Aggression from the Perspective of the Reflective-Impulsive Model: Testing Predictions Using Indirect Measures

Rainer Banse

University of Bonn

Alexander F. Schmidt

University of Luxembourg

Roland Imhoff

University of Cologne

Banse, R., Schmidt, A. F., & Imhoff, R. (in press). Aggression in the perspective of the Reflective-Impulsive Model: Testing predictions using indirect measures. In R. Deutsch, B. Gawronski, & W. Hofmann (Eds.). *Reflective and impulsive determinants of human behavior*. New York, NY: Routledge Psychology Press.

Although aggressive behavior is socially highly undesirable, it is prevalent in contemporary societies as well as human history and media portrayals of human interactions. The insight that lashing out aggressively does not only have severe negative consequences for the target of this behavior but also for the actor likely leads to the valid conclusion that such behavior is often executed against the actor's own self-interest. Searching for an explanation for this puzzling fact has led several scholars to assume a typical self-control conflict between impulsive aggression and an inner antagonist whose job it is to stifle these aggressive impulses (e.g., DeWall, Baumeister, Stillman, & Gailliot, 2007). In this view, impulses are aggressive and need to be controlled by rational and non-aggressive instances reinforcing socially acceptable behavior. In the present chapter we aim to argue that the Reflective-Impulsive Model (RIM) by Fritz Strack and Roland Deutsch (2004) is helpful in overcoming this simplified and mechanistic model by incorporating the idea of individual differences on automatic and controlled levels of information processing that will determine how which individual will behave under which circumstances.

The RIM is an attempt to provide a comprehensive explanatory framework for the prediction of social behavior. Although this work is building on a history of dual-system theories comprising Freud's psychodynamic account (e.g., 1915/1960), Shiffrin and Schneider's (1977) distinction between automatic and controlled information processing, Epstein's (1994) theory of personality, Sloman's (1996) theory of reasoning, as well as Smith & DeCoster's (2000) dual-process model, Strack and Deutsch's endeavor to integrate more than a hundred years of theory and empirical research into a unified model of social behavior is ambitious, if not daring. The RIM has encountered a lot of praise, and, as to be expected for a theory making strong claims, also a lot of criticism (e.g., Keren & Schul, 2009). In the following we will investigate to what extent the RIM has already contributed or could contribute to a better understanding of aggressive behavior and compare it with another prominent model of aggression.

At about the same time when the RIM was published, Anderson and Bushman (2002) proposed a General Aggression Model (GAM) with the intention to integrate a number of domain-specific aggression theories into one comprehensive framework. In a similar vein as the RIM, the GAM tries to integrate cognitive, emotional, and motivational processes to explain the occurrence of aggressive behavior. Both models are similar in many ways; in particular they both emphasize knowledge structures and cognitive processes as major determinants of (aggressive) behavior. Whereas the GAM is formulated in the tradition of appraisal theories of stress and emotion (e.g., Lazarus, 1991), the RIM stands in the more recent tradition of social cognition research that is based on generalized dual-process models distinguishing between associative and rule-based information processing (Payne & Gawronski, 2010). Besides these more superficial aspects of form, the critical questions are whether different predictions can be derived from both models, and whether they differ in their explanatory power.

Explaining Aggression as a Special Case of Social Behavior: Reflective-Impulsive Model vs. General Aggression Model

Drawing on Cognitive Neoassociation Theory (Berkowitz, 1990) and other social information processing accounts of aggression (e.g., Crick & Dodge, 1994; Huesmann, 1988; 1998), Anderson and Bushman (2002) have proposed the GAM that aims to integrate specific or limited “mini-theories” of aggression into a unified theory. In a nutshell, the GAM postulates that aggressive behavior is the result of an immediate automatic appraisal of social situations that elicits specific cognitions, affective reactions, and arousal levels according to the momentarily activated perceptual schemata, associative knowledge structures, and motivations. If this result of the immediate appraisal is perceived to be either satisfying or unimportant, the individual will implement impulsive action that may be aggressive or not depending on the automatically activated behavior scripts. If the result of this immediate appraisal is perceived to be unsatisfactory and there are sufficient resources (e.g., cognitive capacity and time), a more complex re-appraisal will take place that may result in the implementation of more thoughtful

action. Unfortunately, Anderson and Bushman (2002) do not explain which kind of *homunculus* decides whether the outcome of a primary appraisal is satisfactory or not.

The information processing account of the GAM integrates the mainly situation-focused perspective of social psychology and the mainly person-focused perspective of personality psychology. For example, situational cues known to elicit aggression (such as weapons or provocations) can be accounted for by considering them as primes that activate cognitive schemas related to anger, arousal, hostile thoughts, and eventually aggressive behavior in the associative network of the knowledge structure. The individual difference perspective can be accounted for by postulating that repeated exposure to violent situations (e.g., parental abuse, or media violence) as well as repeated practice of violent behavior (e.g., playing violent video games, involvement in fights) will result in a chronically high availability and strong interconnectedness of hostile thoughts, schemata, and behavior scripts, which in turn facilitate the activation and implementation of aggressive behavior scripts. In other words, in the GAM actual aggressive behavior is conceptualized as the result of a momentary activation of aggression-related thoughts, emotions, and behaviors; trait aggressiveness is conceptualized as the degree of connectedness and excitability of the very same aggression-related knowledge structure. In this way the GAM provides a conceptual framework that is able to integrate the results of 30 years of aggression research including experimental and survey findings.

A second important feature of the model is the distinction between two different modes of information processing. The GAM postulates an automatic, spontaneous appraisal process outside of awareness leading to impulsive action, and a more effortful, deliberate appraisal process leading to thoughtful action. Although this distinction and terminology seems to originate in Lazarus' (1991) emotion theory featuring the distinction of primary and secondary appraisal, this aspect of the GAM converges with more general dual-system accounts of human information processing that are of central importance in theories of implicit social cognition

such as Strack and Deutsch's (2004) RIM (see also Epstein, 1994; Sloman, 1996; Smith & DeCoster, 2000).

In their comprehensive model, Strack and Deutsch (2004) postulate that human behavior is controlled by two distinct systems of information processing which can both activate motor schemata as a common final pathway in the elicitation of behavior. Both systems have very different properties. The impulsive system relies on an associative network containing the implicit knowledge base of the individual. Upon a specific input by perception or imagination, activation spreads through the network according to the association strengths between the entities or nodes. This process occurs fast and automatically without allocating attention. The impulsive system is limited to very simple operations but is able to efficiently handle automatized if-then contingencies, or the regularities in the individual's transactions with the environment. The reflective system relies on the same associative knowledge base, but it can handle much more complex computations, by integrating propositional relations between concepts, truth values, and syllogistic reasoning. Contrary to the impulsive system it learns quickly, but it depends on an active allocation of attention and cognitive resources. Motivational and emotional influences on behavior are accounted for by postulating that these activate specific relevant knowledge structures and behavior scripts.

The general architecture of the RIM appears to be compatible with the GAM (Anderson & Bushman, 2002) that postulates frequently shown aggressive behaviors to result in mental representations of the self as aggressive, more accessible scripts of aggressive behavior, and more positive attitudes towards aggressive behavior. The GAM also distinguishes a more reflective or rational mode of information processing leading to instrumental aggressive behavior, and a more automatic mode leading to impulsive aggressive behavior. The differences between GAM and RIM appear to be mainly of terminological nature: to explain automatically elicited or impulsive aggression, the RIM postulates an elicitation by the impulsive mode of information processing that is not counteracted by reflective processing, and the GAM a primary

appraisal[1] process that is not overridden by secondary appraisal processes. Both models have in common that the activation of impulsive aggression can only occur if there are cognitive structures (associations, scripts, or schemas) that make the individual prone to behave aggressively. Given that the very same behavior can be impulsive or thoughtful aggression, it is difficult to investigate the elicitation of impulsive aggression.

For a long time the evidence for "impulsive" aggression research rested on studies showing that taxing psychological resources for effortful self-control as a proxy for secondary appraisal or reflective self-control increased aggressive behavior (e.g., DeWall et al., 2007). These studies, however, could only be interpreted as evidence for the proposed model under the additional assumption that on average individuals' implicit cognitive structure makes them prone to behave aggressively. Only after the emergence of reliable and valid indirect measures such as the Implicit Association Test (IAT, Greenwald, McGhee & Schwartz, 1998) there was a chance to investigate the underlying implicit cognitive structure that is assumed to be causally responsible for impulsive aggressive behavior. Although indirect measures are far from being process pure assessments of the impulsive self-concept, they provide a first access to the causes and not only the effects of automatically activated (hence impulsive) aggressive behavior (Nosek, Hawkins, & Frazier, 2011).

Can Impulsive Aggressiveness Be Assessed With the Aggressiveness-IAT?

Building on prior work (e.g., Banse & Fischer, 2002), Banse, Messer, and Fischer (2015) developed an Aggressiveness-IAT (Agg-IAT). Whereas an aggressiveness questionnaire taps into the introspectively accessible self-concept of aggressiveness, the Agg-IAT is supposed to assess implicit associations between the self and aggressive behavior. At the conceptual level, both assessment methods can be conceived of as measures of trait aggressiveness, either explicit, or implicit, respectively. However, at the empirical level it had to be shown that the Agg-IAT

[1] Notably, in order to be comparable with the RIM, primary appraisal needs to be construed as an immediate and simple associative evaluation as otherwise it would be similar to reflective information processing based on more complex evaluations of relations between cognitive content.

could be used to close the "impulsive aggression assessment gap". In a first step, it had to be demonstrated that (a) the Agg-IAT predicts aggressive behavior, (b) this prediction holds after controlling for direct, questionnaire-based measures of aggressiveness, and (c) the Agg-IAT predicts aggressive behavior under circumstances in which the impulsive system is more likely to control behavior than the reflective system. An empirical demonstration of (a) to (c) can be considered as necessary conditions for an assessment of the automatically activated self-concept of aggressiveness with the Agg-IAT. A sufficient condition would consist of an empirical demonstration that (d) the experimental manipulation of the dominance of the reflective or the impulsive system would lead to a superior prediction of aggressive behavior by a direct or indirect measure of aggressiveness, respectively. This was attempted in a study by Schmidt, Zimmermann, Banse, and Imhoff (2015) that is presented in a later section of this chapter.

Predicting Aggressive Behavior with the Aggressiveness-IAT

The Agg-IAT developed by Banse et al. (2015) is a self-concept or personality variant of the IAT using the labels SELF and OTHERS for the object dimension, and PEACEFUL and AGGRESSIVE for the attribute dimension of the IAT. Diverting from standard self-concept IATs the items used for the OTHER category were not pronouns (other, them, etc.) but occupation labels such as book-keeper, architect, educator, cook, and farmer that had been pretested to be perceived as people with an average level of trait aggressiveness. These labels had been chosen with the intention to link the OTHER category to the general population rather than to more specific sub-groups as, for example, the own in-group when responding to the IAT. For the attribute dimension, two variants were developed. An Agg-IAT$_{\text{Behavior}}$ used behavior-related stimulus words for the peaceful-aggressive dimension such as to talk, reconciliation, dialogue, compromise versus vengeance, to punch, to fight, to beat, to revenge, respectively. For the Agg-IAT$_{\text{Trait}}$ variant, trait adjectives denoting an aggressive or peaceful personality were used (e.g., amicable, conciliatory, peaceful versus abusive, aggressive, threatening).

In a series of three studies (Banse et al., 2015) it was tested whether the Agg-IATs predicted aggressive behavior. In Study 1, penalty records over a whole season of young male ice-hockey players were used to generate an aggressive behavior index (for another sample of volleyball players no penalty index could be calculated because there were too few penalties). In Study 2, a female sample of psychology students completed a modified version of the Taylor Aggression Paradigm (TAP; Taylor, 1967) in which the loudness of sounds inflicted on an ostensible opponent in a competitive game was used as a behavioral measure of aggressiveness. Finally, Study 3 used an observer aggressiveness rating of friends and acquaintances of male students as a proxy of everyday observable aggressive behavior in social situations. Although the Agg-IAT$_{Trait}$ predicted aggressive behavior in female psychology students in Study 2, it did not in ice-hockey players in Study 1. The results of the Agg-IAT$_{Behavior}$ were slightly more consistent across studies.[2] For this reason the trait-variant of the IAT was dropped in Study 3. Across Studies 1 to 3, the Agg-IAT$_{Behavior}$ correlated significantly and substantially with the aggressive behavior indices (r_s = .37, .38, and .42, respectively). Moreover, when the Agg-IAT$_{Behavior}$ was entered into a regression analysis after a standard questionnaire measure of aggressiveness in Studies 1 to 3, it accounted for an additional proportion of 15%, 11%, and 13% of the respective aggressive behavior indices.

In summary, the empirical tests confirmed the hypotheses that the Agg-IAT$_{Behavior}$ not only was consistently correlated with aggressive behavior, but in addition incrementally predicted (retrodicted in Study 1) three different types of aggressive behavior above and beyond

[2] The Agg-IAT$_{Trait}$ consists of trait adjectives such as aggressive, peaceful and synonyms. The Agg-IAT$_{Behavior}$ features nouns and verbs denoting aggressive versus peaceful behavior (e.g., hitting – discussing). The behavior-based approach may be advantageous because an aggressive individual may have developed a strong association between the concepts ME and hitting, threatening, etc. without qualifying the own behavior as aggressive. A self-concept containing the trait word “aggressive” requires some degree of critical self-reflection or social feedback, whereas an association between the self and aggressive behaviors does only require co-activation of both concepts, not critical reflection. It may therefore be that both IATs work with individuals engaging in critical self-reflection (e.g., psychology students), but the Agg-IAT$_{Behavior}$ may be more suitable for individuals that are more action-oriented (e.g., young ice hockey players). However, this speculation warrants systematic empirical testing.

an explicit measure of aggressiveness.[3] An independent study (Grumm, Hein, & Fingerle, 2011) obtained similar results with children aged 9 to 11 years, applying an idiographic variant of the Aggressiveness-IAT that used personalized items (own first name, own family name, own gender etc.) as stimuli for the self-other dimension and only verbs as stimuli for the aggressive-peaceful dimension. They also used a variant of the TAP that was particularly suitable for children. The Agg-IAT and an explicit measure of aggressiveness independently predicted aggressive behavior, the IAT accounted for 5% of the variance.

Overall the evidence is compatible with the notion that Agg-IATs using behavior-based stimuli tap into the associative knowledge structure of the impulsive system as postulated by the RIM. As mentioned above, an incremental prediction of aggressive behavior above and beyond self-reported aggressiveness is a necessary (but not sufficient) condition to show the validity of the RIM's prediction regarding impulsive aggression.

Social Desirability as a Moderator of the Relationship between the Aggressiveness-IAT and Aggressive Behavior

In the literature on implicit prejudice there is a standard moderator effect (Fazio, Jackson, Dunton, & Williams, 1995; Payne, Cheng, Govorun, & Steward, 2005) that confirms the validity of indirect measures of prejudice (Evaluative Priming, IAT, Affective Misattribution Procedure), and that the IAT taps into a more automatic process of information processing than a racism questionnaire that taps into a more controlled mode. In participants with a low motivation to control prejudiced reactions (MCPR, Fazio et al., 1995), there is a positive correlation between the implicit and explicit measure of prejudice. In participants with a high motivation to control prejudiced reactions, however, this relation is zero or even reverses to negative values.

[3] The Agg-IAT$_{Trait}$ did not correlate with the penalty index for ice-hockey players in Study 1, and accounted for 9% of the variance of aggressive behavior assessed with the TAP in Study 2.

In an analogous way we had predicted that the relation between the Agg-IAT and aggressive behavior is moderated by the impression-management facet of social desirability (SD). In terms of the RIM one would expect SD low-scorers to make little effort to control impulsive aggressive tendencies. In consequence, the amount of overt aggressive behavior should covary with the automatic activation of an aggressive self-concept. SD high-scorers, however, should be inclined to show very little socially sanctioned aggression, no matter how strong their aggressive impulses are. They may even over-control strong impulsive aggressive tendencies, rendering implicitly highly aggressive individuals particularly peaceful at the behavioral level. In consequence, in SD low-scorers the relation between the Agg-IAT and aggressive behavior should be positive, but zero or negative in SD high-scorers.

The postulated moderator effect of SD was analyzed using the data of the three behavior prediction studies reported in Banse et al. (2015). In Study 1, a coach judgment of aggressiveness was assessed for ice-hockey players and volleyball players and used as a proxy of observable aggressive behavior. The time penalty index of aggressive behavior (available only for ice-hockey players) was used as a behavioral index of aggressiveness. The two aggressiveness indicators were regressed on SD, the Agg-IAT score (both IAT-variants in separate analyses), as well as the SD x Agg-IAT cross-product terms. A significant Social desirability x $IAT_{Behavior}$ effect ($\beta = -.38$, $p = .002$), accounting for 11% of the variance, was found for the coach judgment of aggressiveness for both ice-hockey players and volleyball players (Figure 1). No moderation effect was found for the penalty index. The same series of moderated regression analyses was conducted with the Agg-IAT_{Trait}. No moderation effects emerged.

In Study 2, the two moderated regression analyses were conducted by regressing the volume scores of the TAP on the Agg-IAT, SD, and their cross product term for both Agg-IAT variants. The resulting regression slopes were similar for both IAT variants, but this time only

the interaction term for the IAT_{Trait} yielded a significant beta weight of -.34 (p = .024), and accounted for a significant 7% of the variability in aggressive behavior.

To further explore the relation between SD and aggressive behavior in the TAP, we investigated whether SD was related to the contingency between the loudness of sounds received by the alleged opponent and the loudness of sounds inflicted upon the opponent. For each individual an indicator of a tit-for-tat strategy was computed as the correlation between received volume level and chosen volume level in response. A positive correlation between such a tit-for-tat strategy and SD (r = .29, p < .05), suggested that SD high-scorers tended to mirror the behavior of the alleged opponent by changing the volume in accordance with the opponent's last move, whereas SD low-scorers behaved as if they were guided by their own level of automatic aggressiveness for choosing the volume inflicted on the opponent. Hence, SD moderated the relation between the indirect measure of aggressiveness and overt aggressive behavior.

The results on the SD moderation effect are inconsistent and far from conclusive. However, three out of ten significant results and one near miss (similar interaction configuration that did not reach statistical significance), and all significant effects in independent samples, are quite a lot if the null-hypothesis were true; by chance one would expect only one significant result in 20. We therefore maintain that these results are at least suggestive and should be followed on in future research. From the perspective of the RIM, a high SD-score may be conceived of as a strong motivation to deliberately control impulsive action if the impulse is socially undesirable. This is almost always the case with regard to impulsive aggression (except perhaps in ice hockey players in a match). Future studies should aim for more statistical power and they should systematically vary behavior indicators and the situational setting of aggression. For example, it may be that rule violations in ice-hockey leading to penalties are almost always impulsive, and leave no room for moderation effects of SD. In the theoretical framework of the RIM the SD moderation effect lends at least indirect support that impulsive aggressiveness has trait character and can be assessed with the Agg-IAT.

Even stronger support for the dual nature of the behavior determination system would require a double dissociation that has already been demonstrated in other psychology domains (e.g., shyness, see Asendorpf, Banse, & Mücke, 2002). The same behavior should result from impulsive information processing under conditions of limited motivation or ability to execute reflective control, and should result from reflective information processing under conditions in which individuals are willing and able to override their impulsive system. Such a perspective combines basic assumptions of the RIM with an individual difference perspective on both systems and a classical social psychological perspective on situational constraints.

Moderation of the Association of Direct and Indirect Measures with Aggressive Behavior

To recap, whether aggressive behavior is executed in a certain situation (e.g., after provocation) depends on the strength of activation that has been triggered by either the reflective or the impulsive system for each behavioral option (i.e., to act out or stifle aggression). In case of a synergistic activation of the same behavioral options via both information processing systems behavior prediction is a straightforward task. However, in the case of antagonistic activation when one system inhibits whereas the other disinhibits aggression prediction is more demanding: Behavior execution will depend not only on individual differences in the two information processing systems but also on their interaction with situational boundary conditions that strengthen either reflective or impulsive precursors of aggression. As outlined in the introductory section of this chapter, demands on cognitive resources are a crucial characteristic of the RIM. Whereas the reflective system is driven by slow, deliberative, and effortful processes (i.e., requiring many cognitive resources), the impulsive system operates fast, automatically, and effortlessly (i.e., requiring few cognitive resources). Hence, the domination of one system over the other can largely be framed as a function of available self-regulatory resources (Hofmann, Friese, & Strack, 2009; Perugini, Richetin, & Zogmaister, 2010; and with a specific focus on aggression Richetin & Richardson, 2008).

Over the last decade, a particularly fruitful line of research has focused on the amount of self-control efforts that are maintained in a certain situation (DeWall et al., 2007; Stucke & Baumeister, 2006; Vohs, Glass, Maddox, & Markman, 2011) or dispositionally (Hofmann, Gschwendner, Friese, Wiers, & Schmitt, 2008) as a boundary condition impacting on aggression. Particularly, ego depletion tasks (ED; Baumeister, Vohs, & Tice, 2007) that decrease self-regulatory efforts after initial exertion of self-control have been used to experimentally manipulate acute self-regulatory load. Across several studies it has consistently been shown that ED increases aggression (DeWall et al., 2007; Stucke & Baumeister, 2006; Vohs et al., 2011). From a RIM perspective, such a pattern is consistent with the idea that the impulsive system is likely to instigate aggression but is typically controlled by a reflective system aiming to comply with social and internal norms of non-violence. To a degree, this notion is compatible with a Freudian interpretation of an (impulsive) id that is inherently aggressive and has to be kept in its limits by the (reflective) ego that negotiates behavior in accordance with reality and (internalized) social norms (super-ego). Importantly, the RIM prediction goes beyond that by allowing an individual difference perspective if we accept the above mentioned Agg-IAT as a valid method to assess the automatically activated aggressiveness self-concept. Although it may very well be the case that the impulsive precursors are on average more aggressive than reflective consideration, a taxation of resources (e.g., via ED) should increase (or decrease) aggressive behavior to the extent that the impulsive system instigates (or inhibits) aggressive behavior tendencies.

In the first study to test this idea directly (Schmidt et al., 2015), we hypothesized a specific double dissociation pattern (Asendorpf et al., 2002; Perugini et al., 2010) that so far has been experimentally demonstrated only for food and alcohol consumption (Friese, Hoffmann, & Wänke, 2008). We supposed that the automatically activated aggressiveness self-concept would elicit reactive aggressive behavior *only in highly ego-depleted* individuals as opposed to aggressive precursors from the reflective system being linked to reactive aggressive behavior

only in the low-ED condition. To this end, we assessed explicit (self-report questionnaire) and implicit ($Agg\text{-}IAT_{Behavior}$) individual differences in the aggressiveness self-concept and experimentally manipulated ED levels. The same TAP from the studies described above has been used to measure behavioral reactive aggression. We conducted a moderated regression analysis regressing TAP mean volume levels after provocation on (a) the effect-coded ED manipulation (-1 for low ED, +1 for high ED), (b) *z*-standardized self-reported aggressiveness and Agg-IAT levels, as well as (c) all resulting interaction terms from these three predictors. As the three-way interaction was not significant, we could demonstrate two independent two-way interactions (alongside main effects for all predictors) corroborating the hypothesized double dissociation (Figure 3). Specifically, simple slope analyses revealed that the Agg-IAT was linked to reactive aggression in highly ego-depleted individuals ($b = .91$, $p < .01$) but not in the low-ED condition ($b = -.06$, $p > .60$). Conversely, self-reported aggressiveness was associated with TAP aggression in the lesser depleted participants ($b = .65$, $p < .05$) but not in the high-ED condition ($b = -.05$, $p > .80$) (Schmidt et al., 2015).

The reported double dissociation pattern underscores the surplus of explanatory power of the RIM for the prediction of aggressive behavior over and above acute self-regulatory impairments: Under conditions of low self-regulatory capacity not every provoked individual will lash out aggressively. Aggressive behavior is indeed dependent on the interplay of individual differences in aggressiveness on automatic and controlled levels as predicted by a dual-systems perspective. Although this model is rather complex, it poses nevertheless an elegant and powerful framework for the explanation of behavioral phenomena (as mirrored in the significant increase from 22% to 37% of explained variance; Schmidt et al., 2015). At the same time, the RIM might shed light on a vexing conundrum in psychology – the debate about how to construe the “true self” (Gawronski, 2009). One might have good reasons to extrapolate the true self from automatic behavior that is shown in situations of impaired self-control (e.g., drunkenness reveals the true persona). However, it is equally reasonable to base such a judgment

on the person's deliberate behavior in situations where self-regulatory resources are intact (e.g., engaging in effortful protest against certain political decisions). From the perspective of the RIM, both versions of the self are equally authentic, but differently valid under distinct boundary conditions. This also explains the differential predictive validities associated with direct and indirect measures tapping into reflective and impulsive behavioral precursors (Perugini et al., 2010).

Conclusions and Future Directions

The evidence reported in this chapter is overall consistent and points to the usefulness of applying the RIM to one of modern societies' most pressing problems: aggressiveness. At the same time, however, the available evidence is far from conclusive. Particularly the gold standard for a dual-process assumption, the double dissociation, has up to now only been empirically demonstrated in a single study (Schmidt et al., 2015) that does not meet the current standards of statistical power (i.e., sample size). This important first demonstration, thus, calls for systematic replications with sufficient statistical power. Robust and replicable double dissociations in predicting aggressive behavior would constitute an important example for the applicability of the RIM to aggressive behavior.

Although the presented results are compatible with the RIM, they beg the – admittedly blasphemous – question whether such a complex model is needed in order to predict the observed data pattern or whether there are theoretically more parsimonious explanations. Reassuringly, there are none (although there is a lively debate whether dual-process models outperform single-process explanations, e.g., Keren & Schul, 2009). Thinking about aggressive behavior in the RIM framework could suggest that there are more controlled forms of aggression (enervated by the reflective system) and more impulsive forms of aggression (enervated by the impulsive system). This seems to be well aligned with the notion that aggression can be either instrumental (with the goal of harming an individual as only a proximate goal in order to achieve a higher-order non-aggressive goal) or hostile (with the goal of harming an individual as an end

goal). However, this dichotomy of aggressive behavior has come under harsh attack, calling for a knowledge structure approach to aggression that does not confound behavioral dichotomies with processing dichotomies and also allows for multi-causal behavior determination (Bushman & Anderson, 2001). Applying this criticism to a RIM-inspired understanding of aggression would imply that aggressive behavior may be elicited to differing degrees by both reflective and impulsive information processing, depending on the underlying cognitive structure, motivational states, and situational influence. As an example, slapping someone's face is likely to be the result of impulsive system processing when cognitive resources are low and social desirability concerns do not play a major role (e.g., in a bar fight late at night), but likely to be determined by more controlled processing and propositional reasoning when both these factors are high (e.g., as a reaction to an insult in a formal setting to prevent face-loss).

Although simple moderation effects in which individuals behave more aggressively after ED or when they have low SD concerns are in principle compatible with a more simplistic dichotomy of aggressive impulses and rational, non-aggressive internalized norms, the double dissociation clearly requires an explanation that integrates the assumptions of substantial inter-individual variation at both of these levels. Individuals whose automatic associations do not prompt aggressive behavioral tendencies after provocation will not lash out even after the most strenuous depletion, whereas individuals whose propositional knowledge suggests aggression as a preferable behavioral strategy will be aggressive even in the absence of any disinhibiting factors and may indeed behave more peacefully after ED (if their implicit self-concept of aggressiveness suggests so).

The RIM represents an overarching model of information processing that allows meaningful predictions regarding aggressive behavior. For the case of aggression, however, very similar predictions could be made on the base of the more specific GAM. Aside from all conceptual similarities, do the two proposals indeed fall into one with merely different terminology, or is there a way to differentiate empirically between the two? One difference lies

in the specification of processes. The RIM clearly states that the impulsive system will always activate a behavioral schema, and the reflective system will provide a behavioral decision if the available resources allow for more effortful reflective processes. The GAM makes very similar predictions with the exception that the more elaborate processing (secondary appraisal) is not only contingent on the available resources but also on the outcome of an evaluation of the primary appraisal. If the primary appraisal is satisfactory, more elaborate processing will not take place. Although this difference might seem relatively modest, it points to a potential loophole in the GAM, as it is not specified which instance evaluates whether the outcome of the primary appraisal is satisfactory and on which bases. This is an important obstacle to empirical observation as one cannot disentangle the results of the primary appraisal from the second one. If aggression is shown this can always be explained by an appraisal on the first stage. Another difference between the two models concerns the order of processing. In the GAM, the primary (automatic) appraisal is finished before the secondary (controlled) appraisal begins, whereas the RIM postulates that both processes work in parallel. Assuming that the different processes could be traced with neuro-cognitive processes, it is conceivable to directly compare these two propositions.

One of the crucial open questions for future research concerns the different sources of explicit vs. implicit self-concepts of aggressiveness. The RIM may serve as a guideline here as it suggests that what we typically refer to as the explicit self-concept is the result of information processing in the reflective system, a system that is highly flexible and also malleable by the mere addition of a negation or other qualifying information. In contrast, the impulsive system is often conceptualized as based on more habitual associative connections, less susceptible to fast and dramatic changes. This would imply that implicit self-concepts of aggressiveness are more stable than their explicit counterparts and that they are also likely to be based on longer-lasting learning histories. One potential implication might be that (repeated) early experiences with own violent behavior leave a stained mark on implicit self-concepts of aggressiveness and therefore

constitute a lasting vulnerability to act out aggressively in situation with low self-control or external constraints in later life.

This dual-process framework could also inform therapeutic attempts in changing problem behavior. It is a puzzling reality of forensic contexts that some offenders persistently behave aggressively although the consequences of this behavior are blatantly negative (Van Gelder, 2013). From the perspective of self-verification theory (Swann, 2011) people strive to maintain their personal self-concepts. According to this notion, fostering prosocial change in individuals with a deeply ingrained aggressive identity is particularly difficult due to individuals' self-verification attempts through ongoing aggressive behavior. In turn, such dysfunctional self-reaffirmation likely results in an interpersonal vicious circle of aggressive behavior in which the aggressor's self-concept will be reinforced. Interventions aiming to reduce aggressiveness typically address knowledge, insight, values, and goals – cognitive structures that operate exclusively at the level of reflective processing and are regarded as crucial drivers of prosocial responsibility taking. Unfortunately, to the extent that the causes of aggression are located at the impulsive level, standard intervention methods will have limited success. The application of the RIM calls for new types of interventions that specifically target implicit knowledge structures located in the impulsive system (e.g., Wiers, Gladwin, Hofmann, Salemink, & Ridderinkhof, 2013).

Moreover, this underscores the necessity to rethink common confrontational therapeutic strategies meant to increase responsibility taking for past transgressions and decreasing external attributions (Ware & Mann, 2012; Maruna & Mann, 2006). This approach could have paradoxical side effects as repeated confrontation with one's past aggressiveness might actively foster an aggressive self-concept on the reflective and the impulsive levels of the RIM by continuously coupling the self with transgressive behavior (i.e., increasing maladaptive self-verification; Swann, 2011). Therefore, in line with findings from research on desistance factors in high-level career-criminals (Maruna, 2002), therapists should better focus their clients on

experiencing prosocial self-aspects in the here and now. Moreover, therapists should tolerate (realistic) external attributions for past transgressions that open possibilities to take responsibility for future behavior change tied to specific risk situations (Ware & Mann, 2012). Replacing “condemnation scripts” with “redemption scripts” (Maruna, 2002) is therapeutically much more conducive to behavioral change than implementing responsibility taking for past transgressions, which are only cementing antisocial self-concepts.

In conclusion, the id as a historical forerunner of the impulsive system is not inhabited by destructive forces alone, as Freud already knew. Allowing an individual difference perspective on the respective degrees that aggressive and non-aggressive dispositions have in determining information processing at an automatic level, the RIM provides a useful guideline for an in-depth exploration of sometimes converging, sometimes diverging pathways to aggressive vs. non-aggressive behaviors. The RIM provides a powerful theoretical framework for aggression research that not only integrates more specialized models such as the GAM, it also has promises for substantial scientific advance by relating the known phenomena of aggressive behavior to state-of-the-art social cognition theory (Bluemke & Teige-Mocigemba, 2015). It is now up to empirical aggression research to live up to these promises.

References

Anderson, C. A., & Bushman, B. J. (2002). Human aggression. *Annual Review of Psychology, 53*, 27-51.

Asendorpf, J. B., Banse, R. & Mücke, D. (2002). Double dissociation between implicit and explicit personality self-concept: The case of shy behavior. *Journal of Personality and Social Psychology, 83*, 380-393.

Banse, R., & Fischer, I. (2002, July). *Implicit and explicit aggressiveness and the prediction of aggressive behavior.* Poster presented at the 11th European Conference on Personality of the European Society for Personality Psychology, Jena, Germany.

Banse, R., Messer, M., & Fischer, I. (2015). Predicting aggressive behavior with the Aggressiveness-IAT. *Aggressive Behavior, 41,* 65-83.

Baumeister, R. F., Vohs, K. D., & Tice, D. M. (2007). The strength model of self-control. *Current Directions in Psychological Science, 16,* 351-355.

Berkowitz, L. (1990). On the formation and regulation of anger and aggression: A cognitive-neoassociationistic analysis. *American Psychologist, 45*, 494-503.

Bluemke, M., & Teige-Mocigemba, S. (2105). Automatic processes in aggression: Conceptual and assessment issues. *Aggressive Behavior, 41*, 44-50.

Bushman, B. J., & Anderson, C. A. (2001). Is it time to pull the plug on the hostile versus instrumental aggression dichotomy? *Psychological Review, 108,* 273-279.

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin, 115*, 74-101.

DeWall, C. N., Baumeister, R. F., Stillman, T. F., & Gailliot, M. T. (2007). Violence restrained: Effects of self-regulation and its depletion on aggression. *Journal of Experimental Social Psychology, 43,* 62-76.

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49,* 709-724.

Fazio, R. H., Jackson, J. R., Dunton, B. C., & Williams, C. J. (1995). Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline? *Journal of Personality and Social Psychology, 69,* 1013-1027.

Friese, M., Hofmann, W., & Wänke, M. (2008). When impulses take over: Moderated predictive validity of implicit and explicit attitude measures in the prediction of food choice and consumption behavior. *British Journal of Social Psychology, 19,* 285-338.

Freud, S. (1915). Das Unbewußte [The unconscious]. In S. Freud (Ed.), *Das Ich und das Es und andere metapsychologische Schriften (1960)* (pp. 73-104). Frankfurt a. M.: Fischer.

Gawronski, B. (2009). Ten frequently asked questions about implicit measures and their frequently supposed, but not entirely correct answers. *Canadian Psychology, 50,* 141-150.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. K. L. (1998). Measuring individual differences in implicit cognition: The Implicit Association Test. *Journal of Personality and Social Psychology, 74,* 1464-1480.

Grumm, M., Hein, S., & Fingerle, M. (2011). Predicting aggressive behavior in children with the help of measures of implicit and explicit aggression. *International Journal of Behavioral Development, 35,* 352-357.

Hampel, R., & Selg, S. (1998). FAF – Fragebogen zur Erfasung von Aggressivitätsfaktoren [FAF–Inventory for the assessment of aggressiveness factors]. Göttingen, Germany: Hogrefe.

Hofmann, W., Friese, M., & Strack, F. (2009). Impulse and self-control from a dual systems perspective. *Perspectives on Psychological Science, 4,* 162-176.

Hofmann, W., Gschwendner, T., Friese, M., Wiers, R. W., & Schmitt, M. (2008). Working memory capacity and self-regulatory behavior: Toward an individual differences perspective on behavior determination by automatic versus controlled processes. *Journal Personality and Social Psychology, 95,* 962-977.

Huesmann, L. R. (1988). An information-processing model for the development of aggression. *Aggressive Behavior, 14,* 13-24.

Keren, G., & Schul, Y. (2009). Two is not always better than one – A critical evaluation of two-system theories. *Perspectives on Psychological Science, 4,* 533-550.

Lazarus, R. S. (1991). *Emotion and adaptation*. New York: Oxford University Press.

Maruna, S. (2002). *Making good: How ex-convicts reform and rebuild their lives.* Washington, DC: American Psychological Association.

Maruna, S., & Mann, R. E. (2006). A fundamental attribution error? Rethinking cognitive distortions. *Legal and Criminological Psychology, 11,* 155-177.

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social cognition: From measures to mechanisms. *Trends in Cognitive Sciences, 15,* 152-159.

Payne, B. K., Cheng, C. M., Govorun, O., & Steward, B. D. (2005). An inkblot for attitudes: Affect misattribution as implicit measurement. *Journal of Personality and Social Psychology, 89,* 277-239.

Payne, B. K., & Gawronski, B. (2010). A history of implicit social cognition: Where is it coming from? Where is it now? Where is it going? In B. Gawronski & B. K. Payne (Eds.). *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 1-15). New York: Guilford Press.

Perugini, M., Richetin, J., & Zogmaister, C. (2010). Prediction of behavior. In B. Gawronski, & B. K. Payne (Eds.), *Handbook of implicit social cognition: Measurement, theory, and applications* (pp. 255-277). New York: Guilford.

Richetin, J., & Richardson, D. S. (2008). Automatic processes and individual differences in aggressive behavior. *Aggression and Violent Behavior, 13,* 423-430.

Richetin, J., Richardson, D. S., & Mason, G. D. (2010). Predictive validity of IAT aggressiveness in the context of provocation. *Social Psychology, 41,* 27-34.

Schmidt, A.F., Zimmermann, P. S., Banse, R., & Imhoff, R. (2015). Ego depletion moderates the influence of automatic and controlled precursors of reactive aggression: A double dissociation. *Social Psychology, 46,* 132-141.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, atomatic attending, and a general theory. *Psychological Review, 84,* 127-190.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3-22.

Smith, E. R., & DeCoster, J. (2000). Dual process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review, 4,* 108-131.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review, 8,* 220-247.

Stucke, T. S., & Baumeister, R. F. (2006). Ego depletion and aggressive behavior: Is the inhibition of aggression a limited resource? *European Journal of Social Psychology, 36,* 1-13.

Swann, W. B., Jr. (2011). Self-verification theory. In P. A. M. Van Lange, A. W. Kruglanksi, & E. T. Higgins (Eds.), *The handbook of theories of social psychology* (Vol. 2, pp. 23-42). London: Sage.

Taylor, S. P. (1967). Aggressive behavior and physiological arousal as a function of provocation and the tendency to inhibit aggression. *Journal of Personality,* 35, 297-310.

Van Gelder, J.-L. (2013). Beyond rational choice: The hot/cool perspective of criminal decision making. *Psychology, Crime, & Law, 19,* 745–763.

Vohs, K. D., Glass, B. D., Maddox, W. T., & Markman, A. B. (2011). Ego depletion is not just fatigue: Evidence from a total sleep deprivation experiment. *Social Psychology and Personality Science, 2,* 166-173.

Ware, J., & Mann, R. E. (2012). How should "acceptance of responsibility" be addressed in sexual offending treatment programs? *Aggression and Violent Behavior, 17,* 279-288.

Wiers, R. W., Gladwin, T. E., Hofmann, W., Salemink, E., & Ridderinkhof, K. R. (2013). Cognitive bias modification and cognitive control training in addiction and related psychopathology: Mechanisms, clinical perspectives, and ways forward. *Clinical Psychological Science, 1,* 192–212.

Figure Captions

Figure 1. Regression of the *z*-standardized coach judgment of aggressiveness on the implicit aggressiveness measure ($IAT_{Behavior}$), social desirability, and their interaction collapsed for ice-hockey players and volleyball players (unpublished data from Banse et al., 2015; Study 1).

Figure 2. Regression of the *z*-standardized aggressive behavior index (volume tuning) on the implicit aggressiveness measure (IAT_{Trait}), social desirability, and their interaction terms for female students (unpublished data from Banse et al., 2015; Study 2).

Figure 3. Regression of the *z*-standardized aggressive behavior index (volume tuning) on the explicit aggressiveness measure (Reactive Aggressiveness subscale from Hampel & Selg, 1998), the implicit aggressiveness measure ($IAT_{Behavior}$), the effect-coded ego depletion condition, and all two-way interactions (Schmidt et al., 2015).

Figure 1

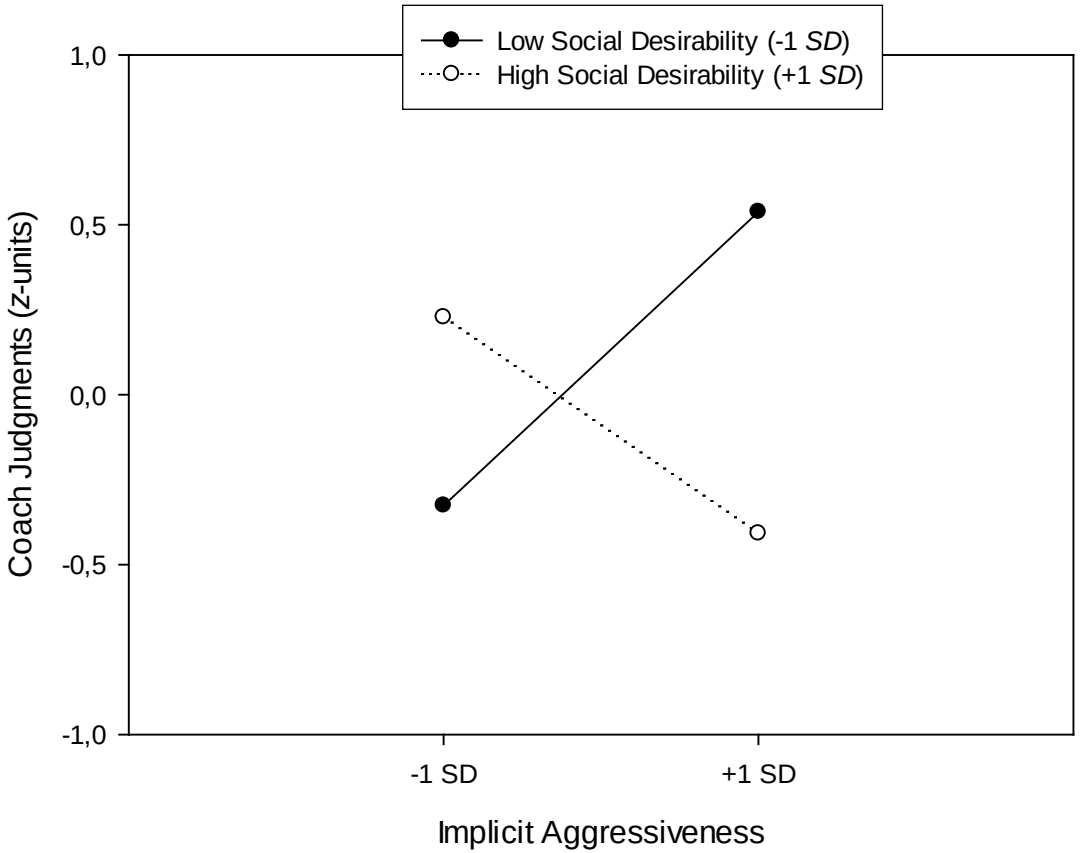

Low Social Desirability (-1 SD)
High Social Desirability (+1 SD)
1,0
0,5
0,0
-0,5
-1,0
Coach Judgments (z-units)
-1 SD
+1 SD
Implicit Aggressiveness

Figure 2

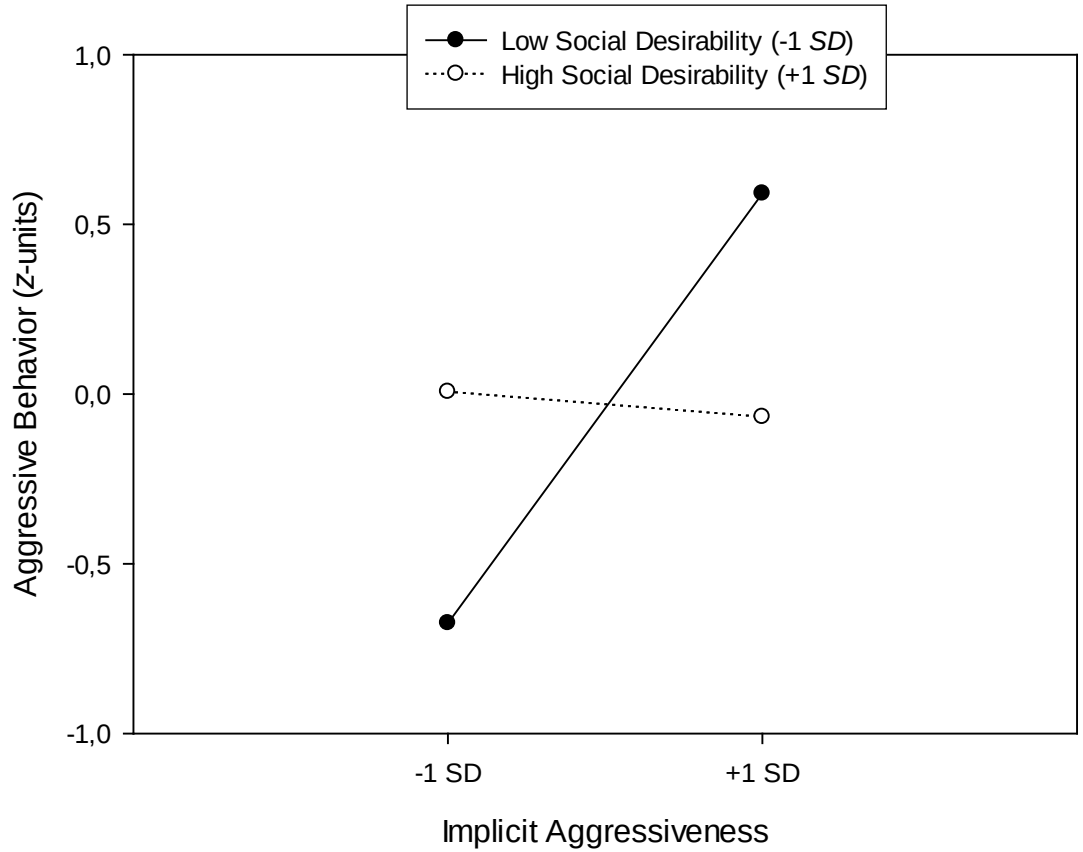
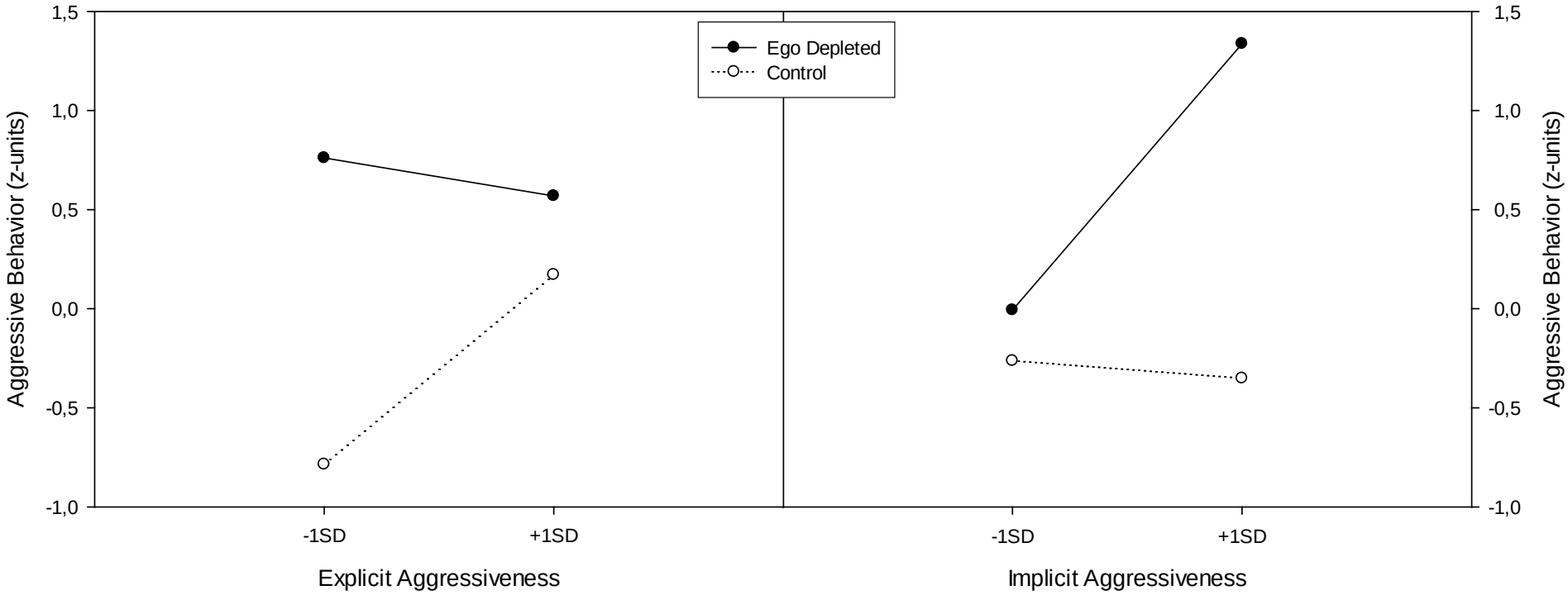

Low Social Desirability (-1 SD)
High Social Desirability (+1 SD)
1,0
0,5
0,0
-0,5
-1,0
Aggressive Behavior (z-units)
-1 SD
+1 SD
Implicit Aggressiveness

Figure 3


Ego Depleted
Control
Aggressive Behavior (z-units)
1,5
1,0
0,5
0,0
-0,5
-1,0
-1SD
+1SD
Explicit Aggressiveness
Implicit Aggressiveness