Are Moral Disengagement, Neutralization Techniques, and Self-Serving Cognitive Distortions the Same? Developing a Unified Scale of Moral Neutralization of Aggression

Journal Article

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

Publication date:
2010

Permanent link:
https://doi.org/10.3929/ethz-b-000059338

Rights / license:
Creative Commons Attribution-NoDerivatives 4.0 International

Originally published in:
International Journal of Conflict and Violence 4(2), https://doi.org/10.4119/ijcv-2833

International Journal
of Conflict and Violence

urn:nbn:de:0070-ijcv-20102101
IJCV: Vol. 4 (2) 2010, pp. 298 – 315

Are Moral Disengagement, Neutralization Techniques, and Self-Serving Cognitive Distortions the Same? Developing a Unified Scale of Moral Neutralization of Aggression

Denis Ribeaud, Institute of Education Science, University of Zurich, Switzerland
Manuel Eisner, Institute of Criminology, University of Cambridge, UK

Vol. 4 (2) 2010

Focus:
Prejudices and Intergroup Differentiation – Within and Between Cultures

Open Section

ISSN: 1864-1385

Are Moral Disengagement, Neutralization Techniques, and Self-Serving Cognitive Distortions the Same? Developing a Unified Scale of Moral Neutralization of Aggression

Denis Ribeaud, Institute of Education Science, University of Zurich, Switzerland
Manuel Eisner, Institute of Criminology, University of Cambridge, UK

Can the three concepts of *Neutralization Techniques*, *Moral Disengagement*, and *Secondary Self-Serving Cognitive Distortions* be conceived theoretically and empirically as capturing the same cognitive processes and thus be measured with one single scale of *Moral Neutralization*? First, we show how the different approaches overlap conceptually. Second, in Study 1, we verify that four scales derived from the three conceptions of *Moral Neutralization* are correlated in such a way that they can be conceived as measuring the same phenomenon. Third, building on the results of Study 1, we derive a unified scale of *Moral Neutralization* which specifically focuses on the neutralization of aggression and test it in a large general population sample of preadolescents (Study 2). Confirmatory factor analyses suggest a good internal consistency and acceptable cross-gender factorial invariance. Correlation analyses with related behavioral and cognitive constructs corroborate the scale's criterion and convergent validity. In the final section we present a possible integration of *Moral Neutralization* in a broader framework of crime causation.

In the past decade the concept of moral disengagement has received increased attention, notably in the field of child and youth development (Hyde, Shaw, and Moilanen 2010; Hymel, Rocke-Henderson, and Bonanno 2005; Paciello et al. 2008). In particular, moral disengagement has been examined as a possible predictor of aggression and delinquency and turns out to be consistently associated with both (Bandura, Barbaranelli, and Caprara 1996; Pelton et al. 2004). Alongside moral disengagement, which was developed relatively recently within the framework of social learning theory (Bandura, 1986; Bandura et al., 1996), other similar concepts were introduced independently in related fields of research. Both the criminological theory of neutralization techniques, formulated back in 1957 by Sykes and Matza, and the notion of self-serving cognitive distortions introduced by Gibbs and colleagues (e.g. Barriga and Gibbs, 1996; Gibbs, Potter, and Goldstein, 1995) appear to describe cognitive processes that are comparable to moral disengagement. These processes assist to self-justify acts that are in conflict with a person's moral beliefs and self-concept and are thus key mechanisms for understanding aggressive and more generally deviant behavior of subjects that view themselves as generally rule-abiding and complying with common moral standards.

Demonstrating conceptual and empirical convergence among concepts developed in related fields of research serves to eliminate unnecessary duplication and to reduce complexity by unifying concepts and terminology. The present research has three interrelated aims in that direction: First, to investigate whether moral disengagement, neutralization techniques, and self-serving cognitive distortions conceptually and empirically capture the same cognitive processes. Should this be the case, the second aim is to develop a unified measure suited for preadolescents and youth that builds on all three concepts and specifically focuses on the neutralization of aggression and violence, and to examine this measure's scale reliability and validity. The third aim is to explore to what extent the new unified concept – labeled *moral neutralization* – can be integrated into a broader framework of crime and violence causation that specifically conceives violence as moral action, i.e., Situational Action Theory (Wikström and Treiber 2009).

We begin by describing and comparing the three theoretical concepts and examining to what extent they converge *conceptually.* Then we review four selected scales derived from moral disengagement, neutralization techniques, and self-serving cognitive distortions and test whether they intersect *empirically* in such a way that they can be regarded as essentially measuring the same. For that purpose we use data from a sample of preadolescents surveyed to pilot and refine a moral neutralization questionnaire in German (Study 1). Next, on the basis of these data, we construct a composite scale derived from the four scales. Finally, in Study 2, we examine the reliability and validity of the scale developed in Study 1 using a large sample of 11-year olds within the prospective longitudinal study *z-proso* (Eisner, Malti, and Ribeaud forthcoming; Eisner and Ribeaud 2005). Validity tests include correlations with well-established behavioral and cognitive outcomes in the domain of aggression and antisocial behavior and also with constructs related to core propositions of Situational Action Theory (Wikström and Treiber 2009).

1. Conceptual Convergence of Neutralization Techniques, Moral Disengagement, and Secondary Self-Serving Biases?

In essence, the three concepts of neutralization techniques, moral disengagement, and self-serving cognitive distortions, which are, in the following, generically grouped under the term moral neutralization, address the same key theoretical question: Through which cognitive processes can an individual who is generally rule-abiding and compliant with moral standards minimize cognitive dissonance, threats to self-concept, and experiences of moral self-sanction when he or she transgresses those standards?

The first authors who tried to answer this question were two American sociologists, Sykes and Matza (1957). Their theoretical effort was driven by their disagreement with Cohen's subculture theory (1955), which understands delinquency as a working-class youth reaction to perceived deprivation. Sykes and Matza's starting point was the simple observation that many delinquents have a middle-class background and moral beliefs as well as basic normative orientations no different to those of non-delinquents. This led them to seek the cognitive processes necessary to overcome the incongruence between internalized norms and beliefs and delinquent behavior. Such processes are viewed as *preceding* a particular delinquent act (Sykes and Matza 1957, 666) and are therefore conceived as being proximally involved in the causation of crime and violence. These processes correspond to the five techniques of neutralization (Table 1):

Denial of responsibility denotes a technique by which "the delinquent can define himself as lacking responsibility for his deviant actions" (667), i.e., the delinquent externalizes the locus of control. For example, a violent interaction might be framed as an accident, as provoked by the victim, or as the product of peer pressure.

Through *denial of injury* perpetrators rationalize the consequences of their acts as not really harmful to the victim. For example, the psychological consequences of verbal bullying might be discounted.

Denial of the victim occurs when "the delinquent accepts the responsibility for his deviant actions and is willing to admit that his actions involve injury" (668). Here, the role of the victim is redefined, for example conceiving the victim as a wrongdoer who deserved a lesson.

Condemnation of the condemners involves shifting attention from the delinquent act to the motives and behavior of those who disapprove such acts (668), for example, portraying authorities as hypocritical or corrupt.

Finally, Sykes and Matza describe the *appeal to higher loyalties* as follows. "Fifth and last, internal and external social controls may be neutralized by sacrificing the demands of the larger society for the demands of the smaller social groups to which the delinquent belongs such as the sibling pair, the gang, or the friendship clique" (669).

More than three decades after the first formulation of a moral neutralization framework by Sykes and Matza "no less a figure than Albert Bandura … developed an important cognitive theory of 'moral disengagement'"(Maruna and Copes 2005, 6). Like Sykes and Matza, Bandura starts from the observation that "people do not ordinarily engage in reprehensible conduct until they have justified to themselves the rightness of their actions" (Bandura et al. 1996, 365), stressing that mechanisms of moral disengagement *precede* immoral acts, and are thus involved in their immediate causation.

Table 1: Overview of concepts of moral neutralization

Cognitive Mechanism	Neutralization Techniques (Sykes and Matza 1957)	Moral Disengagement (Bandura et al. 1996)	Secondary Self-Serving Cognitive Distortions (Barriga and Gibbs 1996)
Cognitive restructuration	· Appeal to higher loyalties · Euphemistic language (implied)	· Moral justification · Euphemistic language · Advantageous comparison	· Minimizing/mislabeling (partially overlap)
Minimizing own agency	· Denial of responsibility	· Displacement of responsibility · Diffusion of responsibility	· Blaming others (partially overlap)
Disregarding/distorting negative impact	· Denial of injury	· Disregarding consequences · Distorting consequences	· Minimizing/mislabeling
Blaming/dehumanizing the victim	· Denial of the victim	· Dehumanization · Attribution of blame	· Minimizing/mislabeling (partially overlap) · Blaming others (partially overlap) · Assuming the worst (partially overlap)
Condemnation of condemner	· Condemnation of condemner		
Assuming the worst			· Assuming the worst

Comparison of the mechanisms of moral disengagement (Bandura et al. 1996) with Sykes and Matza's categories shows a high degree of overlap (Table 1). The first set of disengagement practices labeled *cognitive restructuring* aims to reframe reprehensible conduct as socially acceptable behavior. Bandura and colleagues (1996, 365) differentiate three mechanisms of restructuration: By "*moral justification* detrimental conduct is made personally and socially acceptable by portraying it in the service of valued social or moral purposes" (365). This definition obviously encompasses the *appeal to higher loyalties* described by Sykes and Matza. The second mechanism, *euphemistic language*, is viewed as a "tool masking reprehensible activities or even conferring a respectable status upon them" (365). Although Sykes and Matza fail to mention this mechanism explicitly, euphemization is implicit in their theory. The many terms placed in quotes in their original paper suggest that neutralization is implemented through euphemization.[1] The third mechanism of cognitive restructuration consists in "exploiting *advantageous comparisons* with more reprehensible activities" (365) to neutralize injurious conduct or make it to appear of little consequence.[2]

The second set of disengagement practices encompasses techniques that aim to *displace or diffuse responsibility* for reprehensible acts. In perfect congruence with Sykes and Matza's notion of *denial of responsibility* this implies externalizing the locus of control for socially sanctioned behavior. Typically, people will "view their actions as springing from social pressures or dictates of others" (365) or group decision-making will be used as a means to cognitively diffuse personal responsibility. A third set of disengagement techniques is aimed at *disregarding or distorting the consequences* of antisocial behavior. Note the striking congruence with Sykes and Matza's notion of *denial of injury*.

The last set of disengagement practices relates to a biased perception of the victim. Bandura and colleagues (1996) mention two types of victim-related mechanisms of disengagement. *Dehumanization of the victim* "divests people of

[1] E.g., "…deviant acts are 'accidents' … Vandalism … may be defined … as 'mischief' …" (Sykes and Matza 1957, 667).

[2] Producing conceptual overlap with the mechanism of *distorting consequences* (see next paragraph).

human qualities or attributes bestial qualities to them. Once dehumanized, they are no longer viewed as persons with feelings, hopes, and concerns" (366), while "by *attribution of blame*, people view themselves as faultless victims driven to injurious conduct by forcible provocation [by the victim]" (366).[3] Obviously, these two mechanisms largely coincide with the neutralization technique of *denial of the victim*.

Overall, moral disengagement and neutralization techniques appear to be broadly congruent. The main differences are the more elaborate concept of *moral justification* compared to the narrower concept of the *appeal to higher loyalties*, the lack of a counterpart to *advantageous comparisons* in neutralization theory, and *condemnation of the condemners* in the moral disengagement framework.

The third framework of moral neutralization is rooted in the concept of cognitive distortions or thinking errors (Ellis 1962; Beck 1963) and was developed in the context of young offender rehabilitation by Gibbs and colleagues (Barriga and Gibbs 1996; Barriga et al. 2000; Gibbs et al. 1995). In contrast to Ellis's and Beck's focus on self-*debasing* distortions, Gibbs and colleagues are interested in self-*serving* distortions. They distinguish between primary and secondary distortions: "Primary cognitive distortions are self-centered attitudes, thoughts, and beliefs" (Barriga and Gibbs 1996, 334) and involve "according status to one's views, expectations, needs, rights, immediate feelings and desires to such a degree that the legitimate views, etc. of others (or even one's own long-term best interest) are scarcely considered or are disregarded altogether" (334).[4] Secondary distortions serve to support the primary distortions and "have been characterized as pre- or post-transgression rationalizations that serve to 'neutralize' conscience or guilt" (334). Like neutralization techniques and moral disengagement, Gibbs and colleagues conceive cognitive distortions as potentially preceding antisocial action. As shown below, their account of secondary cognitive distortions (Table 1) shows strong similarities with the other two moral neutralization frameworks.

Blaming others comprises "misattributing blame to outside sources, especially: another person, a group, or a momentary aberration (…); or misattributing blame for one's victimization or other misfortune to innocent others" (Barriga and Gibbs 1996, 334). This distortion overlaps with disengagement mechanisms such as *diffusion and displacement of responsibility* and *attribution of blame*.

The second type of distortion, *minimizing/mislabeling,* consists in "depicting antisocial behavior as causing no real harm, or as being acceptable or even admirable; or referring to others with a belittling or dehumanizing label" (334). Obviously, this concept shares much in common with Bandura's notions of *moral justification*, *euphemistic language*, *advantageous comparisons*, *disregarding or distorting consequences*, and *dehumanization*.

Finally, the notion of *assuming the worst*, which consists in "gratuitously attributing hostile intentions to others, considering a worst-case scenario for a social situation as if it were inevitable; or assuming that improvement is impossible in one's own or others' behavior" (334) partly overlaps with Bandura's concept of *attribution of blame*, but also extends the set of possible neutralization mechanisms.[5]

Overall, our review shows a high degree of congruence among the processes of moral neutralization described in the three frameworks of moral disengagement, neutralization techniques, and self-serving cognitive distortions, thus justifying further enquiry into the empirical overlap between measures derived from them (for a further discus-

[3] This mechanism consists in externalizing the locus of control by locating it in the victim. Accordingly, it represents a special case of displacement of responsibility. Note also that Bandura conceives the construct of hostile attribution of intent (Crick and Dodge 1994) as a possible mechanism of attribution of blame (366). The problem of conceiving hostile attribution as a mechanism of moral disengagment is discussed in the last section of the present paper.

[4] Criminologists will notice the striking similarity between the definition of primary cognitive distortions and Gottfredson and Hirschi's concept of self-control (1990, 89), and particularly with the two constituting dimensions of self-centeredness and impulsivity (i.e., a "here and now" orientation and the inability to defer gratification).

[5] The closeness of this notion to hostile attribution of intent (Crick and Dodge 1994) is not unproblematic in our view, since it tends to conflate moral rationalization with biased information processing.

sion of theoretical approaches in the field of moral neutralization, see Maruna and Copes 2005).

2. Measurement Instruments for Neutralization Techniques, Moral Disengagement, and Self-Serving Cognitive Distortions

All three moral neutralization frameworks have been empirically tested. Some instruments were designed to measure post-hoc neutralization of offences committed by research subjects (e.g., Rogers and Buffalo 1974) while others assess endorsement of neutralizations for selected scenarios of antisocial behavior (e.g., Ball 1966). Most instruments in this field, however, consist of conventional item batteries designed to capture different mechanisms of moral neutralization using Likert scales. Such instruments have the advantage that they are not limited to post-hoc justifications and thus allow offenders to be compared with non-offenders and measurements to be used to predict later offending. Given appropriate wording, these instruments are easier to understand than a scenario-based approach (e.g., Shields and Whitehall 1994) which is an important issue in studies with children.

The preselection of scales for the *z-proso* study was guided by three requirements. First, the scales of interest had to be related to one of the three moral neutralization frameworks presented above. Second, they should measure neutralization of aggressive behavior. Third, they needed to be suited for a preadolescent sample, and later a youth sample. Four scales were selected using these criteria: techniques of neutralization of violence were measured with a brief instrument used for all age groups in the Denver Youth Study (i.e., from age 7 to at least age 20) (Huizinga et al. 2003). In the following, this scale is referred to as NT1. Moral disengagement was assessed with two scales: Scale MD1 is the original 32-item scale designed by Bandura and colleagues (1996, 374) and first used in a general population sample of 10- to 15-year-old Italian adolescent (*M*=11.8) of which an abbreviated version was used by Pelton and colleagues (2004) in an African-American community sample aged between 9 and 14 years (*M*=11.4). Both versions suggest a one-dimensional factor structure of moral disengagement, i.e., the mechanisms of moral disengagement tend to come together in the same persons (Bandura et al. 1996, 367; Pelton et al. 2004, 36), and accordingly both yield high internal consistency (Cronbach's α=.82). The second moral disengagement measure (MD2) specifically examines moral disengagement related to school bullying (Hymel, Rocke-Henderson, and Bonanno 2005) and was tested in a Canadian upper and middle class sample of 8th-, 9th-, and 10th-graders. Out of 51 items, 13 were identified as indicators of the four main mechanisms of moral disengagement (Table 1, column 1, rows 1–4). Factor analysis showed a single factor of moral disengagement, again suggesting that the different mechanisms of moral disengagement tend to converge. The resulting scale yielded a Cronbach's α of .81. Self-serving cognitive distortions were measured with an adapted version of the "How I think" questionnaire (HIT). Unlike the original questionnaire by Gibbs and colleagues (2001), which also encompasses non-violent problem behavior, the adapted Dutch version (van der Velden 2008) specifically focuses on aggression and bullying among children and adolescents of both genders (*M*=11.4 years). Whereas van der Velden (2008) does not report pertinent analyses, two studies (one American by Barriga and Gibbs 1996, 339; the other Dutch by Nas, Brugman, and Koops 2008, 186) that use the original HIT scale in mixed samples of incarcerated and general population male youth (16<*M*<17 years) report strong correlations among the three secondary self-serving cognitive distortions (between .71 and .78), again suggesting a one-dimensional latent construct of moral neutralization.

Generally, measures of moral neutralization correlate with aggressive and delinquent behavior. For example, a study using a neutralization techniques scale similar to the one used in the Denver study reports correlations of r=.40** and r=.41**[6] between neutralization techniques and violence in the American National Youth Survey (Agnew 1994, 580). Bandura and colleagues (1996, 369) report correlations between .13***[7] and .56*** between moral disengagement and aggression, and .20* and .45*** for delinquency. Pelton and colleagues (2004, 36) report similar patterns in their

[6] ***p<.001; **p<.01; *p<.05; $^{\text{n.s.}}$p>.05; $^{\text{n.a.}}$not available

[7] The correlation of r=.06*** reported for teacher-rated aggression in Table 1 is erroneous and should read .13*** (personal communication from Claudio Barbarenelli, 2 July 2010).

sample while Hymel and colleagues (2005, 38) report a highly significant association between bullying and moral disengagement (F(2, 459)=69.57***). Regarding secondary self-serving cognitive distortions, Barriga and Gibbs (1996, 339) report correlations between .23** and .38*** with the Nye Short Self-Report Delinquency Questionnaire and between .43*** and .55*** with the Externalizing Scale of the Youth Self-Report. Similarly, Nas and colleagues (2008, 186) report coefficients between .20* and .29** for correlations among self-serving cognitive distortions and the Teacher Report Form and of .20$^{n.a.}$ and .37$^{n.a.}$ between self-serving cognitive distortions and the Reactive-Proactive Aggression Questionnaire.

3. Study 1: Empirical Overlap and Composite Measure

Study 1 set out to explore the empirical overlap of the different measures of moral neutralization of aggression and violence and, if possible, to derive a composite measure based on the best-fitting items of the different scales.

3.1. Participants and Data Collection

The 142 participants were recruited in seven 4th- and 5th-grade classes in middle-class suburbs near the city of Zurich. Parental consent was obtained for all participants in advance. All contacted parents and children consented to participate. The mean age of the participants was *M*=10.5 years (*SD*=0.68), 52.5 percent were male. The surveys were conducted during regular school hours. Participants were guided through the written questionnaire by two researchers. All questionnaires were completed within 45 minutes.

3.2. Measures

First, the 67 items of the four scales of interest (NT1, MD1, MD2, HIT) were screened and preselected for the goals of the study. The items retained for Study 1 are shown in Table 2. Ten items of the MD1 scale were eliminated: As suggested by Pelton and colleagues (2004), the four *euphemistic language* items were removed because they are inappropriate for children. The other items were removed either because they related to behavioral domains other than violence and aggression or because they turned out to (almost) duplicate items in other scales.[8] Three items were removed from the MD2 scale because of inverse wording or translation problems.

The HIT scale used for the present study is a Dutch adaptation of the original scale that focuses on aggression and verbal bullying (van der Velden 2008). From this 28-item scale we discarded items related to *primary* self-serving cognitive distortions and social desirability as well as five filler items. Two items in the *blaming others* subscale were removed because they presumably measure hostile attribution of intent (Crick and Dodge 1994).[9] Three other items were removed because they strongly overlapped with items from other scales or because of translation problems.

Finally, one item in the NT1 scale was deleted because it overlapped with another.

The 31 items retained from preselection were translated into German (see Table 2 for the English wordings) and used in a paper-and-pencil questionnaire in the Study 1 sample.

3.3. Analysis

Correlational and exploratory factor analyses (EFA) were used.[10] First, all items of a given scale were forced to load on one single factor (Table 2, column 7). To improve the measurement quality of the scale, items with standardized loadings above .4 were selected and their standardized scores were averaged.[11] Then the four scales were correlated with each other and factor analyzed to test the empirical overlap

[8] This implies that the correlations between the scales reported below would likely have been stronger if overlapping items had been retained. Hence, the coefficients presented in the following can be viewed as conservative estimates of the correlations that would have resulted between the full-length original scales.

[9] E.g., "People are always trying to start fights with me." Some authors even explicitly use these items as indicators of hostile attribution bias (Pornari and Woods 2010).

[10] Although confirmatory factor analyses would have been the method of choice, preliminary tests suggested that both the overall sample size and the ratio of the number of parameter estimates to the number of cases were too small to allow proper parameter estimation (see e.g., Bentler and Chou 1987; Hair et al. 2006; Jackson 2003).

[11] The criterion of .4 is somewhat stricter than the one of .3 typically recommended (Bryant and Yarnold 1995) to reduce the number of items for the final scale.

(Table 3). Finally, all preselected indicators of the four scales were forced to load on a single factor (Table 2, column 8). Only items with a loading above .4 were selected for the final integrated moral neutralization scale used in Study 2.

To prevent case deletions in the factor analyses and in the computation of the sum scores all missing values in the items were imputed using the EM imputation algorithm (SPSS 2009). The number of missing cases varied between 0 and 14 per indicator (Table 2).

3.4. Results

First, we examine the properties of each individual scale. The first factor extracted from the ten MD1 items accounts for 21.2 percent of total variance (eigenvalue 2.12). With eigenvalues of 1.54, and 1.17 respectively, the next two factors also account for a substantial share of the total variance. However, the loading structure in the three-factor solution (not shown) does not suggest meaningful factors. Since all items in the one-factor solution load positively and significantly on the single factor, the hypothesis of one-dimensionality is supported by the data. However, only four items meet the strict criterion of a loading above .4 (Table 2, Item ID 1–4) and were kept for scale construction. The resulting scale yields an internal consistency of Cronbach's α=.61 (Table 3).

Factor analysis of the ten MD2 items shows a clearer scree pattern (Cattell and Vogelmann 1977). The first factor accounts for 32.1 percent of the variance, the corresponding eigenvalue of 3.21 being much higher than the eigenvalue of the next two factors (1.18, 1.04). Moreover, all items of the scale load with at least .4 on the single-factor solution, thus clearly suggesting monodimensionality. The resulting scale yields a Cronbach's α of .76. Similarly, the first factor extracted from the HIT items accounts for 32.7 percent of total variance, and the corresponding eigenvalue of 2.62 is again much higher than the eigenvalue of the next two factors (1.12, 1.02), again evidencing a clear scree pattern. All items of this scale also load positively on the single-factor solution. One item had a loading below .4 (ID 28) and was consequently excluded. The derived 7-item scale yields a reliability of .71. Finally, the first factor extracted from the three NT1 items explains 51.4 percent of the variance (eigenvalue 1.54) while the other two factors have eigenvalues below 1 (0.85, 0.61). All three items load with at least .6 on the first factor. The derived scale yields a Cronbach's α of .52.

Table 2: Item wordings, descriptive statistics, and factor loadings in Study 1

Item wording	Generic domain	Scale	N	M	S.D.	Single-factor loading on original scale	Single-factor loading of selected items on total scale	Item ID
It is alright to fight to protect your friends.	Cog. Restruct.	MD1	138	2.51	1.03	.742	.578	1
It is alright to fight when your group's honour is threatened.	Cog. Restruct.	MD1	128	1.75	0.91	.724	.630	2
If someone acts like a jerk, it is ok to treat them badly.	Victim	MD1	141	1.52	0.75	.663	.612	3
It is unfair to blame a child who had only a small part in the harm caused by a group.	Minim. Agency	MD1	136	2.93	1.29	.410	.137	4
A kid who only suggests breaking rules should not be blamed if other kids go ahead and do it.	Minim. Agency	MD1	137	1.88	1.03	.350	--	5
If a group decides together to do something harmful it is unfair to blame any kid in the group for it.	Minim. Agency	MD1	140	3.07	1.28	.332	--	6
Insults among children do not hurt anyone.	Neg. Impact	MD1	137	1.55	0.85	.254	--	7
Teasing someone does not really hurt them.	Neg. Impact	MD1	137	1.54	0.87	.234	--	8
A kid in a gang should not be blamed for the trouble the gang causes.	Minim. Agency	MD1	135	2.84	1.14	.230	--	9
Children do not mind being teased because it shows interest in them.	Neg. Impact	MD1	135	1.54	0.84	.170	--	10
Bullying can be a good way to solve problems.	Neg. Impact	MD2	142	1.34	0.68	.702	.574	11
It's okay to join in when someone you don't like is being bullied.	Cog. Restruct.	MD2	138	1.51	0.78	.656	.599	12
Sometimes it's okay to bully other people.	Cog. Restruct.	MD2	141	1.74	0.88	.649	.622	13
Some kids get bullied because they deserve it.	Victim	MD2	136	1.92	1.02	.634	.570	14
Bullying is just a normal part of being a kid.	Cog. Restruct.	MD2	137	1.89	0.86	.556	.471	15
Some kids need to be picked on just to teach them a lesson.	Neg. Impact	MD2	139	1.65	0.83	.550	.564	16
In my group of friends, bullying is okay.	Cog. Restruct.	MD2	140	1.34	0.59	.483	.471	17
It's okay to pick on losers.	Victim	MD2	142	1.18	0.53	.482	.294	18
Most students who get bullied bring it on themselves.	Victim	MD2	138	2.02	0.86	.475	.458	19
Getting bullied helps to make people tougher.	Neg. Impact	MD2	140	1.81	1.05	.402	.427	20
You should hurt people first, before they hurt you.	Assuming Worst	HIT	138	1.58	0.93	.720	.686	21
People sometimes need to be bashed.	Cog. Restruct.	HIT	139	1.65	0.93	.706	.663	22
Sometimes you have to hurt people if you have a problem with them.	Minim. Agency	HIT	141	1.73	0.82	.667	.605	23
Only a coward would ever walk away from a fight.	Cog. Restruct.	HIT	139	1.91	1.08	.662	.592	24
It's ok to slag other people off, they slag you off too.	Assuming Worst	HIT	141	1.72	0.87	.574	.568	25
It's ok to slag other people off. It doesn't really hurt anybody.	Cog. Restruct.	HIT	142	1.38	0.72	.464	.388	26
If people don't cooperate with me, it's not my fault if someone gets hurt.	Minim. Agency	HIT	131	2.05	1.17	.404	.381	27
If you don't push people around, you will always get picked on.	Assuming Worst	HIT	136	1.71	0.84	.111	--	28
It's ok to get in a physical fight with someone if you have to stand up to protect your rights.	Cog. Restruct.	NT	137	1.82	0.94	.803	.668	29
It's ok to get in a physical fight with someone if they hit you first.	Minim. Agency	NT	137	2.08	1.04	.711	.508	30
It's ok to hurt someone if you didn't mean to or it was an accident.	Minim. Agency	NT	139	2.12	1.01	.627	.418	31

Note: Standardized factor loadings below .4 are shaded in grey. Item IDs of items omitted from the final scale are also shaded in grey.

Table 3: Correlations between different scales of moral neutralization (Study 1)

	Correlations						
	1	2	3	Factor loading	M	S.D.	Alpha
1 MD1				.79	0.00	0.68	.61
2 MD2	.51			.84	0.00	0.56	.76
3 HIT	.56	.77		.90	0.00	0.60	.71
4 NT1	.59	.53	.64	.82	0.00	0.72	.52

As Table 3 shows, the four mean scales derived from the MD1, MD2, HIT, and NT scales are strongly correlated with each other (.51***≤r≤.77***). Accordingly, factorial analysis of these mean scales suggests a one-factor solution, the first factor explaining 70.1 percent of the variance (eigenvalue 2.8) while the other three factors have eigenvalues below 0.6. Similarly, when all items constituting the four moral neutralization scales are factor-analyzed together (Table 2), a clear scree pattern emerges suggesting a one-dimensional factor structure. The first factor accounts for 23.2 percent of the total variance (eigenvalue 7.19) while all other factors have eigenvalues of 2.0 and below. All items load significantly on the first factor. Overall, these results strongly support the hypothesis that neutralization techniques, moral disengagement, and secondary self-serving cognitive distortions converge not only theoretically but also empirically.

For the final version of the instrument the number of items was reduced yet again,[12] and the item wordings were refined, unified, and simplified to better meet the needs of the study population. The resulting 18-item moral neutralization instrument was tested in a second pretest sample of 118 fourth- and fifth-graders (mean age *M*=11.4 (*SD*=0.48); 50.0 percent male). As a result of this analysis, one item with a loading below .4 was removed from the scale (ID 31). After this, only one item reflecting agency minimization remained in the scale (ID 30). This item was also omitted to further shorten and simplify the scale. The final 16-item version of the scale yields an excellent consistency of α=.87 (first pretest sample) and α=.88 (second pretest sample).

4. Study 2: Testing the Composite Scale

Study 2 assessed the internal consistency, cross-gender structural invariance, and criterion validity of the moral neutralization scale developed in Study 1 in a large sample of preadolescents and also includes correlational analyses with two constructs relevant to Situational Action Theory (Wikström and Treiber 2009) to explore possible integration of the moral neutralization concept within this broader criminological framework.

4.1. Participants and Data Collection

Data for this study were collected as part of *z-proso*, a large-scale prospective longitudinal study (Eisner and Ribeaud 2005). Participants were recruited from a stratified random sample of 56 public primary schools in the city of Zurich when they entered grade 1 in 2004. Initial recruitment involved letters to the parents in their native language (nine languages) followed by telephone appointments for personal interviews, again in the parents' native language. Parental consent for the child's participation was obtained at the beginning of the parent interview at the parent's home, as a part of the informed consent procedure (for details on sampling and recruitment see Eisner et al. 2009; Eisner and Ribeaud 2005; Eisner and Ribeaud 2007).[13] At the time of the fourth data collection wave used for the present study, a valid set of moral neutralization data was available for 1,109 participants. This corresponds to a participation rate of 66.2 percent of the gross sample and a retention rate of 81.5 percent of the wave 1 sample.[14] At wave 4, participants were aged *M*=11.33 on average (*SD*=0.37), 50.9 percent were male, 44.4 percent were from migrant families (both parents born abroad). Of the participants 87.5 percent were in fifth grade, 10.3 percent in fourth grade, and 2.2 percent in

[12] Four items with loadings below .4 on the total scale were removed (ID 4, 18, 26, 27). One item was deleted because of its difficult (German) wording (ID 23) and another because it could potentially reflect facts rather than rationalizations (ID 17).

[13] Parental consent for child participation was also obtained for an additional 8.6 percent of the raw sample from parents who refused to participate themselves. Overall, the child participation rate at wave 1 was 82.6 percent.

[14] The considerable drop in participation between wave 1 and 4 is a consequence of the legal necessity to renew parental consent for all participants at wave 4. At this time, many parents refused continuing participation of their children in the study.

another grade, in special education without specified grade, or respective data were missing. Overall, 3.1 percent of the children attended a special education class.

The surveys were conducted during regular school hours in classrooms of public schools. Participants at a given school were pooled across classes to form groups of 5 to 20 children. Participating children were guided through the written questionnaire by two or three researchers. The surveys lasted 90 minutes. The 13.8 percent of the children who had moved out of the city or who were the only project participant in their school were surveyed individually at their home.

Selected behavioral outcomes were also measured at the parent and teacher levels. Among the 1,109 cases with a valid moral neutralization measure, there were 994 cases with a completed parent questionnaire and 1,009 with a completed teacher questionnaire. Parents, usually the mothers, were surveyed at home with standardized computer-assisted face-to-face interviews which lasted about an hour. Participants were offered an incentive worth approximately €25 per interview. Since 57 percent of the parents in the gross sample belonged to migrant communities, interviews were also conducted in the most important minority languages (Albanian, English, Italian, Portuguese, Serbian/Croatian/Bosnian, Spanish, Tamil, and Turkish). Details of the multilingual survey procedure are described in Eisner and Ribeaud (2007). Teacher assessments consisted of one-page paper-and-pencil questionnaires that included questions on the child's behavior, on the child's social role in the class, and his/her academic achievement.

4.2. Measures

Moral neutralization was measured with the 16-item scale developed in Study 1. Eight items refer to mechanisms involving *cognitive restructuring*, three are related to *distortion/disregard of negative consequences*, three relate to *blaming the victim*, and two involve *assuming the worst*. As to behavioral domains, eight items relate to bullying and verbal aggression, five relate to physical aggression, and two relate to aggression in general.

Only questionnaires with a valid entry for at least 10 of the 16 items were retained for further analysis. The 59 cases with one to six missing values were imputed using the EM algorithm (SPSS 2009). Scale properties are presented in the results section.

A first set of behavioral outcomes used to assess the criterion validity of the moral neutralization instrument was measured with the *Social Behavior Questionnaire* developed by Tremblay and colleagues (1991). The *Social Behavior Questionnaire* is similar to the *Child Behavior Checklist* (Achenbach and Ruffle 2000) and is adapted from the *Preschool Behavior Questionnaire* (Behar and Stringfield 1974) and the *Prosocial Behavior Questionnaire* (Weir and Duveen 1981). For the present study we used an age-adapted written version for the child survey while parents were administered the face-to-face adult version and teachers completed an abbreviated written version (for more details see Ribeaud and Eisner 2010). All versions are based on 5-point Likert scales. The *prosociality* subscale elicits altruistic and empathic behavior (child version (C): 8 items, Cronbach's α=.79; parent version (P): 10 items, α=.83; teacher version (T): 7 items, α=.92). Moreover, the *Social Behavior Questionnaire* also differentiates between two basic types of aggression, namely, *indirect/covert aggression* (C: 3, α=.76; P: 5, α=.82; T: n.a.) and *direct/overt aggression* (C: 9, α=.76; P: 12, α=.82; T: 11, α=.93).

Further behavioral outcomes include a bullying scale covering four types of bullying (verbal, physical, exclusion, hiding/destroying property) measured at the child level (4, α=.75) and three indices of *delinquency and serious problem behavior* encompassing truancy, substance use (alcohol, tobacco, cannabis), theft, vandalism, carrying a weapon, and assault (C: 11, α=.67; P: 9, α=.37; T: 8, α=.48).

Two indicators are related to social skills. To assess *aggressive conflict resolution schemata* participants were asked what they usually do in a conflict with other children. Answers were recorded on 5-point Likert scales (4 items, α=.70). Within the same instrument we also assessed *socially competent conflict resolution schemata* (4 items, α=.65).

Finally, two indicators related to cognitive predispositions were included because of their specific relevance to Situational Action Theory (Wikström and Treiber 2009). *Low self-control* was assessed using a scale derived from Grasmick and colleagues (1993), with two items for each of the five domains of *risk-seeking*, *impulsivity*, *self-centeredness*, *preference for physical activities*, and *low frustration tolerance* (10 items, α=.74). *Intrinsic benefits* and *discounting of moral costs of offending* were measured with a scenario-based instrument assessing decision-making . Participants were presented three scenarios depicting the following situations: reacting violently to a provocation, threatening a schoolmate to get his mobile phone, and shoplifting chewing gum. For each situation respondents answered questions about the perceived internal and external (i.e., social) costs and benefits. The *intrinsic benefits* of offending were assessed by asking how good the respondents would feel in the depicted situation, with high values corresponding to feeling very good. *Discounting of moral costs* was assessed by asking respondents how bad they would find it to act as depicted, with low values indicating feeling very bad about offending. All responses were recorded on 4-point Likert scales (6 items, α=.73).

4.3. Analysis

The internal consistency of the moral neutralization measure developed in Study 1 was assessed with confirmatory factor analysis and the invariance of the factor structure tested across gender groups with AMOS 6.0 software (Arbuckle, 2005). Then convergent and divergent validity of the derived moral neutralization scale was assessed using Pearson correlations with selected behavioral and cognitive constructs.

4.4. Results

4.4.1. Internal Consistency

The moral neutralization construct's internal consistency was assessed in a one-factor structure in which all 16 items of the scale were forced to load on a single factor. This initial solution yields a near-acceptable fit of CFI=.926, RMSEA=.055 (χ^2=452.1; df=104; N=1109; p<.001). Modification indices suggested that freeing-up six covariances among error terms could significantly improve model fit (χ^2=194.0; df=6).[15] This increases the fit indices of the adapted model to CFI=.966, RMSEA=.038.

Table 4: Descriptive statistics and standardized loadings on the Moral Neutralization factor (N=1209)

					Standardized factor loadings		
Item[a]	Item wording[b]	Domain	M	S.D.	*All*	*Boys*	*Girls*
1	It is alright to fight to protect your friends.	Cog. Restruct.	2.23	0.99	.48	.44	.53
2	It is alright to beat somebody who doesn't respect your friends.	Cog. Restruct.	1.40	0.68	.69	.71	.57
12	It's okay to join in when someone you don't like is being bullied.	Cog. Restruct.	1.47	0.71	.56	.57	.49
13	Sometimes it's okay to bully other people.	Cog. Restruct.	1.56	0.75	.61	.63	.61
15	Bullying is just a normal part of being a kid.	Cog. Restruct.	1.91	0.92	.45	.45	.46
22	People sometimes need to be bashed.	Cog. Restruct.	1.56	0.84	.65	.68	.50
24	Only a coward would ever walk away from a fight.	Cog. Restruct.	1.80	1.04	.51	.51	.42
29	It's ok to get in a physical fight with someone if you have to stand up to protect your rights.	Cog. Restruct.	1.67	0.85	.60	.62	.47
11	Many problems can be solved with violence.	Neg. Impact	1.27	0.66	.46	.48	.31
16	Some kids need to be picked on just to teach them a lesson.	Neg. Impact	1.50	0.77	.68	.70	.63
20	Getting bullied helps to make people tougher.	Neg. Impact	1.77	0.89	.38	.39	.37
3	If someone acts like a jerk, it is ok to treat them badly.	Victim	1.50	0.69	.67	.69	.60
14	Some kids get bullied because they deserve it.	Victim	1.85	0.91	.58	.59	.54
19	Most students who get bullied bring it on themselves.	Victim	2.14	0.90	.38	.38	.35
21	You should hurt people first, before they hurt you.	Assum. Worst	1.51	0.81	.60	.62	.47
25	It's ok to slag other people off, they slag you off too.	Assum. Worst	1.79	0.90	.53	.53	.54

[a] see Table 2; [b] wordings may slightly differ from those in Study 1 due to refinements.

[15] The six covariances relate to items with identical keywords and/or similar meaning.

As shown in Table 4, the standardized loadings range between .38 and .67 in the full sample. Both the level of model fit and the loading structure confirm one-dimensionality. Overall, the 16-item scale of moral neutralization (*M*=1.78, *SD*=0.49) used for further analysis yields a consistency coefficient of α=.87. Tests of structural invariance (Table 5) provide limited confirmation of invariance across genders. Although standardized factor loadings are within similar ranges for boys (.38 to .71, see Table 4) and girls (.31 to .63), constraining the factor loadings to equality across genders yields a highly significant decrease in model fit (χ^2=119.5; df=16; see Table 5). The decrease is further exacerbated when error terms (χ^2=320.0; df=32) and error covariances (χ^2=345.0; df=38) are also constrained to equality. However, the less strict tests of model fit based on fit indices suggest that constraining factor loadings to equality is acceptable (CFI=.942; RMSEA=.033; see Table 5), while imposing further restrictions (equal error terms, equal error covariances) results in poor CFI values.

Table 5: Tests of factorial invariance across gender groups

Model	CFI	RMSEA	χ^2	df	χ^2/df	Diff. in χ^2	Diff. in DF	p
Unconstrained across groups	.966	.026	343.4	196	1.75			
Equal (unstandardized) loadings	.942	.033	462.9	212	2.18	119.5	16	<.001
Equal loadings and equal error terms	.900	.041	663.4	228	2.91	320.0	32	<.001
Equal loadings, equal error terms and equal error covariances	.896	.042	688.4	234	2.94	345.0	38	<.001

4.4.2. Criterion Validity

Table 6 displays the correlations between the moral neutralization scale and selected constructs for the entire sample and for both genders separately. The first row shows a marked correlation with gender (r=-.25***[16]), moral neutralization being more prevalent among boys than among girls. With one exception, all correlations with behavioral outcomes are highly significantly correlated in the expected direction in the entire sample. While prosociality is consistently and significantly negatively correlated with moral neutralization across informants (child measure (C): r=-.27***; parent (P): r=-.10***; teacher (T): r=-.15***), both direct (C: r=.59***; P: r=.10**; T: r=.27***) and indirect aggression (C: r=.46***; P: r=.04^{ns}; T: n.a.) are significantly positively associated with moral neutralization (except parent-reported indirect aggression). Moreover, self-reported bullying (r=.42***) and delinquency and problem behavior as reported by all three informant groups are also highly significantly correlated with moral neutralization (C: r=.31***; P: r=.11**; T: r=.21***). The children's behavioral self-ratings correlate much better with (self-rated) moral neutralization than the teachers' and the parents' ratings. The scale's specific focus on aggressive outcomes is reflected in stronger correlations with the aggression and bullying scales compared – for a specific type of informant – to general delinquency/problem behavior.

These results corroborate the predictive validity of the moral neutralization scales in the domain of aggressive and, more generally, antisocial behavior, the latter as a consequence of the strong association between aggressive outcomes and other forms of deviance (not shown).

Construct validity is also corroborated by the positive correlations of moral neutralization with aggressive conflict resolution schemata (r=.55***) and by the less pronounced negative correlation with competent conflict resolution schemata (r=-.22***). Finally, low self-control is strongly correlated with moral neutralization (r=.51***). Also, a favorable perception of the costs and benefits of offending is similarly highly correlated with moral neutralization (r=.48***) which likely reflects that moral neutralization affects the cost-benefit assessment of offending.

[16] ***p<.001; **p<.01; *p<.05; ^{ns}p>.05

Gender-specific results show that the correlations found for the entire sample can, by and large, be reproduced in both genders, so it would not appear that moral neutralization mediates gender-effects. These results also suggest that moral neutralization is similarly correlated with behavioral and cognitive outcomes in girls and in boys, providing further corroboration of the construct validity of the moral neutralization scale.

Table 6: Correlations of moral neutralization with selected constructs

	All	Boys	Girls
Gender (1=male; 2=female)	-.248***	--	--
Prosociality (child)	-.269***	-.243***	-.156***
Prosociality (parent)	-.100***	-.106*	.029
Prosociality (teacher)	-.149***	-.135**	-.005
Direct/overt aggression (child)	.585***	.603***	.465***
Direct/overt aggression (parent)	.097**	.075	.049
Direct/overt aggression (teacher)	.268***	.264***	.162***
Indirect/covert aggression (child)	.457***	.459***	.411***
Indirect/covert aggression (parent)	.038	.088*	.029
Bullying (child)	.417***	.382***	.380***
Delinquency and problem behavior (child)	.314***	.290***	.239***
Delinquency and problem behavior (parent)	.108***	.105*	.022
Delinquency and problem behavior (teacher)	.209***	.190***	.175***
Aggressive conflict resolution strategies (child)	-.550***	-.557***	-.440***
Competent conflict resolution strategies (child)	-.223***	-.221***	-.187***
Low self-control (child)	.514***	.524***	.453***
Intrinsic benefits and discounting of moral costs (child)	.475***	.475***	.357***
	994≤N≤1109	505≤n≤564	483≤n≤545

***p<.001; **p<.01; *p<.05

5. Discussion and Conclusions

Our research confirms that the three concepts of *Neutralization Techniques* (Sykes and Matza 1957), *Moral Disengagement* (Bandura et al. 1996), and secondary *Self-Serving Cognitive Distortions* (Barriga and Gibbs 1996) essentially capture the same cognitive processes. A conceptual review broadly supports the convergence hypothesis by demonstrating that the three approaches identify (under different labels) *cognitive restructuration*, *minimizing own agency*, *disregarding/distorting negative impact*, and *blaming/dehumanizing the victim* as the four key mechanisms forming a cluster of cognitive processes serving to cognitively overcome dissonance between individual moral standards and behavioral transgressions.[17] This set of processes, labeled moral neutralization in the present study, is important for individuals to maintain their moral self-concept without experiencing moral self-sanctions, and thus allowing transgressions of moral norms at reduced psychological costs. Importantly, all three approaches identify these processes as *preceding* specific antisocial actions and thus conceive moral neutralization as facilitating such actions. So all three approaches conceive moral neutralization as a factor in the (proximal) causation of antisocial action.

Factor analyses of 31 items derived from a selection of moral neutralization measures tested in a small-scale study (Study 1) corroborate empirical convergence of the different formulations of moral neutralization and confirm the finding from previous research (e.g., Bandura et al. 1996) that the key mechanisms of moral neutralization tend to appear together in the same persons.

[17] The self-serving cognitive distortions approach additionally identifies the mechanism of *assuming the worst* which is partly related to *attribution of blame* but is more general in assuming negative outcomes as legitimation for the transgression of moral rules.

The 16-item scale of moral neutralization focusing on neutralization of aggression and bullying constructed in Study 1 was found to be internally consistent, invariant across genders and valid when tested in a large sample of 11-year olds (Study 2). Confirming previous research we found a higher prevalence of moral neutralization among boys (Bandura et al. 1996; van der Velden 2008) and marked positive correlations with aggressive, violent, and delinquent behavior (Agnew 1994; Bandura 1996; Barriga and Gibbs 1996; Hymel et al. 2005; Nas et al. 2008; Pelton et al. 2004). Conversely, moral neutralization was confirmed to be negatively correlated with prosocial behavior (Bandura et al. 1996). These correlations remained fairly stable across genders, suggesting that the scale has the same predictive power in both gender groups. Concerning the sources of information about behavioral outcomes, the children's self-ratings were much better correlated with (self-assessed) moral neutralization than the teachers' and parents' ratings. This finding is in line with validation studies of moral disengagement which also find higher correlations for the children's self-assessments (Bandura et al. 1996, 369; Pelton et al. 2004, 36). The scale's criterion validity was further corroborated by its marked correlation with conflict resolution strategies, which is also found for each gender separately and which confirms earlier findings on a linkage between moral disengagement and social competence (Pelton et al. 2004, 36).

5.1. Theoretical Outlook

Our conceptual and empirical analyses suggest that moral disengagement, neutralization techniques, and (secondary) self-serving cognitive distortions describe the very same cognitive processes and that these processes tend to cluster within the same persons. For the sake of scientific parsimony it seems justified to subsume these processes under the single label of moral neutralization and to derive a single scale informed by all the original conceptualizations.

From this unifying point, theoretical criminology needs to integrate the concept into a broader theoretical frame. As suggested by Maruna and Copes (2005) it makes little sense to construct an etiology of deviance or aggression on the sole basis of neutralization techniques (or, correspondingly, moral neutralization).[18] Because of its understanding of crime and violence as *moral* action and its focus on the most *proximal* mechanisms of crime/violence causation, Situational Action Theory (Wikström 2004; Wikström and Treiber 2009) offers a promising framework to integrate the concept of moral neutralization. Wikström and Treiber posit that acts of crime and violence are the product of an interaction between situational characteristics (temptations, provocations, moral context[19]) and individual decision making, viewing individual decision-making as largely determined by an individual's morality and ability to exercise self-control. In a given situation of temptation or provocation with a given moral context, acts of violence are expected 1) when an individual has not internalized the moral rules relevant in the corresponding situation so that acting violently is viewed as a legitimate option or 2) when an individual is unable to exercise self-control when confronted with temptation or provocation and hence unable to act in accordance with his or her moral beliefs.

Within this framework the concept of moral neutralization is useful for understanding another mechanism that *facilitates* violent or, more generally, immoral action. Specifically, we posit that an individual able to cognitively neutralize the incongruence between his or her moral beliefs and acts that conflict with those beliefs is also more likely to engage in immoral action. In other words, moral neutralization allows internalized moral rules to be temporarily discarded and makes them appear irrelevant in specific situations.[20] It is expected that such a mechanism will substantially lower the psychological costs of violence and thus also lower the individual pressure to exercise self-control. This view is also

[18] Maruna and Copes (2005) suggest a theoretical integration that differs substantially from what we propose. In essence, they conceive neutralization techniques/moral neutralization as *post*-transgression mechanisms that are important for understanding persistence of or desistance from criminal behavior. Although we agree on the relevance of such mechanisms, we believe that moral neutralization is also important in the immediate *pre*-transgression phase. In line with Bandura, our starting point is that "people do not ordinarily engage in reprehensible conduct until they have justified to themselves the rightness of their actions" (Bandura et al. 1996, 365).

[19] "A moral context is defined as the action-relevant moral rules that apply to a setting and their level of enforcement" (Wikström and Treiber 2009, 91).

[20] With regard to neutralization techniques, Agnew (1994, 567–568) supplies valuable evidence in support of this hypothesis.

in line with the concept of Gibbs and colleagues (1995) that secondary self-serving cognitive distortions (or, more generally, moral neutralization) are "pre- or post-transgression rationalizations [that] reduce the stresses from the consequences of the primary distortions" (Barriga and Gibbs 1996, 334), where the notion of primary distortions shares much in common with Gottfredson and Hirschi's concept of self-control (1990; see footnote 4).

The strong correlations between moral neutralization and both self-control and favorable perception of the costs and benefits of offending supplies preliminary empirical support for our conception of the mechanisms linking self-control and moral neutralization in the causation of aggressive and otherwise antisocial behavior. However, further research is needed to conclusively elucidate the mechanisms connecting these three constructs in the immediate causation of violence and, more generally, immoral action. Further extensions of the theory should also encompass situational characteristics – or elements of the moral context – that are likely to trigger specific moral neutralizations (e.g., being with a group of friends is likely to trigger diffusion of responsibility).

5.2. Need for Conceptual Clarification

Our review of the different conceptualizations of moral neutralization shows that some authors fail to clearly differentiate between processes of moral neutralization and biased social information processing. In particular, we found that hostile attribution of intent (e.g. Crick and Dodge 1994) was identified as a mechanism of *blaming the other* (Bandura et al. 1996) or of *assuming the worst* (Barriga and Gibbs 1996). Other authors have already stressed the fundamental difference between biased information processing and cognitive processes related to aggression beliefs and aggression legitimation (Zelli et al. 1999). For that reason we dropped items likely to measure biased social perception rather than self-serving legitimations from our scale in the preselection procedure. Future research should better take into account such delimitation problems to increase the conceptual clarity and, consequently, the discriminant validity of corresponding measurements.[21]

5.3. Limitations and Future Directions for Research

The moral neutralization scale presented in this article suffers from several limitations. First, unlike most other scales reviewed above, our moral neutralization scale focuses specifically on the neutralization of aggression and violence rather than on a broader range of antisocial and/or immoral behaviors, and its predictive scope is accordingly narrower than that of more general scales. Second, the findings are limited to a general population of preadolescents. Results from younger and older age groups and from high-risk populations are needed for a fuller assessment of the scale's properties. Third, given the cross-sectional nature of our data, the direction of the relationship between moral neutralization, aggression, and other proximal factors involved in the causation and perpetuation of aggression is not clear. From a theoretical point of view experimental and longitudinal research aimed precisely at unraveling pre- and post-transgression mechanisms involving moral neutralization would be highly desirable.

Finally, our review of different scales in the field of moral neutralization showed that they were validated with samples of very different ages, in a range between 10 and 20 years. However, in most studies the age of the participants and their level of moral development are not an issue. Hence, both theory and research would likely benefit to focus on the emergence and consequent development of moral neutralization patterns in the life course[22] and to link these patterns with other relevant developmental processes, such as moral development, the emergence and consolidation of self-control and, of course, with trajectories of aggression and violence.

[21] Similar conceptual blur is also likely in other domains such as the differentiation between lack of empathy and conscious denial of injury.

[22] To our knowledge, only one study specifically focuses on the developmental precursors of moral disengagement (Hyde, Shaw, and Moilanen 2010) while another analyzes trajectories of moral disengagement (Paciello et al. 2008). However, since in this study measurement of moral disengagement started as late as age 14, the decisive stage of preadolescent development remains unexplored.

References

Achenbach, Thomas M., and Thomas M. Ruffle. 2000. The Child Behavior Checklist and Related Forms for Assessing Behavioral/Emotional Problems and Competencies. *Pediatrics in Review* 21 (1): 265–71.

Agnew, Robert. 1994. The Techniques of Neutralization and Violence. *Criminology* 32 (4): 555–80.

Arbuckle, James L. 2005. *Amos 6.0 User's Guide*. Chicago: SPSS.

Ball, Richard A. 1966. An Empirical Exploration of Neutralization Theory. *Criminology* 4 (2): 22–32.

Bandura, Albert. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, Albert, Claudio Barbaranelli, Gian Vittorio. Caprara, and Concetta Pastorelli. 1996. Mechanisms of Moral Disengagement in the Exercise of Moral Agency. *Journal of Personality and Social Psychology* 71 (2): 364–74.

Barriga, Alvaro Q., and John C. Gibbs. 1996. Measuring Cognitive Distortion in Antisocial Youth: Development and Preliminary Validation of the »How I Think« Questionnaire. *Aggressive Behavior* 22 (5): 333–43.

Barriga, Alvaro Q., Jennifer R. Landau, Bobby L. Stinson, Albert L. Liau, and John C. Gibbs. 2000. Cognitive Distortion and Problem Behaviors in Adolescents. *Criminal Justice and Behavior* 27 (1): 36–56.

Beck, Aaron. 1963. Thinking and Depression: I, Idiosyncratic Content and Cognitive Distortions. *Archives of General Psychiatry* 9: 324–33.

Behar, Lenore, and Samuel Stringfield. 1974. A Behavior Rating Scale for the Preschool Child. *Developmental Psychology* 10 (5): 601–10.

Bentler, Peter M., and Chih-Ping Chou. 1987. Practical Issues in Structural Modeling. *Sociological Methods & Research* 16 (1): 78–117.

Bryant, Fred B., and Paul R. Yarnold. 1995. Principal-Components Analysis and Exploratory and Confirmatory Factor Analysis. In *Reading and Understanding Multivariate Statistics*, ed. Laurence G. Grimm and Paul R. Yarnold, 99–136. Washington D.C.: APA.

Cattell, Raymond B., and S. Vogelmann. 1977. Comprehensive Trial of Scree and KG Criteria for Determining the Number of Factors. *Multivariate Behavioral Research* 12 (3): 289–325.

Cohen, Albert. 1955. *Delinquent Boys: The Culture of the Gang*. Glencoe, IL: Free Press.

Crick, Nicki R., and Kenneth A. Dodge. 1994. A Review and Reformulation of Social Information-Processing Mechanisms in Children's Social Adjustment. *Psychological Bulletin* 115 (1): 74–101.

Eisner, Manuel, Tina Malti, and Denis Ribeaud. Forthcoming. Large-Scale Criminological Field Experiments: The Zurich Project on the Social Development of Children. In *The Sage Handbook of Criminological Research Methods*, ed. David Gadd, Susanne Karstedt, and Steven F. Messner. London: Sage.

Eisner, Manuel, Joseph Murray, Denis Ribeaud, Tuba Topcuoglu, Lila Kazemian, and Sytske Besemer. 2009. The Event History Calendar as an Instrument for Longitudinal Criminological Research. *Monatsschrift für Kriminologie und Strafrechtsreform* 92 (2/3): 137–59.

Eisner, Manuel, and Denis Ribeaud. 2005. A Randomised Field Experiment to Prevent Violence: The Zurich Intervention and Prevention Project at Schools, ZIPPS. *European Journal of Crime, Criminal Law and Criminal Justice* 13 (1): 27–43.

Eisner, Manuel, and Denis Ribeaud. 2007. Conducting a Criminological Survey in a Culturally Diverse Context. *European Journal of Criminology* 4 (3): 271–98.

Ellis, Albert. 1962. *Reason and Emotion in Psychotherapy*. New York: L. Stuart.

Gibbs, John C., Alvaro Q. Barriga, and Granville Bud Potter. 2001. *The How I Think Questionnaire*. Champaign, IL: Research Press.

Gibbs, John C., Granville Bud Potter, and Arnold P. Goldstein. 1995. *The EQUIP Program: Teaching Youth to Think and Act Responsibly through a Peer-Helping Approach*. Champaign, IL: Research Press.

Gottfredson, Michael R., and Travis Hirschi. 1990. *A General Theory of Crime*. Palo Alto: Stanford University Press.

Grasmick, Harold G., Charles R. Tittle, Robert J. Bursik, Jr., and Bruce J. Arneklev. 1993. Testing the Core Empirical Implications of Gottfredson and Hirschi's General Theory of Crime. *Journal of Research in Crime and Delinquency* 30 (1): 5–29.

Hair, Joseph F., William C. Black, Barry J. Babin, Rolph E. Anderson, and Ronald L. Tatham. 2006. *Multivariate Data Analysis*, 6th ed. Upper Saddle River, NJ: Pearson Prentice Hall.

Huizinga, David, Anne Wylie Weiher, Rachele Espiritu, and Finn Esbensen. 2003. Delinquency and Crime: Some Highlights from the Denver Youth Survey. In *Taking Stock of Delinquency*, ed. Terence P. Thornberry and Marvin D. Krohn, 47–91. New York: Kluwer Academic and Plenum.

Hyde, Luke W., Daniel S. Shaw, and Kristin L. Moilanen. 2010. Developmental Precursors of Moral Disengagement and the Role of Moral Disengagement in the Development of Antisocial Behavior. *Journal of Abnormal Child Psychology* 38 (2): 197–209.

Hymel, Shelley, Natalie Rocke-Henderson, and Rina A. Bonanno. 2005. Moral Disengagement: A Framework for Understanding Bullying among Adolescents. In *Peer Victimization in Schools: An International Perspective*, ed. Oyaziwo Aluede, Adriana G. McEachern, and Maureen C. Kenny (guest eds.). *Journal of Social Sciences* special issue no. 8: 1–11.

Jackson, Dennis L. 2003. Revisiting Sample Size and Number of Parameter Estimates: Some Support for the N:q Hypothesis. *Structural Equation Modeling: A Multidisciplinary Journal* 10 (1): 128–41.

Maruna, Shadd, and Heith Copes. 2005. What Have We Learned from Five Decades of Neutralization Research? In *Crime and Justice: A Review of Research*, vol. 32, ed. Michael Tonry, 221–320. Chicago: University of Chicago Press.

Nas, Coralijn N., Daniel Brugman, and Willem Koops. 2008. Measuring Self-Serving Cognitive Distortions with the »How I Think« Questionnaire. *European Journal of Psychological Assessment* 24 (3): 181–89.

Paciello, Marinella, Roberta Fida, Carlo Tramontano, Catia Lupinetti, and Gian Vittorio Caprara. 2008. Stability and Change of Moral Disengagement and Its Impact on Aggression and Violence in Late Adolescence. *Child Development* 79 (5): 1288–1309.

Pelton, Jennifer, Mary Gound, Rex Forehand, and Gene Brody. 2004. The Moral Disengagement Scale: Extension with an American Minority Sample. *Journal of Psychopathology and Behavioral Assessment* 26 (1): 31–39.

Pornari, Chrisa D., and Jane Wood. 2010. Peer and Cyber Aggression in Secondary School Students: The Role of Moral Disengagement, Hostile Attribution Bias, and Outcome Expectancies. *Aggressive Behavior* 36 (2): 81–94.

Ribeaud, Denis, and Manuel Eisner. 2010. Risk Factors for Aggression in Preadolescence: Risk Domains, Cumulative Risk, and Gender Differences – Results from a Prospective Longitudinal Study in a Multiethnic Urban Sample. *European Journal of Criminology* 7 (6): 460–98.

Rogers, Joseph W., and M. D. Buffalo. 1974. Neutralization Techniques: Toward a Simplified Measurement Scale. *Pacific Sociological Review* 17 (3): 313–31.

Shields, Ian W., and Georga C. Whitehall. 1994. Neutralization and Delinquency among Teenagers. *Criminal Justice and Behavior* 21 (2): 223–35.

SPSS. 2009. *SPSS Missing Values 17.0*. Chicago: SPSS Inc.

Sykes, Gresham M., and David Matza. 1957. Techniques of Neutralization: A Theory of Delinquency. *American Sociological Review* 22 (6): 664–70.

Tremblay, Richard E., Rolf Loeber, Claude Gagnon, Pierre Charlebois, Serge Larivée, and Marc LeBlanc. 1991. Disruptive Boys with Stable and Unstable High Fighting Behavior Patterns During Junior Elementary School. *Journal of Abnormal Child Psychology* 19 (3): 285–300.

van der Velden, Marcella. 2008. *Morele domeinverschuiving en denkfouten bij kinderen in het zesde en achtste leerjaar van de basisschool.* Unpublished Master Thesis, University of Utrecht: Utrecht.

Weir, Kirk, and Gerard Duveen. 1981. Further Development and Validation of the Prosocial Behaviour Questionnaire for Use by Teachers. *Journal of Child Psychology and Psychiatry* 22 (4): 357–74.

Wikström, Per-Olof. 2004. Crime as Alternative: Towards a Cross-Level Situational Action Theory of Crime Causation. In *Advances in Criminological Theory*, vol. 13, *Beyond Empiricism: Institutions and Intentions in the Study of Crime*, ed. Joan McCord, 1–37. New Brunswick: Transaction.

Wikström, P. O., and K. Treiber. 2009. Violence as Situational Action. *International Journal of Conflict and Violence* 3 (1): 75–96.

Zelli, Arnaldo, Kenneth A. Dodge, John E. Lochman, and Robert D. Laird. 1999. The Distinction between Beliefs Legitimizing Aggression and Deviant Processing of Social Cues: Testing Measurement Validity and the Hypothesis that Biased Processing Mediates the Effects of Beliefs on Aggression. *Journal of Personality and Social Psychology* 77 (1): 150–66.

Denis Ribeaud
dribeaud@ife.uzh.ch

Manuel Eisner
mpe23@cam.ac.uk