preceding the experiment. Children with low or high scores on either negative emotionality or surgency[1] were selected for the experiment because oversampling extreme scores on the susceptibility marker (at both the high and low ends) increases power to detect interaction effects (Preacher, Rucker, MacCallum, & Nicewander, 2005). In this way, we ended up with four temperament groups: children low on both dimensions, children high on negative emotionality and low on surgency, children low on negative emotionality and high on surgency, and children high on both dimensions. The selected children were then randomly assigned to an experimental or control group, stratified according to their temperament group.

Hypotheses

In this article, we introduce a new approach to testing differential susceptibility by combining a between-groups experiment and a within-participant design. Focusing on both negative and positive microenvironments and outcomes, the current study provides one of the first *experimental* tests of a true crossover interaction. If predictions of differential susceptibility theory are correct and apply to situational reactivity, children scoring higher on negative emotionality should show more pronounced changes in affect, intentions, and behavior when receiving positive feedback as well as negative feedback. This offers a pioneering test of the hypothesis that the same children who respond most strongly to negative changes in their environment also respond most strongly to positive changes.

We hypothesized that children in the experimental group would increase in positive affect and prosocial intentions and behavior after the positive manipulation. Likewise, we hypothesized that children in the experimental group would increase in negative affect and antisocial intentions and behavior after the negative manipulation. We expected these changes to be stronger in the experimental group compared with the control group. Importantly, we hypothesized changes in outcome measures on receiving feedback to be moderated by temperament group. If the predictions of differential susceptibility are correct, among children high on negative emotionality, receiving positive as well as negative feedback should produce more pronounced changes in outcomes compared with their counterparts low on negative emotionality.

[1] Initial selection of the sample took place based on both negative emotionality and surgency and is discussed as such in the Introduction and Method. However, because surgency as a susceptibility marker is not supported based on more recent literature (Slagt et al., 2016), we did not test for moderation by surgency.

Method

Participants

Information about the study was distributed to parents of children in Grades 1 and 2 at regular elementary schools in the province of Utrecht, The Netherlands. Parents could voluntarily sign their children up for the study at a website, where they gave active informed consent, filled out their contact information, and completed a short screening questionnaire inquiring about children's negative emotionality and surgency. In this way, 280 children signed up for the study. The experiment was part of this larger longitudinal study.

For the experiment, we selected a subsample of 192 children on their low/high scores on negative emotionality and surgency (see "Measures" section below for a description of these scales), using an extreme group approach (Preacher et al., 2005). This way, a priori power to detect interactions between temperament group and change from pretest to posttest within the experimental group ($n = 96$; $p < .05$; comparison among four temperament groups; small to medium effect of $f = .17$) would be sufficient (>80%; Faul, Erdfelder, Lang, & Buchner, 2007). The children were selected by first splitting negative emotionality scores and surgency scores into three groups: the lowest 42%, the middle 16%, and the highest 42%. Next, we made a cross table based on these two categorized temperament measures, which yielded four approximately equal-sized groups in the upper right, upper left, lower right, and lower left cells of the cross table that together contained 192 children. Children in these cells—low on both negative emotionality and surgency (low–low), high on both negative emotionality and surgency (high–high), high on negative emotionality and low on surgency (high N), and low on negative emotionality and high on surgency (high S)—were selected for the experiment (see Online Supplement 1 in Supplementary material for comparisons among the four groups). Two of the families declined participation in the experiment and wanted to take part only in the questionnaire part of the study, leaving a final sample of 190 children for the experiment.

Because only negative emotionality was tested as a susceptibility marker, the division into four temperament groups described above was used only during sample selection and not during further analyses. Instead, for analyses, we combined the two temperament groups high on negative emotionality into a single high negative emotionality group and combined the two temperament groups low on negative emotionality into a single low negative emotionality group (see Online Supplement 1 for comparisons between the two groups). Because of the bimodal distribution of negative emotionality among the children participating in the experiment, we decided to create this dichotomous version of negative emotionality. This new grouping variable allowed us to examine the moderating effect of negative emotionality regardless of children's level of surgency.

Participating children were boys (54.7%) and girls (45.3%) between the ages of 3.77 and 6.14 years at the start of the study ($M = 4.76$, $SD = 0.57$). Most of the children (96.3%) were born in The Netherlands. Parents who completed the screening questionnaire were mothers (90%) and fathers (10%) between the ages of 25.29 and 51.92 years ($M = 37.92$, $SD = 4.33$). Most parents were born in The Netherlands (93.7%) and were married or cohabiting (94.7%). Parents were highly educated, with 5.3% having no high school diploma or having finished lower vocational education, 16.8% having finished intermediate vocational education, and 77.9% having finished higher vocational education or university. Gross annual household income was less than the national mode (€35,000) for 4.7% of families, between 1 and 1.5 times the national mode for 12.1% of families, 1.5 to 2 times the national mode for 30.5% of families, and more than 2 times the national mode for 38.4% of families (14.2% of families did not report their income). Children participating in the experiment ($n = 190$) did not differ from children not participating in the experiment ($n = 90$) on child gender, parent gender, child age, parent age, country of birth child, or country of birth parent or on marital status, education level, or income of parents, as indicated by chi-square and independent-samples *t*-tests.

Procedure

Parents filled out a screening questionnaire, on which children were randomly assigned to either the experimental condition or the control condition, stratified by temperament group (see Fig. 1 for a graphical display of the research design).

The home visits during which the experiment took place occurred approximately 4 months later. Children were visited at home twice by an experimenter (either the first author or a trained research assistant). The experimenter was blind to children's temperament group status. During the first visit, the study was explained, after which observations of parent–child interactions took place (not relevant for this paper). Next, children participated in the first half of the experiment, consisting of a pretest, an experimental manipulation (for children in the experimental group), and a posttest. During the second visit, the second half of the experiment took place, again consisting of a pretest, an experimental manipulation (for children in the experimental group), and a posttest. Home visits were scheduled approximately 2 weeks apart to prevent carryover effects between visits.

The manipulation consisted of role-play scenarios using puppets (Cimpian, Arce, Markman, & Dweck, 2007; Kamins & Dweck, 1999). Children chose a puppet to represent themselves and a puppet to represent their parent (puppets of both sexes were available). Children then used the child puppet to act out their part in the scenarios, whereas the experimenter handled the parent puppet and narrated the scenarios. After a short warm-up period during which children got the chance to practice role-playing with the puppets, the manipulation started.

The manipulation consisted of two scenarios per visit acted out using the puppets. In each scenario, the parent puppet asked the child puppet to draw a different object (an apple, a tree, a cat, or a bus). Small pencils were used as pretend crayons; no actual drawing or pictures were involved. In the experimental group, during one home visit, the child puppet successfully completed the requested drawing and received positive feedback from the parent puppet ("That looks like an *apple/tree.* You did a good job drawing. I'm very proud of you!" in each of the two scenarios). During another home visit, the child puppet made a mistake (omitting ears on a cat and wheels on a bus) and received negative feedback from the parent puppet ("That doesn't look like a *cat/bus;* it doesn't have any *ears/wheels.* You did a bad job drawing, I'm very disappointed in you!" in each of the two scenarios). The order of manipulations (positive feedback during the first visit and negative feedback during the second visit or vice versa) was counterbalanced. For debriefing, the mistake scenarios were completed successfully at the end of the negative feedback visit.

In the control condition, children participated in a similar procedure. They were also visited at home twice and role-played similar scenarios using puppets. The only crucial difference was that the child puppet received no feedback from the parent puppet in these scenarios. In the remainder of the article, we refer to the visit during which children role-played apple and tree scenarios as the "positive feedback visit" and refer to the visit during which children role-played cat and bus scenarios as the "negative feedback visit."

Pretend stories were used so that no judgments or criticism was given directly to the children, but the scenarios were vivid, so that all children would feel as though they were performing the task and receiving the feedback. The study procedure was approved by the faculty's ethics committee. Children received a bubble blower at the end of the second visit to thank them for their participation.

Measures

Dependent measures were assessed at pretest and posttest of the two different visits, that is, at four time points.

Positive affect

Positive affect was assessed using a single item asking children "At this moment, do you feel happy or not?" If children answered "happy," they were then asked "How happy? Do you feel a little happy or very happy?" The question was presented verbally as well as supported visually by emoticons representing the emotion (Russell, 1990). Answers were scored as 1 = no, 2 = yes, a little, and 3 = yes, very.

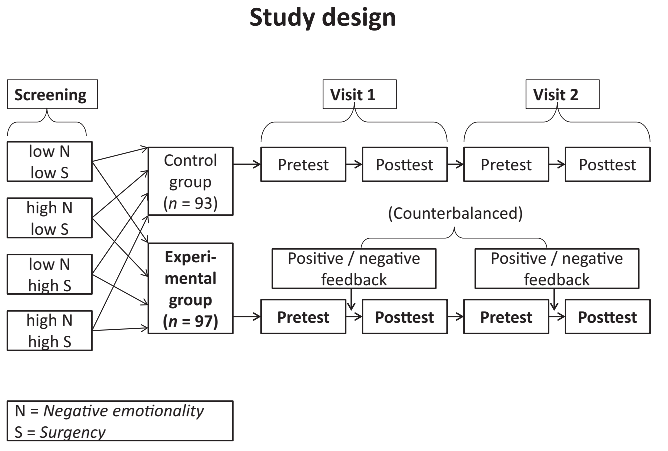

Fig. 1. Study design.

Prosocial intentions

Prosocial intentions were measured using vignettes adopted from Rotenberg and Eisenberg (1997). Two vignettes were used, each involving the child and a same-sex peer. Line drawings were used to depict the stories. The stories described situations that would elicit prosocial intentions. One vignette involved Santa Claus visiting and bringing presents, but one child gets a present that is broken. The other vignette involved a child playing with his or her toys, and then the child accidently spills a drink over the toys. The complete wording of the Santa Claus vignette was as follows: "Santa Claus comes to visit and brings toys. This girl/boy gets a present that is broken. See, her doll's arm is broken [for a girl]. See, his train is broken [for a boy]. Imagine you're there too. Look, you see the girl/boy with the broken doll/train." Children were then asked "What would you do?" Children's responses were coded verbatim. At each visit, one vignette was administered during pretest and the other vignette during posttest (counterbalanced across samples). Preliminary analyses showed that the order in which these two versions were administered did not affect results.

Two independent coders who had not conducted home visits and who were blind to temperament and experimental group status coded the verbatim answers on a 3-point scale (adapted from Vaish, Carpenter, & Tomasello, 2009): 2 = behaviors or verbal expressions meant to help or comfort the other child in the vignette (e.g., share own toys, ask an adult to mend the toy, comfort the other child, help to clean up), 1 = showing distress or attending to the situation (e.g., feeling sorry for the other child, wanting to help but not knowing how), and 0 = absence of prosocial intentions (e.g., shows no involvement or interest in the child in the vignette). To assess interrater reliability, 25% of the answers were coded by both coders. Interrater reliability was high, with Cohen's kappa = .81 and interrater agreement = 93%.

Prosocial behavior

Prosocial behavior was measured by a sharing task (Iannotti, 1985). Each child was given a choice of either M&Ms (candy) or raisins. Seven of the preferred items were given to the children while the experimenter indicated that they "could leave some for another child I'll visit later." Several options were presented (eating all, giving some, and giving all of the candy or raisins). The children were then left alone to put the candy in an envelope. The number of candies shared, between 0 and 7, was the prosocial behavior score.

Negative affect

Negative affect was assessed using three items asking children whether they felt angry, sad, or scared. All three items were formulated the same way (e.g., "At this moment, do you feel *angry* or not?" If children answered "angry," they were then asked "How angry? Do you feel a little angry or very angry?"). The questions were presented verbally as well as visually by emoticons representing the relevant emotions (Russell, 1990). Answers were scored as 1 = no, 2 = yes, a little, and 3 = yes, very. Ordinal factor analysis in Mplus supports a single factor structure of the three negative affect items at each of the four time points, with strong measurement invariance across time, $\chi^2(69) = 76.30$, $p = .26$, comparative fit index (CFI) = .997, Tucker–Lewis index (TLI) = .997, root mean square error of approximation (RMSEA = .024). Factor loadings varied between .93 and .99, and the negative affect factor explained between 87% and 98% of the variance in the items at each time point. Therefore, we averaged the three negative affect items into a single negative affect score at each time point. Internal consistency was high, with ordinal coefficient alpha (Zumbo, Gadermann, & Zeisser, 2007) at .96 or .97.

Antisocial intentions

Antisocial intentions were measured using vignettes adopted from Murphy and Eisenberg (1997). Two vignettes were used, each involving the child and a same-sex peer. Line drawings were used to depict the stories. The stories described situations that would elicit antisocial intentions. One vignette involved the child wanting to play with a cool puzzle, but then another child comes and takes the puzzle. The other vignette involved the child building a sand castle, but then another child runs past and steps on it, damaging the castle. The complete wording of the puzzle vignette was as follows: "You're at school. In the corner of the classroom, you spot a very cool puzzle. You really want to play with that puzzle. But then another boy/girl comes. The boy/girl takes the puzzle that you wanted and starts playing with it." Children were then asked "What would you do?" Children's responses were coded verbatim by the experimenter. If children did not respond, or responded with "I don't know," the experimenter used prompts such as "Would you, for instance, say something angry or would you not? What then?" At each visit, one vignette was administered during pretest and the other vignette during posttest (counterbalanced). Preliminary analyses show that the order of administration did not affect results.

Two independent coders coded the verbatim answers on a 3-point scale: 2 = behaviors or verbal expressions that likely inflict physical or emotional pain onto the other child in the vignette (e.g., destroying something, hitting, aggressive language), 1 = behaviors or verbal expressions that likely inflict light emotional pain onto the other child (e.g., becoming angry, trying to take the puzzle from the other child), and 0 = absence of antisocial intentions (e.g., find something else to play with, rebuild the sand castle). To assess interrater reliability, 25% of the answers were coded by both coders. Interrater reliability was high, with Cohen's kappa = .87 and interrater agreement = 96%.

Antisocial behavior

Antisocial behavior was measured by an adaptation of the hot sauce paradigm and the sharing task[2] (Iannotti, 1985; Lieberman, Solomon, Greenberg, & McGregor, 1999). Each child was presented with 14 stickers, 7 of which were intact and 7 of which were torn. The experimenter explicitly pointed out to the child that some of the stickers were intact and others were torn and also mentioned that most children did not like torn stickers but that most children did like intact stickers. Next, the experimenter indicated that the child "could choose 7 stickers to give to another child I'll visit later." The child was told that he or she did not get to keep the remaining 7 stickers. The child was then left alone to put the stickers in an envelope. The number of torn stickers given to another child, between 0 and 7, was the antisocial behavior score. Because children might be prone to testing effects if they need to do the same task twice during the same visit, we employed two versions of this task: one with stickers and one with intact and torn empty balloons. One was administered during pretest and one during posttest (counterbalanced). Preliminary analyses show that administration order did not affect results.

[2] We based this task on existing tasks and adapted it to the age group used in our study. The new task used here correlates in the expected direction with other measures used in the study, specifically with negative affect and antisocial intentions (Table 2).

Temperament

Children's temperament was assessed using the Children's Behavior Questionnaire–Short Form (Putnam & Rothbart, 2006; Rothbart et al., 2001). We measured Anger/Frustration ("has temper tantrums when s/he doesn't get what s/he wants"), Soothability ("is very difficult to soothe when s/he has become upset"), Fear ("is afraid of burglars or the boogie man"), Sadness ("cries sadly when a favorite toy gets lost or broken"), Impulsivity ("usually rushes into an activity without thinking about it"), Activity Level ("seems always in a big hurry to get from one place to another"), Approach ("becomes very excited while planning for trips"), and High Intensity Pleasure ("likes going down high slides or other adventurous activities"). Items could be answered on a 7-point scale ranging from 1 (*extremely untrue of your child*) to 7 (*extremely true of your child*), with higher scores indicating higher negative emotionality or surgency. A *not applicable* response option was also available. Scale scores were created by averaging applicable item scores. Following previous research (Rothbart et al., 2001), the Anger/Frustration, reversed Soothability, Fear, and Sadness scales were then averaged into a negative emotionality score ($\alpha = .82$). The Impulsivity, Activity Level, Approach, and High Intensity Pleasure scales were averaged into a Surgency score ($\alpha = .90$).

Analyses

Analyses were run in Mplus 7.2 (Muthén & Muthén, 1998–2012). To answer our research questions, we needed to model within-person changes as well as predict between-persons variation in change between specific time points of interest, that is, from pretest to posttest around the negative feedback manipulation and from pretest to posttest around the positive feedback manipulation. To this end, we estimated latent change score models (McArdle, 2009).

For each of the dependent variables (negative affect, positive affect, prosocial intentions, antisocial intentions, prosocial behavior, and antisocial behavior) we first estimated an unconditional latent change score model (see Fig. 2). In these models, changes from pretest to posttest at the two visits were represented by two latent change scores. The unit of these latent change scores is standardized unit change relative to the variability observed at the pretests.[3] After having modeled change in the dependent measures for the sample as a whole, we proceeded to predict variation in latent change scores by experimental condition and temperament group. That is, we first checked whether random assignment was successful by regressing latent pretest scores on experimental condition (0 = control and 1 = experimental). Then, to check whether the manipulation was successful, we regressed latent change scores on experimental condition. Next, we added main effects of negative emotionality (0 = low and 1 = high). Finally, to test whether experimentally induced changes in the dependent variables would be larger among children high on negative emotionality, we added interactions between negative emotionality and experimental condition.

To estimate the latent change score models, we used a Bayesian estimator, which has a better small-sample performance compared with maximum likelihood estimators and is well equipped to handle non-normally distributed data (Asparouhov & Muthén, 2010; Lee & Song, 2004; see van de Schoot et al., 2014, for an applied example). More information about the Bayesian estimation procedure, the interpretation of the statistics it yields, and the settings we used can be found in Online Supplement 2 of the Supplementary material.

We reestimated the models described in this section using ordinal regression analyses with $N = 1000$ bootstrap resamples. This yielded similar results as those described in the Results section below.[4] Finally, to control for inflation of Type I error rates, we applied a false discovery rate (FDR) procedure (Benjamini & Hochberg, 1995).

[3] Observed posttest scores were converted to z-scores using the means and standard deviations from the corresponding pretest scores for proper scaling.

[4] Results are available from the first author on request.

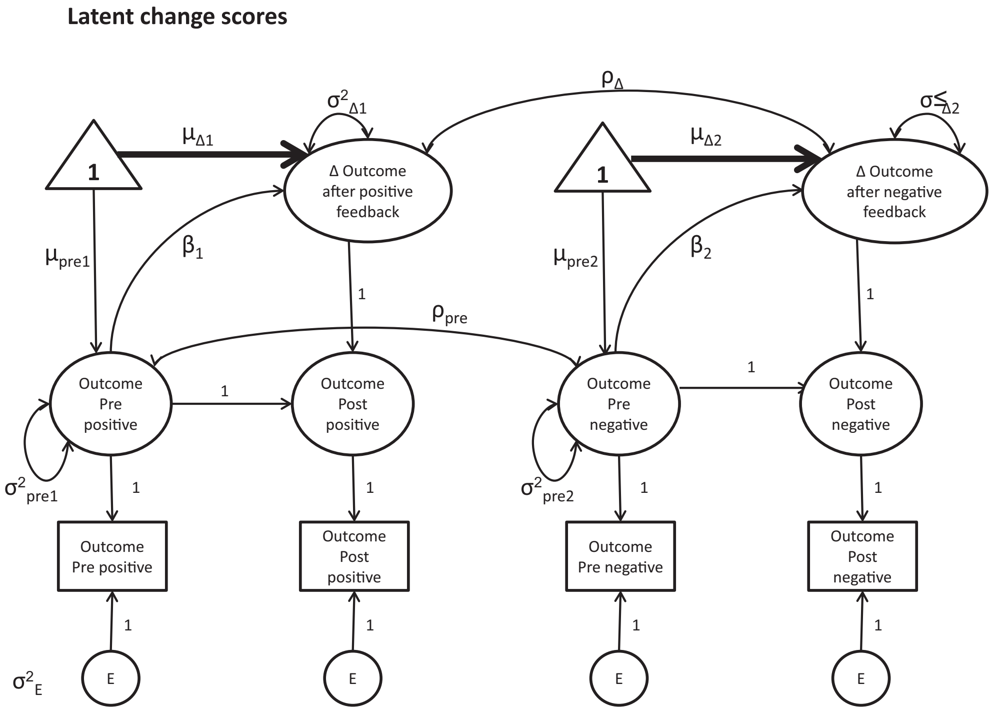

Fig. 2. Unconditional latent change score model. The model contains the following parameters: the change from pretest to posttest at each visit, that is, the latent change scores $\mu_{\Delta 1}$ and $\mu_{\Delta 2}$ (these latent change scores are the parameters of interest that are predicted by experimental condition and temperament group in later models); the latent means at the pretest of each visit (μ_{pre1} and μ_{pre2}); the variances of the latent pretest means (σ^2_{pre1} and σ^2_{pre2}) and latent change scores ($\sigma^2_{\Delta 1}$ and $\sigma^2{}_{\Delta 2}$); the latent posttest score being a one-to-one function of both the latent pretest score and the latent change score (paths are fixed to 1 to achieve this); the regression of the latent change scores on their corresponding pretest scores (β_1 and β_2) (i.e., do children's pretest scores predict how they respond to the experimental manipulation?); the correlation between the two pretest scores (ρ_{pre}) (i.e., are pretest scores at the two visits correlated?); the correlation between the two latent change scores (ρ_{Δ}) (i.e., is the way children respond to positive feedback related to how they respond to negative feedback?); errors (σ^2_{E}) that are assumed to have a mean of 0 and to have variances that are equal across time (i.e., homoscedasticity).

Results

Descriptive results

Descriptive statistics of the dependent variables are displayed in Table 1 for both the experimental and control groups. All measures showed moderate to strong rank-order stability within and across the two visits (see Online Supplement 3 in Supplementary material). Correlations of dependent measures at pretest are presented in Table 2. The more negative affect children showed at pretest, the less positive affect they reported. In addition, children who reported more negative affect tended to display more antisocial behavior and antisocial intentions and less prosocial behavior. Children's positive affect was not related to their antisocial and prosocial behaviors and intentions. Children's antisocial behavior and antisocial intentions were weakly associated with each other, whereas their prosocial behavior and prosocial intentions were unassociated. Finally, children who showed more antisocial intentions tended to show less prosocial behavior.

Unconditional latent change score models

To examine whether children varied in how much they changed from pretest to posttest, we first estimated unconditional latent change score models (see Fig. 2). Each of the latent change scores

Table 1
Means and standard deviations for dependent variables split by experimental condition.

Dependent variable	Control group				Experimental group			
	Pretest positive feedback *M* (*SD*)	Posttest positive feedback *M* (*SD*)	Pretest negative feedback *M* (*SD*)	Posttest negative feedback *M* (*SD*)	Pretest positive feedback *M* (*SD*)	Posttest positive feedback *M* (*SD*)	Pretest negative feedback *M* (*SD*)	Posttest negative feedback *M* (*SD*)
Positive affect	2.65 (0.62)	2.51 (0.72)	2.56 (0.72)	2.45 (0.78)	2.44 (0.72)	2.65 (0.68)	2.49 (0.73)	2.04 (0.91)
Prosocial intentions	1.69 (0.69)	1.74 (0.64)	1.65 (0.73)	1.71 (0.66)	1.61 (0.74)	1.61 (0.74)	1.59 (0.75)	1.65 (0.73)
Prosocial behavior	1.58 (1.92)	1.66 (2.08)	1.61 (1.98)	1.51 (2.00)	1.57 (1.90)	1.23 (1.87)	1.86 (2.25)	1.38 (2.09)
Negative affect	1.08 (0.34)	1.08 (0.27)	1.10 (0.35)	1.11 (0.42)	1.10 (0.37)	1.12 (0.40)	1.09 (0.32)	1.30 (0.55)
Antisocial intentions	0.39 (0.66)	0.37 (0.62)	0.38 (0.62)	0.48 (0.75)	0.48 (0.71)	0.49 (0.74)	0.39 (0.67)	0.51 (0.75)
Antisocial behavior	0.89 (1.79)	1.01 (1.91)	1.03 (1.76)	1.18 (2.14)	1.32 (2.23)	1.61 (2.45)	1.13 (1.89)	1.14 (1.97)

Note. The possible range of scores for positive and negative affect was 1 to 3, for prosocial and antisocial intentions was 0 to 2, and for prosocial and antisocial behavior was 0 to 7.

Table 2
Correlations between measures at pretest.

	1	2	3	4	5	6	7	8
1. Positive affect	–	.12	−.06	−.31***	.04	−.07	.03	.03
2. Prosocial intentions	−.02	–	−.02	−.06	.03	−.05	.02	.00
3. Prosocial behavior	.01	−.04	–	−.03	.03	−.07	.01	−.13
4. Negative affect	−.16*	−.20**	.00	–	.28***	.25***	−.11	.03
5. Antisocial intentions	.04	−.00	−.20**	.24**	–	.20**	−.01	.09
6. Antisocial behavior	−.06	.01	.02	.13	.12	–	−.04	.02
7. Negative emotionality	.02	.02	.01	−.09	−.09	.03	–	.21**
8. Surgency	.05	.01	−.00	.05	.11	.06	.21**	–

Note. All correlations are Spearman rank correlations. Correlations below the diagonal refer to the pretest before positive feedback; correlations above the diagonal refer to the pretest before negative feedback.
* $p < .05$.
** $p < .01$.
*** $p < .001$.

showed significant variance, indicating that variation across children exists in how much they change from pretest to posttest (see Online Supplement 4 for fit statistics and parameter estimates of unconditional latent change score models). In subsequent models, we tried to predict this variation by condition, temperament group, and their interaction.

Random assignment and counterbalancing check

To check whether random assignment was successful, we tested whether experimental condition predicted latent pretest means. No associations emerged between condition and children's latent pretest scores (i.e., all of the 95% posterior probability intervals [PPIs] contained zero), indicating that random assignment was successful. Next, we examined whether counterbalancing of manipulation order affected the results. To this end, we predicted latent change scores by condition, manipulation order (positive feedback–negative feedback or vice versa), and their interaction. These analyses showed that the effect of experimental condition on latent change scores did not depend on manipulation order.

Manipulation check

To examine whether the manipulation had an effect on children, we predicted changes in dependent measures by experimental condition. This was done by adding regression paths from experimental condition to the latent change scores in each of the six unconditional models. The fit of the models predicting positive and negative affect improved by adding these paths (positive affect: ΔDIC (deviance information criterion) = −9.69, ΔBIC (Bayes information criterion) = −5.27, Bayes factor = 13.94; negative affect: ΔDIC = −5.85, ΔBIC = −3.12, Bayes factor = 4.76). In contrast, models predicting the other four outcomes did not improve by adding main effects of condition on latent change scores (prosocial intentions: ΔDIC = 2.99, ΔBIC = 9.57, Bayes factor = 0.01; prosocial behavior: ΔDIC = 1.06, ΔBIC = 3.56, Bayes factor = 0.17; antisocial intentions: ΔDIC = 3.25, ΔBIC = 8.16, Bayes factor = 0.02; antisocial behavior: ΔDIC = 0.41, ΔBIC = 7.73, Bayes factor = 0.02). Parameter estimates of these models are presented in Table 3. These show that children in the control group did not change on any of the dependent measures from pretest to posttest. Children in the experimental group, however, did change. Specifically, on hearing negative feedback, children in the experimental group increased in negative affect (0.61*SD* compared with the pretest) and decreased in positive affect (−0.61*SD* compared with the pretest), as indicated by their significant latent change scores (see Table 3). Moreover, these changes were stronger in the experimental group than in the control group, as indicated by significant associations between condition and latent change scores. Whether children belonged to the experimental group or the control group explained 9.4% of the variance in changes in negative affect and 12.9% of the variance in changes in positive affect during the negative feedback visit.

Table 3
Main effects of experimental condition on latent change scores.

Dependent variable	Latent change in control group		Latent change in experimental group		Experimental condition predicting latent change (β)	
	Estimate β (*SD*)	95% PPI	Estimate β (*SD*)	95% PPI	Estimate β (*SD*)	95% PPI
Positive affect; visit positive feedback	−0.09 (0.10)	[−0.29, 0.11]	0.20 (0.10)	[0.00, 0.41][*,a]	0.29 (0.14)	[0.01, 0.57][*,a]
Positive affect; visit negative feedback	−0.17 (0.11)	[−0.39, 0.06]	−0.61 (0.11)	[−0.83, −0.39][*]	−0.45 (0.15)	[−0.75, −0.14][*]
Prosocial intentions; visit positive feedback	0.10 (0.08)	[−0.07, 0.26]	−0.02 (0.08)	[−0.19, 0.14]	−0.12 (0.12)	[−0.35, 0.10]
Prosocial intentions; visit negative feedback	0.09 (0.09)	[−0.08, 0.27]	0.06 (0.09)	[−0.11, 0.23]	−0.03 (0.12)	[−0.27, 0.20]
Prosocial behavior; visit positive feedback	0.05 (0.08)	[−0.11, 0.22]	−0.19 (0.08)	[−0.35, −0.03][*,a]	−0.24 (0.11)	[−0.47, −0.02][*,a]
Prosocial behavior; visit negative feedback	−0.09 (0.08)	[−0.25, 0.07]	−0.20 (0.08)	[−0.36, −0.05][*]	−0.12 (0.11)	[−0.33, 0.10]
Negative affect; visit positive feedback	−0.05 (0.09)	[−0.22, 0.13]	0.09 (0.09)	[−0.08, 0.26]	0.14 (0.12)	[−0.10, 0.38]
Negative affect; visit negative feedback	0.04 (0.13)	[−0.21, 0.29]	0.61 (0.12)	[0.37, 0.85][*]	0.57 (0.17)	[0.23, 0.91][*]
Antisocial intentions; visit positive feedback	−0.07 (0.09)	[−0.25, 0.11]	0.05 (0.09)	[−0.12, 0.23]	0.13 (0.13)	[−0.13, 0.38]
Antisocial intentions; visit negative feedback	0.19 (0.11)	[−0.03, 0.41]	0.15 (0.11)	[−0.07, 0.23]	−0.04 (0.14)	[−0.32, 0.24]
Antisocial behavior; visit positive feedback	0.03 (0.07)	[−0.11, 0.18]	0.16 (0.07)	[0.02, 0.31][*,a]	0.13 (0.10)	[−0.07, 0.33]
Antisocial behavior; visit negative feedback	0.09 (0.08)	[−0.06, 0.25]	−0.01 (0.08)	[−0.16, 0.15]	−0.09 (0.11)	[−0.31, 0.12]

[a] According to the FDR procedure, this latent change/path is not significant.
[*] Parameter estimate is significantly different from zero; that is, the 95% PPI does not contain zero.

Table 4
Interaction effects of Experimental Condition * Negative Emotionality Group on latent change scores.

Dependent variable	Visit positive feedback		Visit negative feedback	
	Estimate β (*SD*)	95% PPI	Estimate β (*SD*)	95% PPI
Positive affect	0.03 (0.29)	[−0.53, 0.60]	−0.14 (0.31)	[−0.75, 0.47]
Prosocial intentions	0.03 (0.24)	[−0.43, 0.49]	0.08 (0.24)	[−0.40, 0.56]
Prosocial behavior	0.04 (0.23)	[−0.41, 0.50]	0.02 (0.22)	[−0.42, 0.44]
Negative affect	−0.12 (0.24)	[−0.59, 0.36]	−0.43 (0.35)	[−1.11, 0.26]
Antisocial intentions	−0.72 (0.25)	[−1.21, −0.23][*,a]	−0.47 (0.28)	[−1.02, 0.08]
Antisocial behavior	−0.01 (0.20)	[−0.40, 0.39]	0.18 (0.22)	[−0.25, 0.61]

[a] According to the FDR procedure, this path is not significant.
[*] Parameter estimate is significantly different from zero; that is, the 95% PPI does not contain zero.

On hearing positive feedback, children in the experimental group increased in positive affect and antisocial behavior and decreased in prosocial behavior, but these changes were no longer significant after controlling for multiple testing.

Testing differential susceptibility

To test differential susceptibility theory, we analyzed whether experimentally induced changes in outcome variables would be larger among children high on negative emotionality.

Before adding interactions between negative emotionality and experimental condition, we first added main effects of negative emotionality to each model. Most of the models did not improve by adding main effects of negative emotionality on latent change scores (positive affect: ΔDIC = 3.06, ΔBIC = 8.75, Bayes factor < 0.01; prosocial intentions: ΔDIC = 1.89, ΔBIC = 8.59, Bayes factor = 0.01; prosocial behavior: ΔDIC = 1.22, ΔBIC = 7.68, Bayes factor = 0.02; antisocial intentions: ΔDIC = 3.77, ΔBIC = 7.74, Bayes factor = 0.02; antisocial behavior: ΔDIC = 3.13, ΔBIC = 9.71, Bayes factor < 0.01). Only the fit of the model predicting negative affect improved slightly by adding these paths (negative affect: ΔDIC = −8.83, ΔBIC = −1.81, Bayes factor = 2.48). However, after controlling for Type I error, no associations emerged between negative emotionality and any of the latent change scores. That is, there were no main effects of children's negative emotionality on changes in any of the dependent variables.

Next, we added interactions between negative emotionality and experimental condition (see Table 4). After controlling for multiple testing, we found no interaction effects, indicating that experimentally induced changes in dependent measures were similar across children low and high on negative emotionality. Thus, the extent to which children changed on dependent measures in response to positive and negative feedback did not depend on their negative emotionality.

Exploratory follow-up analyses

We did not find support for differential susceptibility theory in our main analyses, testing moderation by negative emotionality. To test the robustness of our results, we wanted to focus on within-person changes, regardless of children's negative emotionality. Potentially, some children change in response to both negative and positive feedback (i.e., they are susceptible "for better and for worse"), but they do not score high on traditional susceptibility markers (e.g., negative emotionality, stress reactivity, a certain genotype). To this end, we calculated the reliable change index (Christensen & Mendoza, 1986) for children in the experimental group. Reliable change refers to change over time (e.g., from pretest to posttest) that is greater than would be expected from random variation alone. It is calculated by dividing the difference between the pretest and posttest scores by the standard error of the difference between the two scores. If the reliable change index is greater than |1.96|, the difference is said to be reliable.

Because we had six dependent measures and responses to positive and negative feedback could potentially be expressed in different outcomes (e.g., increases in positive affect in response to positive

Table 5
Reliable change in response to the feedback manipulations within the experimental group ($n = 97$).

		Reliable change in response to negative feedback		
		Reliable change in unexpected direction[a]	No reliable change	Reliable change in expected direction[b]
Reliable change in response to positive feedback	Reliable change in unexpected direction[b]	0.00%	3.09%	0.00%
	No reliable change	2.06%	80.41%	10.31%
	Reliable change in expected direction[a]	0.00%	4.12%	0.00%

[a] Increases in positive affect, prosocial intentions, and/or prosocial behavior and/or decreases in negative affect, antisocial intentions, and/or antisocial behavior.
[b] Increases in negative affect, antisocial intentions, and/or antisocial behavior and/or decreases in positive affect, prosocial intentions, and/or prosocial behavior.

feedback, increases in antisocial behavior in response to negative feedback), we focused on change in an aggregate score of the six dependent measures. That is, two observed change scores were first calculated: One reflected observed change in response to positive feedback (with positive scores denoting increases in positive affect, prosocial intentions, and prosocial behavior and decreases in negative affect, antisocial intentions, and antisocial behavior and with negative scores denoting the opposite), and the other reflected observed change in response to negative feedback (with positive scores denoting increases in negative affect, antisocial intentions, and antisocial behavior and decreases in positive affect, prosocial intentions, and prosocial behavior and with negative scores denoting the opposite). Thus, the observed change scores were calculated in such a way that a positive score denoted a "change in the expected direction" on one or more dependent measures, a score of zero denoted no change, and a negative score denoted a "change in the opposite direction from what was expected."

Next, these observed change scores were converted to reliable change scores reflecting meaningful reliable change. We were interested in whether some children in the experimental group would display reliable change in the "expected direction" in response to both positive feedback and negative feedback. Results are displayed in Table 5. Although there was a group of 10.31% who showed reliable change in the expected direction in response to negative feedback, this group did not show reliable change in response to positive feedback. Likewise, although 4.12% showed reliable change in the expected direction in response to positive feedback, this group did not show reliable change in response to negative feedback. We tested whether the observed distribution of reliable changers and non-changers differed from the expected random distribution (i.e., 2.5% decrease, 95% stable, and 2.5% increase) using chi-square tests. The observed distribution did not differ from the expected distribution for reliable change in response to positive feedback, $\chi^2(2) = 1.21$, $p = .55$, but it did for reliable change in response to negative feedback, $\chi^2(2) = 24.30$, $p < .001$. In sum, using reliable change scores to examine within-person change, we found no support for a "susceptible" group of children. Instead, we found a group of 10.31% of the children in the experimental group who were "vulnerable" to negative feedback only. Differential susceptibility "for better and for worse," therefore, was not supported by our data.

Discussion

A key assumption underlying differential susceptibility theory is that a subset of individuals exist who display enhanced susceptibility to both negative and positive environmental conditions. This study represents the first attempt to directly test this assumption by exposing the same children to both negative and positive feedback using puppet role-plays. Following calls by Ellis and colleagues (2011) and van IJzendoorn et al. (2015), we introduced an experimental within-participant design to test differential susceptibility, manipulating the microenvironment of children in the experimental group both "for better and for worse." This powerful method provides a closer test of the assumption of "for better and for worse" than the between-participants designs that have been used to date.

The results show that on hearing negative feedback, children in the experimental group increased in negative affect and decreased in positive affect, and they did so more strongly than children in the control group, corroborating previous findings (Kamins & Dweck, 1999; Zentall & Morris, 2010). On hearing positive feedback, children in the experimental group increased in positive affect and decreased in prosocial behavior more strongly than children in the control group, although these findings disappeared after controlling for multiple testing. Crucially, the extent to which children in the experimental group changed in response to positive or negative feedback did not depend on their temperament; children who scored high on negative emotionality did not respond more strongly to positive or negative feedback compared with their counterparts who scored low on this trait. In addition, although approximately 10% of the children in the experimental group changed reliably in response to negative feedback ("vulnerable" children), these same children did not change reliably in response to positive feedback. In sum, although we found support for a subset of "vulnerable" children, we could not find support for a subset of "susceptible" children.

The findings obtained in this study concur with some intervention studies that found no interaction effects (e.g., Andersson et al., 2013; Bockting, Mocking, Lok, Koeter, & Schene, 2013). At the same time, our findings oppose findings from meta-analyses of studies using different designs such as between-participants interventions and correlational studies (Slagt et al., 2016; van IJzendoorn et al., 2015). We must acknowledge that this is the first study of its kind, directly testing "for better and for worse" in the same children. At least two groups of explanations for our findings exist that need to be addressed in future research before definitive conclusions can be drawn. The first pertains to the environmental stimuli that should be used, and the second pertains to the timing of differential susceptibility experiments.

Choosing stimuli in differential susceptibility experiments

Both the strength and evolutionary relevance of environmental stimuli used to test susceptibility are important to consider. Administering feedback using puppet role-play scenarios has been effective in previous studies (e.g., Cimpian et al., 2007; Kamins & Dweck, 1999), and in our study it was effective as well, changing children's positive and negative affect. However, because of ethical concerns, we chose a fairly gentle manipulation. Although we exposed children to both negative and positive stimuli, these stimuli were not extreme and, as such, do not represent the "full" range of environments. This may have limited the opportunity to find individual differences in responsiveness to both negative environments and positive environments and consequently, if they do exist, to find children who are responsive to both types of environments. Especially the positive manipulation yielded small main effects that disappeared after controlling for multiple testing. Stronger environmental stimuli may be required before individual differences in responsiveness to the environment become apparent. This is also suggested by a recent meta-analysis on experimental tests of genetic susceptibility, which revealed that macrotrials showed more evidence of differential susceptibility than microtrials (van IJzendoorn et al., 2015). However, with stronger stimuli, especially negative stimuli, come ethical objections. The balance between the optimal strength of environmental exposure and ethical treatment of participants, therefore, remains a difficult one.

Not only the strength of environmental stimuli but also their evolutionary relevance (i.e., whether they convey information about potential threats or resources that affect survival chances) may matter. Although both positive and negative feedback are informative in general (because they convey information about the availability of social support), they were delivered within puppet role-plays narrated by the experimenter. Probably feedback delivered by parents in real life constitutes a more salient and relevant stimulus from an evolutionary perspective and could lead to larger differences among children in how they would respond. Tentative support for this idea is provided by a functional magnetic resonance imaging (fMRI) study showing that adults higher on sensory processing sensitivity processed information about close others' emotions more thoroughly compared with information about strangers' emotions (Acevedo et al., 2014).

Finally, effects of negative feedback were more persuasive than those of positive feedback. This may be a manifestation of the negativity bias, which refers to the well-established phenomenon that something positive will generally have less of an impact on a person's behavior and cognition than something equally emotional but negative (Baumeister, Bratslavsky, Finkenauer, & Vohs, 2001). Survival requires urgent attention to possible negative outcomes, but it is less urgent with regard to positive outcomes. In the extreme, negative events are more threatening than positive events are beneficial. Hence, it would be adaptive to be psychologically designed to respond to negative events more strongly than to positive events. Future studies testing susceptibility to both negative and positive events within person could take this phenomenon into account by perhaps expecting a higher response threshold for positive stimuli compared with negative stimuli.

Taking into account time and timing in differential susceptibility experiments

When studying differential susceptibility, time is an important yet underappreciated consideration (Pluess, Stevens, & Belsky, 2013). On what time scale, and when during life, do differences in susceptibility make themselves known? Differential susceptibility theory asserts that susceptible individuals should experience sustained developmental change in response to environmental exposures (Ellis et al., 2011; Pluess, 2015). This experiment tapped into "short-term" reactivity. Whether reactivity also implies developmental susceptibility, and vice versa, is an open question (Stamps, 2016). The within-participant experimental design we used is suitable for examining reactivity but is harder to apply to ontogenetic plasticity. Apart from ethical objections to exposing individuals to conditions that are so harsh that they can have a lasting impact on development, carryover effects would also become a serious problem. That is, if individuals, after prolonged exposure and adjustment to one type of environment, would suddenly find themselves in opposite environmental conditions, adjustment to these new conditions would likely be affected by the previous environment. Therefore, examining the notion of a subset of individuals who are susceptible to harsh as well as supportive conditions seems to be possible only when focusing on short-term change, and it is unclear how this relates to long-term change "for better and for worse."

Furthermore, children's susceptibility at any given point in time may depend on their previously experienced environment (Boyce & Ellis, 2005; Ellis, Del Giudice, & Shirtcliff, 2013). Specifically, an initial propensity for susceptibility "for better and for worse" early in life may, for some children, develop into a biased susceptibility toward contextual adversity (i.e., vulnerability) or contextual support (i.e., vantage sensitivity) depending on specific environments encountered early in life (i.e., stress or support) (Pluess, 2015; see Cleveland et al., 2015, for an example). Translated to our experiment, repeated exposure to negative or positive parenting may make children more susceptible to either threatening or supportive cues in their environment and, as such, more likely to respond to only one of the experimental manipulations. Alternatively, instead of increasing susceptibility, repeated exposure to extremely negative parenting might result in a blunted response to negative feedback, whereas repeated exposure to extremely supportive environments could render children indifferent to positive feedback. The latter may have occurred in our relatively well-functioning, high socioeconomic status community sample. We found a group of children who responded especially strongly to negative feedback (perhaps because this was "out of the ordinary" for them), but these children did not respond especially strongly to positive feedback (perhaps because this was nothing special for them). In testing short-term susceptibility to both negative and positive stimuli, future research should try to take into account children's previous parenting experiences and include children from a wider range of socioeconomic backgrounds.

Conclusion

Differential susceptibility theory proposes that a subset of individuals exists who display enhanced susceptibility to both negative (risk-promoting) and positive (development-enhancing) environmental conditions. In exposing the same individuals to experimentally induced positive as well as negative changes in their microenvironment, we failed to find support for this assumption. However, because of the explanations for our findings described above, in particular the small manipulation effects, it would be premature to reject differential susceptibility theory based on this study. Instead, future research is needed to substantiate our findings.

Apart from being an empirical test of differential susceptibility, this article serves as a methodological primer on how to test differential susceptibility. It introduced a new approach that is characterized by the manipulation of children's environment "for better and for worse" using a within-person design. This enables a proper test of the idea that a subset of individuals would be generally susceptible (see Ellis et al., 2011). More broadly, an experimental design has been emphasized as a powerful way to test person-by-environment interactions (Bakermans-Kranenburg & van IJzendoorn,, 2015; van IJzendoorn & Bakermans-Kranenburg, 2012; van IJzendoorn et al., 2015). Not only children's environment (i.e., feedback) but also their response to this environment was sampled from positive as well as negative sides, enabling detection of potential crossover interactions. In addition, random assignment to experimental conditions was stratified according to temperament scores, ruling out temperament–outcome associations as a possible confound (Belsky et al., 2007; Ellis et al., 2011). Finally, we used novel statistical techniques to test for differential susceptibility (latent change score models and reliable change indexes) that directly get at within-person changes in response to both negative and positive environments and that fit the new research design we used.

Building on these strengths, this study provides two clear lessons that can guide future differential susceptibility research. First, within-person designs that expose the same individuals to both negative and positive stimuli (nanotrials or microtrials) are crucial in testing differential susceptibility theory's proposition that a subset of individuals will be susceptible "for better and for worse." Although such designs are incapable of showing long-term developmental effects of susceptibility to social context, they are able to test whether, even within experimental random assignment designs, differential susceptibility is operable and consequential. Thus, such designs are needed to demonstrate proof of principle.

Second, despite the surge of interest in differential susceptibility theory during the past decade, we hardly know what characterizes susceptible individuals if they exist. Are they developmentally susceptible or also susceptible in the short term? Are they susceptible only early in life or throughout life (cf. Belsky & Pluess, 2013; Windhorst et al., 2015)? And finally, if susceptible individuals exist in any of the forms described above, what markers can best be used to identify them? Within-person experimental designs provide a powerful tool to study these questions.

Acknowledgments

Support for this research was provided by the Netherlands Organisation for Scientific Research (NWO Grant 406-11-030), the ISSBD-JF Mentored Fellowship Program for Early Career Scholars, a Fulbright scholarship, and an NWO Visitor Travel Grant (NWO Grant 040.11.494). We extend our sincere thanks and appreciation to the student assistants who helped collect the data and to the families who participated in this study.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jecp.2016.10.004.

References

Acevedo, B. P., Aron, E. N., Aron, A., Sangster, M., Collins, N., & Brown, L. L. (2014). The highly sensitive brain: An fMRI study of sensory processing sensitivity and response to others' emotions. *Brain and Behavior, 4*, 580–594.

Andersson, E., Rück, C., Lavebratt, C., Hedman, E., Schalling, M., Lindefors, N., et al (2013). Genetic polymorphisms in monoamine systems and outcome of cognitive behavior therapy for social anxiety disorder. *PLoS One, 8*(11), e79015.

Asparouhov, T., & Muthén, B. (2010). *Bayesian analysis of latent variable models using Mplus: Technical report*. Retrieved from <http://statmodel.com/papers.shtml>.

Bakermans-Kranenburg, M. J., & van IJzendoorn (2015). The hidden efficacy of Gene × Environment experiments from a differential susceptibility perspective. *Annual Review of Psychology, 66*, 381–409.

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*, 323–370.

Belsky, J. (1997a). Theory testing, effect-size evaluation, and differential susceptibility to rearing influence: The case of mothering and attachment. *Child Development, 68*, 598–600.

Belsky, J. (1997b). Variation in susceptibility to rearing influences: An evolutionary argument. *Psychological Inquiry, 8*, 182–186.
Belsky, J. (2005). Differential susceptibility to rearing influences: An evolutionary hypothesis and some evidence. In B. Ellis & D. Bjorklund (Eds.), *Origins of the social mind: Evolutionary psychology and child development* (pp. 139–163). New York: Guilford.
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., Hsieh, K. H., & Crnic, K. (1998). Mothering, fathering, and infant negativity as antecedents of boys' externalizing problems and inhibition at age 3 years: Differential susceptibility to rearing experience? *Development and Psychopathology, 10*, 301–319.
Belsky, J., & Pluess, M. (2009). Beyond diathesis–stress: Differential susceptibility to environmental influences. *Psychological Bulletin, 135*, 885–908.
Belsky, J., & Pluess, M. (2013). Genetic moderation of early child-care effects on social functioning across childhood: A developmental analysis. *Child Development, 84*, 1209–1225.
Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society B: Methodological, 57*, 289–300.
Blair, C. (2002). Early intervention for low birth weight preterm infants: The role of negative emotionality in the specification of effects. *Development and Psychopathology, 14*, 311–332.
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.
Boyce, W. T., Chesney, M., Alkon, A., Tschann, J. M., Adams, S., Chesterman, B., et al (1995). Psychobiologic reactivity to stress and childhood respiratory illnesses: Results of two prospective studies. *Psychosomatic Medicine, 57*, 411–422.
Boyce, W. T., & Ellis, B. J. (2005). Biological sensitivity to context: I. An evolutionary–developmental theory of the origins and functions of stress reactivity. *Development and Psychopathology, 17*, 271–301.
Cassidy, J., Woodhouse, S. S., Sherman, L. J., Stupica, B., & Lejuez, C. W. (2011). Enhancing infant attachment security: An examination of treatment efficacy and differential susceptibility. *Development and Psychopathology, 23*, 131–148.
Christensen, L., & Mendoza, J. L. (1986). A method of assessing change in a single subject: An alteration of the RC index. *Behavior Therapy, 17*, 305–308.
Cimpian, A., Arce, H., Markman, E. M., & Dweck, C. S. (2007). Subtle linguistic cues impact children's motivation. *Psychological Science, 18*, 314–316.
Cleveland, H. H., Schlomer, G. L., Vandenbergh, 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) genetic variants. *Development and Psychopathology, 27*, 51–67.
Del Giudice, M., Ellis, B. J., & Shirtcliff, E. A. (2011). The adaptive calibration model of stress responsivity. *Neuroscience and Biobehavioral Reviews, 35*, 1562–1592.
Ellis, B. J., Boyce, W. T., Belsky, J., Bakermans-Kranenburg, M. J., & van IJzendoorn, M. H. (2011). Differential susceptibility to the environment: An evolutionary–neurodevelopmental theory. *Development and Psychopathology, 23*, 7–28.
Ellis, B. J., Del Giudice, M., & Shirtcliff, E. A. (2013). Beyond allostatic load: The stress response system as a mechanism of conditional adaptation. In T. P. Beauchaine & S. P. Hinshaw (Eds.), *Child and adolescent psychopathology* (2nd ed., pp. 251–284). New York: John Wiley.
Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.
Fox, E., Zougkou, 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.
Iannotti, R. J. (1985). Naturalistic and structured assessments of prosocial behavior in preschool children: The influence of empathy and perspective taking. *Developmental Psychology, 21*, 46–55.
Kamins, M. L., & Dweck, C. S. (1999). Person versus process praise and criticism: Implications for contingent self-worth and coping. *Developmental Psychology, 35*, 835–847.
Klein Velderman, M., Bakermans-Kranenburg, M. J., Juffer, F., & Van IJzendoorn, M. H. (2006). Effects of attachment-based interventions on maternal sensitivity and infant attachment: Differential susceptibility of highly reactive infants. *Journal of Family Psychology, 20*, 266–274.
Lee, S. Y., & Song, X. Y. (2004). Evaluation of the Bayesian and maximum likelihood approaches in analyzing structural equation models with small sample sizes. *Multivariate Behavioral Research, 39*, 653–686.
Lieberman, J. D., Solomon, S., Greenberg, J., & McGregor, H. A. (1999). A hot new way to measure aggression: Hot sauce allocation. *Aggressive Behavior, 25*, 331–348.
McArdle, J. J. (2009). Latent variable modeling of differences in changes with longitudinal data. *Annual Review of Psychology, 60*, 577–605.
Murphy, B. C., & Eisenberg, N. (1997). Young children's emotionality, regulation, and social functioning and their responses when they are targets of a peer's anger. *Social Development, 6*, 18–36.
Muthén, L. K., & Muthén, B. O. (1998). *Mplus user's guide* (7th ed.). Los Angeles: Muthén & Muthén.
Pluess, M. (2015). Individual differences in environmental sensitivity. *Child Development Perspectives, 9*, 138–143.
Pluess, M., & Belsky, J. (2010). Children's differential susceptibility to effects of parenting. *Family Science, 1*, 14–25.
Pluess, M., & Boniwell, I. (2015). Sensory-processing sensitivity predicts treatment response to a school-based depression prevention program: Evidence of vantage sensitivity. *Personality and Individual Differences, 82*, 40–45.
Pluess, M., Stevens, S., & Belsky, J. (2013). Differential susceptibility: Developmental and evolutionary mechanisms of gene–environment interactions. In M. Legerstee, D. W. Haley, & M. H. Bornstein (Eds.), *The infant mind: Origins of the social brain* (pp. 77–96). New York: Guilford.
Preacher, K. J., Rucker, D. D., MacCallum, R. C., & Nicewander, W. A. (2005). Use of the extreme groups approach: A critical reexamination and new recommendations. *Psychological Methods, 10*, 178–192.
Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment, 87*, 103–113.

Quas, J. A., Bauer, A., & Boyce, W. T. (2004). Physiological reactivity, social support, and memory in early childhood. *Child Development, 75*, 797–814.

Roisman, G. I., Newman, D. A., Fraley, R. C., Haltigan, J. D., Groh, A. M., & Haydon, K. C. (2012). Distinguishing differential susceptibility from diathesis–stress: Recommendations for evaluating interaction effects. *Development and Psychopathology, 24*, 389–409.

Rotenberg, K. J., & Eisenberg, N. (1997). Developmental differences in the understanding of and reaction to others' inhibition of emotional expression. *Developmental Psychology, 33*, 526–537.

Rothbart, M. K., Ahadi, S. A., Hershey, K. L., & Fisher, P. (2001). Investigations of temperament at 3–7 years: The Children's Behavior Questionnaire. *Child Development, 72*, 1394–1408.

Russell, J. A. (1990). The preschooler's understanding of the causes and consequences of emotion. *Child Development, 61*, 1872–1881.

Rutter, M. (2006). *Genes and behavior: Nature–nurture interplay explained*. Oxford, UK: Blackwell.

Scott, S., & O'Connor, T. G. (2012). An experimental test of differential susceptibility to parenting among emotionally-dysregulated children in a randomized–controlled trial for oppositional behavior. *Journal of Child Psychology and Psychiatry, 53*, 1184–1193.

Slagt, M. I., Dubas, J. S., Deković, M., & van Aken, M. A. G. (2016). Differences in sensitivity to parenting depending on child temperament: A meta-analysis. *Psychological Bulletin, 142*, 1068–1110.

Stamps, J. A. (2016). Individual differences in behavioural plasticities. *Biological Reviews, 91*, 534–567.

Suomi, S. J. (1997). Early determinants of behaviour: Evidence from primate studies. *British Medical Bulletin, 53*, 170–184.

Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective taking and its relation to prosocial behavior in toddlers. *Developmental Psychology, 45*, 534–543.

van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & van Aken, M. A. G. (2014). A gentle introduction to Bayesian analysis: Applications to research in child development. *Child Development, 85*, 842–860.

van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2012). Differential susceptibility experiments: Going beyond correlational evidence—Comment on beyond mental health, differential susceptibility articles. *Developmental Psychology, 48*, 769–774.

van IJzendoorn, M. H., & Bakermans-Kranenburg, M. J. (2015). Genetic differential susceptibility on trial: Meta-analytic support from randomized controlled experiments. *Development and Psychopathology, 27*, 151–162.

Windhorst, D. A., Mileva-Seitz, V. R., Linting, M., Hofman, A., Jaddoe, V. W. V., Verhulst, F. C., et al (2015). Differential susceptibility in a developmental perspective: DRD4 and maternal sensitivity predicting externalizing behavior. *Developmental Psychobiology, 57*, 35–49.

Zentall, S. R., & Morris, B. J. (2010). "Good job, you're so smart": The effects of inconsistency of praise type on young children's motivation. *Journal of Experimental Child Psychology, 107*, 155–163.

Zuckerman, M. (1999). *Vulnerability to psychopathology: A biosocial model*. Washington, DC: American Psychological Association.

Zumbo, B. D., Gadermann, A. M., & Zeisser, C. (2007). Ordinal versions of coefficients alpha and theta for Likert rating scales. *Journal of Modern Applied Statistical Methods, 6*, 21–29.