Report

The nature of the association between moral neutralization and aggression
a systematic test of causality in early adolescence

Author(s):
Ribeaud, Denis; Eisner, Manuel

Publication Date:
2014

Permanent Link:
https://doi.org/10.3929/ethz-a-010056646

Rights / License:
In Copyright - Non-Commercial Use Permitted

This page was generated automatically upon download from the ETH Zurich Research Collection. For more information please consult the Terms of use.
THE NATURE OF THE ASSOCIATION BETWEEN MORAL NEUTRALIZATION AND AGGRESSION: A SYSTEMATIC TEST OF CAUSALITY IN EARLY ADOLESCENCE

Denis Ribeaud*, Chair of Sociology, Swiss Federal Institute of Technology at Zurich, Switzerland; Manuel Eisner, Institute of Criminology, University of Cambridge, United Kingdom;

*Corresponding author:
Denis Ribeaud
Criminological Research Unit
Chair of Sociology
ETH Zurich, RZ E4.1
Clusiusstrasse 59
CH-8092 Zurich
Switzerland
E-Mail: ribeaudd@ethz.ch
Voice: +41 44 632 97 32
Fax: +41 44 632 13 41

Acknowledgements:
This study was supported by grants of the Jacobs Foundation and of the Swiss National Science Foundation.
THE NATURE OF THE ASSOCIATION BETWEEN MORAL NEUTRALIZATION AND AGGRESSION: A SYSTEMATIC TEST OF CAUSALITY IN EARLY ADOLESCENCE
ABSTRACT
The present contribution examines possible causal linkages among moral neutralization (moral neutralization) – a generic term for the related concepts of neutralization techniques, moral disengagement, and self-serving cognitive distortions – and aggressive behavior using a set of repeated measures at ages 11.4 and 13.7 in a culturally diverse urban sample (N=1032).
First, correlational analysis shows a strong cross-sectional association between moral neutralization and aggression. Second, fixed-effects regressions suggest a substantial within-individual association implying that the cross-sectional association cannot be explained away by population heterogeneity. The within-individual association also remains stable when controlling for a number of potential confounders, suggesting a direct causal relationship. Third, path analyses reveal near-zero lagged effects of moral neutralization on aggression when controlling for antecedent aggression and vice-versa, thus suggesting no longer-term independent causal effects in either direction. Moreover, synchronous effects of moral neutralization on aggression when controlling for antecedent aggression and vice-versa are same-sized and highly significant. Overall, results suggest a close short-term interdependence of both constructs.
INTRODUCTION & STATE OF RESEARCH

The present contribution examines possible causal linkages between moral neutralization and aggressive behavior in early adolescence. Moral neutralization is a generic term for a set of closely related concepts from different fields of research such as the *techniques of neutralization* introduced in the field of criminology by Sykes and Matza (1957), *moral disengagement* as an element of Bandura’s social cognitive theory (e.g., Bandura, Barbaranelli, & Caprara, 1996), or *secondary self-serving cognitive distortions* proposed by Gibbs and Barriga (e.g., Barriga & Gibbs, 1996) in the field of developmental psychology and young offender treatment. Ribeaud and Eisner (2010a) offer a detailed overview on the conceptual and empirical overlap among these concepts. Other closely related concepts include the social psychological framework of *moral hypocrisy* (e.g., Batson, Thompson, Seuferling, Whitney, & Strongman, 1999) and *moral rationalizations* (Tsang, 2002). In substance, moral neutralization refers to a set of cognitive processes by means of which an individual who is generally rule-abiding and compliant with moral standards can minimize cognitive dissonance, threats to self-concept, and experiences of moral self-sanction when she or he transgresses those standards (Ribeaud & Eisner, 2010a). Put more simply, moral neutralization refers to (self-)justifications of moral transgressions and entails four key mechanisms, namely (1) cognitive restructuring or reframing of reprehensible behavior, (2) minimizing one’s own agency or responsibility, (3) disregarding or distorting the negative impact of detrimental behavior, and (4) blaming, dehumanizing or denying the victim.

In the last two decades much evidence has been brought forth in support of a substantial association between moral neutralization and detrimental behavior in general, and aggressive behavior in particular (for overviews see, e.g., Fritsche, 2005; Gini, Pozzoli, & Hymel, 2013; Obermann, 2011; Ribeaud & Eisner, 2010a). However, most of this research is cross-sectional (Fritsche, 2005; Maruna & Copes, 2005) and thus fails to establish the nature of the temporal and causal order between moral neutralization and aggressive behavior. We are aware of only three significant studies that
analyzed the relationship longitudinally. Agnew (1994) found a small, yet significant independent effect of prior neutralizations on later violence ($\beta=.08$) when controlling for prior violence and other possible confounders in a representative adolescent sample. There was a time lag of one year between both measurements. Paciello, Fida, Tramontano, Lupinetti, and Caprara (2008) found a correlation between trajectories of moral disengagement in adolescence on later aggression at age 20. Hyde, Shaw, and Molainen (2010) found a substantial ($\beta=.34$) independent effect of moral disengagement at age 15 on antisocial behavior one to two years later when controlling for social information processing. Importantly, however, the model did not control for antecedent aggression. Disentangling the temporal order is important because it allows gathering key evidence on the causal direction of the link between moral neutralization and aggression and other detrimental behavior. Only if moral neutralization precedes aggression it can be conceived as a cause or at least as a facilitator of detrimental behavior. Otherwise, it would rather need to be conceived as a consequence or “cognitive reflection” of detrimental behavior, in the sense of ex-post rationalizations. In this respect, the different approaches considered here provide mixed hypotheses. According to Bandura’s social cognitive theory “people do not ordinarily engage in reprehensible conduct until they have justified to themselves the rightness of their actions” (Bandura et al., 1996, p. 365). In the same vein, and more generally, Bandura (1991) assumes that “most human behavior, being purposive, is regulated by forethought” (p. 248). Hence, in this perspective processes of moral disengagement are explicitly conceived as preceding detrimental behavior and as being causally involved in its generation (see Bandura et al. (1996, p. 366-367) for description of the assumed causal model). Similarly, but only as a possibility, in their neutralization theory Sykes and Matza assume that “there is also reason to believe that [justifications for deviance] precede deviant behavior and make deviant behavior possible” (Sykes & Matza, 1957, p. 666). Finally, when Barriga and Gibbs (1996) state that “secondary cognitive distortions have been characterized as pre- or post-
transgression rationalizations that serve to ‘neutralize’ conscience or guilt and thereby to prevent damage to the self-image following antisocial behavior” (p. 334) their reception of the two previous approaches remains ambivalent as regards the temporal order that relates the two constructs. This theoretical ambivalence regarding the temporal order calls for empirical elucidation. Obviously, only experimental or longitudinal designs are suited to test assumptions on the temporal order and to assess the causal links that relate aggression and moral neutralization (see also Maruna and Copes, 2005, p. 45). Such contributions still being scarce, the present paper aims at further fill this gap.

AIM OF THE PRESENT STUDY

The aim of this study is to determine, within a longitudinal framework, to what extent moral neutralization and aggression are directly causally linked to each other and to examine the temporal order underlying such a causal relationship in early adolescence. To this purpose, we first assess the cross-sectional association between aggression and moral neutralization. Once this association has been corroborated, the main focus is on a sequential test of hypotheses regarding the nature of the relationship. First, we explore if the relationship is spurious, i.e., whether it can be “explained away” by unobserved and by observed time-varying and time-invariant covariates that previous research has identified as key predictors of aggression and delinquency in a large range of risk domains (see e.g., Farrington, 1998; Hawkins et al., 1998; Ribcaud & Eisner, 2010b; Wikström & Butterworth, 2006). Specifically, we consider self-control as a key personality characteristic related to aggression and delinquency (Gottfredson & Hirschi, 1990) as well as parenting behavior, substance use, adult media use, deviant peers, unstructured leisure activities, peer victimization, parental SES, migration background, gender, and age. The focus of this part of the analysis is on within-individual change which has, to our knowledge, not yet been analyzed in the research on the link between moral neutralization and aggressive behavior.
Should there remain a stable relationship when controlling for these factors, i.e., should there be evidence for a direct causal within-individual relationship, we will secondly scrutinize its direction and timing. Specifically, using cross-lagged models, we will compare the effect sizes of moral neutralization on later aggressive behavior with those of aggressive behavior on later moral neutralization. Moreover, using synchronous reciprocal models, we will examine the short-term effects of both constructs on each other.

Within-individual change being at the core of this study, it appears judicious to focus on early adolescence, a biographical stage characterized by change and transition in many of the above-mentioned risk domains (e.g., Steinberg & Silverberg, 1986).

METHODS

Sample & Participants

The analyses are based on data from the Zurich Project on the Social Development of Children and Youths, an experimental prospective longitudinal study on the development of aggressive and other antisocial behavior that was set-up in an culturally diverse urban context in Europe (e.g., Eisner, Malti, & Ribeaud, 2012; Ribeaud & Eisner, 2010b). The initial target sample consisted of 1675 children who entered one of 56 randomly selected public schools in the city of Zurich, Switzerland, at age 7 in 2004. Overall, the sample is representative of the city’s child population. So far, five data collection waves have been completed between age 7.5 and 13.7. For the present study, we used data from child assessments in data collection waves 4 and 5 at ages M=11.4 years and M=13.7 years, respectively (henceforth referred to as “age 11” and “age 13”). These two data collections have been selected because the two key measures used in this article – i.e., the moral neutralization scale and the extended written version of the aggressive behavior assessment instrument – were introduced at wave 4 when the survey format changed from personal interviews to a self-completion
questionnaire. Overall, 1032 cases with complete data at both data collection waves were available for analysis corresponding to 61.6 percent of the target sample, and 75.8 percent of the initial sample. 51.1 percent were male, 89.4 percent were born in Switzerland and 44.8 percent had a migration background, i.e., both biological parents were born abroad, chiefly in former Yugoslavian Republics, Sri Lanka, Germany, Portugal, and Turkey. At age 13, 71.4 percent were living with both biological parents.

Consent Procedure & Data Collection

Prior to data collection ages 11 and 13 parents were informed in written form. At age 11 they were requested to sign a consent form (active consent procedure), at wave 5 they were given the opportunity to refuse their child’s participation in the study (passive consent procedure). At the beginning of each data collection participants were informed in detail about the study and about their rights, in particular the right not to answer particular questions. Participants were then asked to sign a written informed consent form. Data were collected by means of paper and pencil surveys in classrooms with 5 to 15 participants and lasted for about 90 minutes. Participants were guided through the questionnaire by 2 or 3 trained staff members. At age 11 the data collection was carried out during regular school lessons whereas at age 13 it took place during leisure time. For this reason participants were given a participation incentive in cash worth 30 US$.

Measures

Moral neutralization was measured with the 16-item instrument described by Ribeaud and Eisner (2010a, 2012 #747). It is based on items of instruments derived from the three theoretical approaches described in the introduction, i.e. moral disengagement, neutralization techniques, and secondary self-serving biases. Specifically, the items stem from Bandura et al.’s (1996) moral
disengagement scale, Hymel, Rocke-Henderson, and Bonanno’s (2005) bullying-focused moral disengagement scale, Huizinga and colleagues short neutralization techniques scale used in the Denver Youth Survey (e.g., Huizinga, Weiher, Espiritu, & Esbensen, 2003), as well as from a Dutch adaptation of Barriga and Gibbs (1996) “How I think” questionnaire that specifically focuses on self-serving cognitive distortions related to aggressive behavior (van der Velden, 2008). The scale covers the four key mechanisms of moral neutralization, i.e., cognitive restructuring (7 items), blaming the victim (3 items), distorting negative impact (3 items), and minimizing own agency (2 items). All items were measured on a 4-point Likert scale. Confirmatory factor analyses suggest a stable one-dimensional factor structure across waves with high and stable reliability coefficients of $\alpha=.87$ at age 11 and $\alpha=.89$ at age 13. For the purpose of the present research a mean-score scale was used.

Aggression was measured with the 12-item subscale of aggression of an adapted adolescent version of the Social Behavior Questionnaire (Tremblay et al., 1991). It assesses physical, proactive, reactive, and indirect aggression in the last six months on a 5-point Likert scale. Confirmatory factor analyses suggest a stable one-dimensional second-order factor structure, the first level representing the four sub-dimensions of aggression. We found stable reliability coefficients of $\alpha=.81$ at age 11 and $\alpha=.86$ at age 13. Again, a mean-score scale was derived for the present research.

Covariates. Time-varying covariates used in the fixed-effects regression models include low self-control (10-item mean-score scale, adapted from Grasmick, Tittle, Bursik, and Arneklev (1993)), substance use (3-item variety index of last year tobacco, alcohol, and cannabis use), aversive parenting (5-item mean-score scale derived from the Alabama Parenting Questionnaire (Shelton, Frick, & Wootton, 1996) on harsh and inconsistent parenting), adult media use (3-item variety index of watching adult horror, action, and other movies), deviant friends (mean-score scale across two 6-item variety indices of substance use, violence and theft among two best friends), unstructured
leisure activities (8-item mean-score scale of unstructured and unsupervised out-of-home leisure activities), peer victimization (4-items mean-score scale of four types of peer victimization). The time-invariant covariates used in the fixed-effects regression models were sex (coded 0 for males, 1 for females), date of birth, parental SES (International Socio-Economic Index of Occupational Status (Ganzeboom, Degraaf, Treiman, & Deleuw, 1992)), parental educational achievement (10-level scale), and migration status (0 if at least one parent was born in Switzerland, 1 otherwise). Descriptive statistics are available on request from the first author.

**Analytical Strategy**

In what follows we adopt a multiple-stage analytical strategy. First, we describe the development of moral neutralization and aggression across the period of observation as well as their cross-sectional association at both points of measurement. Second, the causal nature of the relationship is tested within the framework of within-individual change models. Specifically, using a hierarchical modelling approach, three two-period fixed-effects regression models (Allison, 2009, pp. 6-12) are estimated. At each step a block of additional covariates is included in the model. In the first model, we assess to what extent the baseline association found can be accounted for by unobserved population heterogeneity (Nagin & Paternoster, 2000), i.e., by time-stable, unobserved differences in the study population that affect the levels of both aggression and moral neutralization at any point in time. In the framework of fixed-effects regressions, unobserved differences are controlled for by using each individual as his/her own control. De facto, in the case of two-period fixed-effects models, within-individual difference scores of the dependent variable are regressed on within-individual difference scores of the independent variable (Allison, 2009, p. 14).
Should this first model suggest substantial *within-individual* covariation between aggression and moral neutralization, i.e., that the relationship is not attributable to population heterogeneity, in a second step the model would be extended by selected time-varying independent variables. That is, in this second model we test if the within-individual covariation can be accounted for by other *time-variant* characteristics or processes such as new friends, variations in the parenting style, or changes related to leisure activities, substance use or media use. Finally, in a third model, an additional set of variables that controls for *time-varying effects* of selected *time-invariant predictors* is included in the model (e.g., time-varying effects of gender or SES).

Having gathered evidence for the (direct) causal nature of the relationship, in a third set of analyses we will examine this relationship’s timing and direction, by means of (1) cross-lagged and (2) synchronous reciprocal effects models. Essentially, the cross-lagged model tests if moral neutralization at age 11 has an independent effect on aggression at age 13 when controlling for aggression at age 11, and vice-versa, whereas the synchronous reciprocal effects model tests if moral neutralization at age 13 has an independent effect on aggression at age 13 when controlling for aggression at age 11, and vice-versa.

**RESULTS**

*Descriptive Results*

Between ages 11 and 13 the moral neutralization score highly significantly increased ($t(1031)=19.5$, $p<.001$) from $M=1.68$ ($SD=0.48$) to $M=2.02$ ($SD=0.54$). Similarly, the aggression score increased ($t(1031)=15.4$, $p<.001$) from $M=1.52$ ($SD=0.43$) to $M=1.78$ ($SD=0.56$), suggesting that both constructs follow a similar trend between age 11 and 13. The two constructs are also comparatively stable across time with cross-wave correlations of $r=.397$ as to moral neutralization and $r=.440$ as to aggression. Furthermore there is a pronounced cross-sectional correlation between both constructs
that remains stable across time, with correlation coefficients of \( r = .611 \) at age 11 and of \( r = .652 \) at age 13. Finally, moral neutralization at age 11 is substantially correlated with later aggression at age 13 (\( r = .305 \)) and vice-versa (\( r = .297 \)). All reported correlations are significant at \( p < .001 \).

**Fixed-Effects Regressions**

Having established a strong cross-sectional, i.e., *inter-*individual correlation between aggression and moral neutralization we now examine whether more rigorous evidence for a causal nature of this association can be found or whether it can be accounted for by observed and unobserved covariates. As outlined, we carried out a series of two-period fixed-effects regression models using the dummy-variable method implemented in STATA 11 (*xi:xtreg [...], fe procedure*) (Allison, 2009). Substantially, this implies that we look at within-individual change. Since at this stage of analysis the direction of the relationship is not yet of interest, all models were calculated one time with aggression as the dependent variable (left column in Table 1) and the other time with moral neutralization as the dependent variable (right column in Table 1).\(^1\) First, we tested if the association can be explained by unobserved population heterogeneity. In other terms, we examined if the association reflects unobserved differences in the study population that account for the association. This is achieved by regressing *within-individual* change scores of moral neutralization on *within-individual* change scores of aggression and vice-versa (Model 1 in Table 1). The corresponding regression weights of moral neutralization (\( B = 0.551; SE(B) = 0.025 \)) and aggression (\( B = 0.596; SE(B) = 0.027 \)) are highly significant. This means that the association between aggression and moral neutralization cannot be accounted

\(^1\) The critical ratio (\( B / SE(B) \)) and the associated p-value of the effects of moral neutralization/aggression remain constant across the left-hand and right-hand models, i.e., the information of interest in both columns is redundant.
for by time-invariant effects of unobserved population heterogeneity since the association can also be observed as a change-on-change association within individuals.

[Table 1 about here]

Second, we further included a set of time-varying covariates that previous research has identified as key predictors of aggressive behavior and juvenile delinquency, the assumption being that within-individual change in these variables might potentially account for the within-individual association found in Model 1. The coefficients of moral neutralization and aggression in Model 2 (Table 1) suggest that these covariates attenuate the association between aggression and moral neutralization to some extent. However, the effects are still considerable ($B=0.427; SE(B)=0.026$, and $B=0.477; SE(B)=0.029$, respectively) and highly significant. When further extending the model by allowing for time-variant effects of time-invariant covariates – including age, sex, SES, parental educational achievement, and migration status – the association between moral neutralization and aggression remains almost unaffected as compared to Model 2 and highly significant (Model 3 in Table 1). Hence, provided that key covariates have not been omitted, the results of the fixed-effects regression are in support of a direct causal relationship between moral neutralization and aggression.

**Cross-lagged Models**

The next set of models focuses on the direction and on the temporal structure of the hypothesized causal relationship between moral neutralization and aggression. In order to ease model parameterization and to facilitate comparisons across effect sizes, all analyses presented in what follows are based on z-standardized variables and consequently standardized coefficient values are reported. With regard to the notation used for the model specifications and for the presentation of
the results the reader is asked to refer to the generic path model represented in Figure 1. All paths models were estimated with the AMOS 20 structural equation modeling software using maximum likelihood estimators.

In the baseline parameterization of the cross-lagged model all parameters are freely estimated except $\beta_5$ and $\beta_6$ which are constrained to zero. This initial saturated model ($\chi^2=0; df=0$) allows to test if moral neutralization at age 11 has an independent effect on aggression at age 13 when controlling for aggression at age 11 and vice-versa. With $\beta_3=.058$ ($p=.098$) and $\beta_4=.086$ ($p=.017$) these effects are very weak and only partially significant. Since constraining $\beta_3$ and $\beta_4$ to equality does not decrease model fit significantly ($\chi^2=0.218; df=1; p=.641$) we find no evidence suggesting that the lagged effect is stronger in one direction than in the other, i.e., there is no clear indication as to the main direction of the causal relationship. Finally, constraining both $\beta_3$ and $\beta_4$ to zero results in a significant decrease in model fit compared to the initial saturated model ($\chi^2=13.3; df=2; p=.001$). However, the decrease in CFI from 1.000 to 0.992 suggests that this restriction affects the model fit only to a very limited extent, so that it appears to be an acceptable model specification. In sum, the cross-lagged effects of moral neutralization on aggression and of aggression on moral neutralization are equal and they are very close to zero.

Synchronous Reciprocal Effects Models

Given the lack of longer-term cross-lagged effects we may now look at synchronous reciprocal effects. Importantly, in two-wave designs, such models can only be estimated if cross-lagged effects are (near) zero (Finkel, 1995, p. 34). Accordingly, in the synchronous reciprocal effects model all
parameters are freely estimated, except $\beta_3$ and $\beta_4$, which are constrained to zero. In substance, the synchronous reciprocal effects model hence tests whether moral neutralization at age 13 has an independent effect on aggression at the same age when controlling for aggression at age 11 and vice-versa.

In the initial saturated model ($\chi^2=0; df=0$) both synchronous regression paths are (near) significant ($\beta_5=.170; p=.067; \beta_6=.213; p=.007$). Constraining both regression weights to equality does not decrease model fit significantly ($\chi^2=0.218; df=1; p=.641$). However, the significance of the parameters is increased by this constraint ($\beta_5 = \beta_6 = .194; p<.001$). In essence, there appears to be a significant synchronous reciprocal effect of the same size in either direction. Again, there is no evidence for a clear causal direction between aggression and moral neutralization. Eventually, we also tested a model without correlated errors, i.e., $\rho^2=0$ ($\chi^2=7.168; df=1; p=.007$) which results in stronger reciprocal effects ($\beta_5=.339; p<.001; \beta_6=.338; p<.001$). However, in this model too, constraining the reciprocal effects to equality hardly affects the model fit ($\Delta \chi^2=0.000; df=1; p=.995$).

In sum, the synchronous effects models suggest substantial and equal effects in either direction.

**SUMMARY & CONCLUSIONS**

In this contribution we examined the nature of the association between moral neutralization and aggression based on self-reported data. In particular, we examined to what extent this association can be understood as causal in nature as well as the timing and direction of this relationship.

First, we found a pronounced and stable cross-sectional, inter-individual association between moral neutralization and aggression at ages 11 and 13. With correlations clearly above $r=.5$ this association turned out to be much stronger than what was found in most earlier studies. For example, in their recent meta-analysis, Gini et al. (2013) report a mean correlation of $r=.28$ between aggression and moral disengagement. The exceptionally high correlations found in the present study are likely due to shared method variance (i.e., self-reports), to the use of highly reliable multiple-indicator scales
for both moral neutralization and aggression, and also, importantly, to a moral neutralization scale that is specifically targeted at aggressive behavior.

There was also a considerable amount of within-individual change in both constructs across time which allowed modeling the within-individual relationship between both constructs. Within-individual models of change have the advantage to control for population heterogeneity, i.e., for unobserved differences in the sample population that account for both moral neutralization and aggression. The corresponding fixed-effects regression models suggested that over a period of two years in early adolescence changes in moral neutralization substantially covaried with changes in aggression within individuals, which is much stronger evidence for a causal relationship between the two constructs than between-individual correlations. We further explored if the within-individual association was possibly not genuine but rather reflecting other processes of within-individual change known to be associated with the development of aggressive and delinquent behavior, such as shifts in parenting behavior, changes related to life-style and leisure activities (e.g., onset of substance use, association with delinquent peers, changes in respect of the use of adult media contents etc.), or episodes of peer victimization. Many of these within-individual processes turned out to be associated with changes in moral neutralization and/or aggression. However, the within-individual association between moral neutralization and aggression remained highly stable and significant when controlling for these potentially confounding processes, thus suggesting a direct causal relationship between moral neutralization and aggression. This effect also remained unaltered when we further controlled for time-varying effects of time-invariant variables such a sex, SES, or migration background.

In a next step we thus examined the temporal and directional order of this relationship within the framework of two-period paths models. The first set of models showed near-zero lagged effects of moral neutralization on aggression when controlling for antecedent aggression and vice-versa, thus
suggesting that there are no substantial longer-term independent causal effects in either direction. Note, however, that the effect size we found ($\beta = .07$) was virtually the same as the one found by Agnew (1994) in his study. In substance this implies that previous moral neutralization is not – or very limitedly – predictive of shifts in aggressive behavior just as antecedent aggressive behavior does not appear to substantially predict changes in moral neutralization in the longer term (i.e., two years).

The second set of paths models examined synchronous effects of moral neutralization on aggression when controlling for antecedent aggression and vice-versa. This analysis showed highly significant effects of the same size in either direction. Both the lack of lagged effects and the substantial and equal reciprocal synchronous effects suggest a close short-term interdependence of both constructs.

Note that it is a limitation of the present study, and of all similarly designed longitudinal studies, that they cannot clearly identify cause-effect sequences that occur at time intervals shorter than the time between data collection waves. Thus, the findings suggest reciprocal causal effects and they suggest that causal effects had a delay of less than 2 years.

Taken together the key findings of this research, i.e., the very substantial direct within-individual association of change in moral neutralization with change in aggression along with the reciprocal synchronous effects indicate that moral neutralization and aggression are intrinsically tied to each other, i.e., there is not one that can be viewed as genuinely exogenous to the other, as is typically implied when moral neutralization is modeled as a predictor of aggression in most extant research.

In this new perspective moral neutralization could be conceived as the cognitive and aggression as the behavioral expression of the same phenomenon. Specifically, in the process of (aggressive) decision-making moral neutralization might be envisaged as facilitating aggressive behavior by providing ex-ante justifications whereas aggressive behavior would in turn induce ex-post legitimations that allow a smooth integration of norm-breaking behavior into an apparently intact moral self-concept. This interpretation is in line with Matza’s (1964) conception of soft determinism,
where effect and cause are not related in a deterministic and unidirectional way to each other. Instead, a cause (e.g., moral neutralization) affects an outcome (e.g., aggression) in a way that leaves room for individual agency in the process of decision making. In turn, the outcome affects the initial cause in a process of feedback. This conception also comes close to Bandura’s general notion of reciprocal determinism in which cognitive, behavioral and environmental factors dynamically interact and influence each other bidirectionally (see, e.g., Bandura, 1991). Hence, social cognitive theory already offers a framework that would allow to integrate the findings of the present study and to extend the current unidirectional causal model of the relationship between moral disengagement and detrimental behavior proposed by Bandura (e.g., Bandura et al., 1996).

Overall, our findings suggest that future research and theory development should primarily focus on dynamic, reciprocal processes, whereas unidirectional causal models appear of limited relevance. To test such dynamic models it will be important to go beyond the limitations of the present study in several ways. Specifically, in order to assess the generalizability of our findings, the present research would benefit from replication in samples of different ages and cultures. Moreover, longitudinal analyses with three and more data waves would allow for more refined causal models of within-individual change. Also, repeated measures at much shorter intervals would further advance our knowledge on the shorter-term dynamics that link moral neutralization and aggression. Finally, experimental designs, and especially designs which entail “hot” decision-making, would offer a promising alternative way to understand the short-term dynamics underlying this link.
REFERENCES


TABLE 1

Fixed-effects regression models, effect sizes of moral neutralization on aggression (left side) and vice-versa (right side)

<table>
<thead>
<tr>
<th>Dependent: Aggression</th>
<th></th>
<th></th>
<th></th>
<th>Dependent: Moral Neutralization</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B</td>
<td>SE (B)</td>
<td>p (B)</td>
<td></td>
<td>B</td>
<td>SE (B)</td>
<td>p (B)</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td>Model 2 (Model 1 + time-varying covariates (*))</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Moral Neutralization</td>
<td>0.551</td>
<td>0.025</td>
<td>.000</td>
<td></td>
<td>0.551</td>
<td>0.025</td>
<td>.000</td>
</tr>
<tr>
<td>Aggression</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td></td>
<td>0.596</td>
<td>0.027</td>
<td>.000</td>
</tr>
<tr>
<td>Time Dummy</td>
<td>0.071</td>
<td>0.016</td>
<td>.000</td>
<td></td>
<td>0.187</td>
<td>0.016</td>
<td>.000</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.593</td>
<td>0.042</td>
<td>.000</td>
<td></td>
<td>0.775</td>
<td>0.042</td>
<td>.000</td>
</tr>
<tr>
<td>*Low Self-Control</td>
<td>0.171</td>
<td>0.030</td>
<td>.000</td>
<td></td>
<td>0.173</td>
<td>0.031</td>
<td>.000</td>
</tr>
<tr>
<td>*Substance Use</td>
<td>0.039</td>
<td>0.050</td>
<td>.441</td>
<td></td>
<td>0.084</td>
<td>0.053</td>
<td>.111</td>
</tr>
<tr>
<td>*Aversive Parenting</td>
<td>0.120</td>
<td>0.028</td>
<td>.000</td>
<td></td>
<td>0.115</td>
<td>0.030</td>
<td>.000</td>
</tr>
<tr>
<td>*Adult Media Use</td>
<td>-0.007</td>
<td>0.035</td>
<td>.831</td>
<td></td>
<td>0.003</td>
<td>0.036</td>
<td>.934</td>
</tr>
<tr>
<td>*Deviant Friends</td>
<td>0.289</td>
<td>0.082</td>
<td>.000</td>
<td></td>
<td>0.195</td>
<td>0.087</td>
<td>.025</td>
</tr>
<tr>
<td>*Unstructured Leisure Activities</td>
<td>0.013</td>
<td>0.015</td>
<td>.385</td>
<td>0.028</td>
<td>0.015</td>
<td>.067</td>
<td></td>
</tr>
<tr>
<td>*Peer Victimization</td>
<td>0.071</td>
<td>0.015</td>
<td>.000</td>
<td>-0.023</td>
<td>0.016</td>
<td>.147</td>
<td></td>
</tr>
<tr>
<td>Time Dummy</td>
<td>0.028</td>
<td>0.019</td>
<td>.136</td>
<td>0.115</td>
<td>0.019</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>0.119</td>
<td>0.070</td>
<td>.092</td>
<td>0.407</td>
<td>0.073</td>
<td>.000</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 continued on next page.
TABLE 1 (continued)

<table>
<thead>
<tr>
<th></th>
<th>Dependent: Aggression</th>
<th></th>
<th>Dependent: Moral Neutralization</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>(Model 2 + time-invariant covariates (**))</td>
<td></td>
</tr>
<tr>
<td>Moral Neutralization</td>
<td>0.419</td>
<td>0.026</td>
<td>.000</td>
<td>--</td>
</tr>
<tr>
<td>Aggression</td>
<td>--</td>
<td>--</td>
<td>--</td>
<td>0.477</td>
</tr>
<tr>
<td>*Low Self-Control</td>
<td>0.168</td>
<td>0.029</td>
<td>.000</td>
<td>0.171</td>
</tr>
<tr>
<td>*Substance Use</td>
<td>0.031</td>
<td>0.050</td>
<td>.538</td>
<td>0.094</td>
</tr>
<tr>
<td>*Aversive Parenting</td>
<td>0.132</td>
<td>0.028</td>
<td>.000</td>
<td>0.108</td>
</tr>
<tr>
<td>*Adult Media Use</td>
<td>-0.028</td>
<td>0.035</td>
<td>.412</td>
<td>0.006</td>
</tr>
<tr>
<td>*Deviant Friends</td>
<td>0.289</td>
<td>0.082</td>
<td>.000</td>
<td>0.196</td>
</tr>
<tr>
<td>*Unstructured Leisure Activities</td>
<td>0.017</td>
<td>0.015</td>
<td>.235</td>
<td>0.028</td>
</tr>
<tr>
<td>*Peer Victimization</td>
<td>0.065</td>
<td>0.015</td>
<td>.000</td>
<td>-0.023</td>
</tr>
<tr>
<td>**Age</td>
<td>0.000</td>
<td>0.000</td>
<td>.535</td>
<td>0.000</td>
</tr>
<tr>
<td>**Sex</td>
<td>0.026</td>
<td>0.006</td>
<td>.002</td>
<td>0.021</td>
</tr>
<tr>
<td>**Migration Background</td>
<td>0.063</td>
<td>0.029</td>
<td>.030</td>
<td>-0.020</td>
</tr>
<tr>
<td>**Parental Education Level</td>
<td>-0.011</td>
<td>0.006</td>
<td>.076</td>
<td>-0.001</td>
</tr>
<tr>
<td>**SES (ISEI)</td>
<td>0.001</td>
<td>0.001</td>
<td>.523</td>
<td>-0.001</td>
</tr>
<tr>
<td>Time Dummy</td>
<td>-0.765</td>
<td>1.382</td>
<td>.580</td>
<td>-0.524</td>
</tr>
<tr>
<td>Intercept</td>
<td>0.125</td>
<td>0.070</td>
<td>.075</td>
<td>0.420</td>
</tr>
</tbody>
</table>
FIGURE 1

Generic path model of the causal relationship between moral neutralization (MN) and aggression (AGGR) from age 11 to 13.