Article

The Implicit Component of Moral Disengagement: Applying the Relational Responding Task to Investigate Its Relationship With Cheating Behavior

Personality and Social Psychology Bulletin
2022, Vol. 48(1) 78–94

DOI: 10.1177/0146167220984293
journals.sagepub.com/home/pspb


R. Fida[1], V. Ghezzi[2*], M. Paciello[3*], C. Tramontano[4*], F. Dentale[2], and C. Barbaranelli[2]

Abstract

This article aims to conceptualize, for the first time, an implicit form of moral disengagement and investigate its role in relation to cheating behavior. In line with the implicit social-cognition models, we argue that the implicit moral disengagement would represent an unintentional, automatic, and less accessible form of the mechanisms bypassing the moral self-regulatory system. We anticipate that in situations implying on-the-spot decisions and where individuals might suffer no consequences for the misconduct, the implicit moral disengagement would predict the actual behavior while the explicit moral disengagement would predict self-reported conduct. The results of three empirical studies provide support for the theorization of an implicit moral disengagement and its assessment through a newly developed implicit measurement procedure using the relational responding task. Results of the structural equation models, including both implicit and explicit moral disengagement, demonstrated that only the implicit one was associated with the actual misconduct.

Keywords

moral disengagement, implicit social-cognitions, ethical behavior, morality, RRT, relational responding task

Received January 30, 2020; revision accepted December 4, 2020

Different theoretical models have been developed with the aim of identifying factors affecting individuals' moral and ethical behavior (e.g., Ellemers et al., 2019). Bandura's (1991) moral agency theory introduced moral disengagement (MD), defined as a set of social-cognitive mechanisms that temporarily silence the internal moral and normative control, allowing individuals to misbehave without abdicating and thus preserving their own moral self.

Although Bandura (1999) acknowledged the potential role of automatic processes,[1] these processes were neither discussed nor hypothesized in relation to MD. On the contrary, we believe that an implicit MD does exist and operates together with an explicit and intentional form of MD. In line with the implicit social-cognition models (see Gawronski & Payne, 2010, for a review), the implicit MD would represent an automatic, unintentional, and less accessible form of the justification mechanisms.

The conceptualization of the implicit MD is timely considering the increasing body of literature (e.g., Chugh & Kern, 2016; Reynolds et al., 2010), which suggests that to understand human moral and ethical conduct, it is insufficient to exclusively consider explicit components. Indeed, literature on misbehavior has attested that this type of conduct can be acted outside people's awareness (e.g., Bazerman & Gino, 2012; Chugh et al., 2005). Some authors suggest that bounded ethicality often involves forms of implicit self-serving biases that fall at the fringes of conscious awareness (Sezer et al., 2015).

Conversely, it is essential to also consider automatic components (e.g., Lapsley & Hill, 2008). Implicit and explicit components are not mutually exclusive, and both could explain individuals' behavior (Marquardt, 2010; Marquardt & Hoeger, 2009; Perugini & Leone, 2009). The implicit social-cognition models demonstrate that explicit processes better predict deliberative behaviors, and the implicit

[1]University of East Anglia, Norwich, UK
[2]Sapienza University of Rome, Italy
[3]Uninettuno Telematic International University, Rome, Italy
[4]Coventry University, UK

*These authors contributed equally to this work and should all be regarded as first authors.

Corresponding Author:
R. Fida, Norwich Business School, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK.
Email: r.fida@uea.ac.uk

processes better predict the more spontaneous ones (e.g., Perugini, 2005; Perugini & Leone, 2009).

The present research offers a theoretical contribution to Bandura's moral agency theory, and, more broadly, to the debate in moral psychology, by postulating an implicit component of MD. Theorizing the implicit MD does not imply a claim that individuals cannot be held responsible for their actions. Rather, we suggest that the existence of implicit MD may acknowledge the potential automaticity of these mechanisms. Acknowledging this and empirically proving its impact on misconduct would be consistent with Bandura's (2016) claim that individuals might have variable *gradients* of awareness and intentionality on their behavior and would offer a more comprehensive theorization of moral functioning.

Furthermore, this article offers a methodological contribution by presenting a newly developed valid and reliable MD implicit measure. In doing so, we capitalize on the extensive literature on implicit measures, and in particular, on the relational responding task (RRT; De Houwer et al., 2015). This also contributes to addressing the exclusive use of self-report measures in moral psychology, raised as a highly critical element in a recent literature review (Ellemers et al., 2019). By presenting an implicit measure of MD, we do not seek to undermine the value of existing self-report assessment tools. Rather, we agree with Nosek and colleagues' (2011, p. 155) claim that "neither implicit nor explicit measures have an advantage in being the 'truer' measure of one's thoughts and feelings [. . .]; both are valid assessments of unique aspects of social-cognition."

This article includes three studies describing the development and validation of an implicit MD measure (Studies 1, 2, and 3), and testing whether and how explicit and implicit MD work differently in relation to cheating behavior (Studies 2 and 3). In line with the implicit social-cognition models and previous studies on automatic processes in the moral domain (e.g., Perugini, 2005; Perugini & Leone, 2009), we expected the implicit MD to better predict the actual misconduct and the explicit MD to predict the self-reported one. Since MD measures must be tailored to one specific domain (Bandura, 2016), we focused on academic cheating behaviors, given MD is an important predictor (e.g., Fida et al., 2018).

The Unexplored Implicit Side of Moral Disengagement

Bandura highlighted that people can keep their conduct in line with their principles and systems of norms due to their self-regulatory capabilities. However, the self-regulatory moral system does not ensure behavioral consistency. Indeed, moral control could be selectively "deactivated" by MD (Bandura, 1991), allowing the self-regulatory moral system to be bypassed. Studies have consistently supported that the more individuals morally disengage the more they misbehave (e.g., Fida et al., 2018; Newman et al., 2020).

MD has been invariably conceptualized and operationalized as an explicit construct. Although Bandura (2008, p. 114) acknowledges that human action "contains both cognitively guided and automatic aspects as well as top-down and bottom-up processing," he has never theorized, operationalized, and investigated an implicit MD component. In line with the literature on social-cognition (De Houwer, 2014; Gawronski & Payne, 2010; Strack & Deutsch, 2004) and the current debate on bounded ethicality (e.g., Chugh & Kern, 2016), we believe that this component exists and needs to be assessed.

Implicit MD may capture the justification processes that might have been learnt and routinized over time (e.g., Hyde et al., 2010; Paciello et al., 2008), and this could operate automatically. It would represent not only a "footprint" left by past personal experiences of morally disengaging but also the "mark" left by the repeated exposure to specific social models and situations. If we assume that the self-system is an organized structure of knowledge in which processes can operate at both implicit and explicit levels (Payne & Gawronski, 2010), the implicit MD is the strength of the automatic self-absolving processes that are available to individuals and that might make their engagement in misconduct more likely.

Accepting the existence of an implicit MD sets a challenge in its assessment. MD has been invariably measured using self-report scales. However, "people are highly motivated to protect their self-views of being a moral person" (Ellemers et al., 2019, p. 3). Hence, when completing a self-report measure of MD, individuals may respond untruthfully in an attempt to preserve their moral image and what they want to project externally. Research on moral psychology has mostly relied on self-report measures (Ellemers et al., 2019), which can only partially assess moral processes and are unable to capture implicit ones (Chugh et al., 2005; Perugini & Leone, 2009; Sezer et al., 2015). If we recognize the impact of previous sedimented experiences and of tacit vicarious learning in relation to MD, we need to resort to implicit measures to access content that would otherwise not be captured by any self-report assessment strategies.

Implicit Social-Cognition and Morality

Over the last few decades, implicit social-cognition models have suggested the distinction between implicit and explicit processes (Gawronski & Payne, 2010). The former are described as unintentional, unaware, spontaneous, associative, and requiring little cognitive effort. It has been suggested that they are based on mental associations, namely, simple mnemonic links between specific target categories and specific attributes which can be activated without deliberative effort (e.g., Gawronski & Bodenhausen, 2006). In contrast, explicit processes are described as intentional, aware, deliberative, and propositional, and require high levels of attention. They are perceived as propositional judgments based on reflexive processes (Strack & Deutsch, 2004).

Different experimental paradigms were developed to measure automatic mental associations (De Houwer & Moors, 2010). Among these, the Implicit Association Test (IAT; Greenwald et al., 1998) is considered the most used and tested. The IAT is a computer-administered task designed to measure the strength of automatic mental associations between two opposing target concepts (e.g., Self vs. Others) and two opposing attributes (e.g., Honest vs. Dishonest). In each trial, participants are instructed to categorize a stimulus (e.g., a word or an image) as quickly and accurately as possible into the two possible target categories and two possible attributes (Greenwald et al., 1998).

Notwithstanding the important role of IAT measures in predicting actual behavior, this type of instrument can only measure the mere automatic associations between two target categories (e.g., Self vs. Others) and two attributes (e.g., Honest vs. Dishonest), without considering the possible different relationships between them (De Houwer, 2014). For instance, an automatic association between Self and Honest may be interpreted by applying different logical relationships such as "I am honest" or "I should be honest." Several authors have recently suggested that implicit evaluations depend on an automatic activation of propositions and not on automatic mental associations (e.g., De Houwer, 2014), as proposed in the classical dual models. In the classical dual models, propositional evaluations are conceived as intrinsically reflective and separated from implicit processes that are based on automatic associative activation (e.g., Strack & Deutsch, 2004). De Houwer (2014) showed that propositional evaluations can be formed (e.g., Heider et al., 2015) and retrieved automatically. Hence, to develop an implicit measure of MD, we relied on instruments that included statements as stimuli that captured implicit propositional evaluations.

Other models have suggested different explanations for the relationships of implicit versus explicit measures on spontaneous versus deliberative behaviors (e.g., Perugini, 2005). Examples include the so-called additive and double dissociation models. The former assumes that implicit and explicit measures can provide unique contributions in the prediction of both behaviors (e.g., Perugini et al., 2010). However, it has also been demonstrated that in some cases, only one measure (implicit or explicit) offers an additive unique contribution to both type of behaviors (i.e., partial dissociation model, see Perugini, 2005). The double dissociation model assumes that implicit measures are expected to predict only spontaneous behaviors, whereas explicit measures are expected to predict only deliberative ones.

To the best of our knowledge, only one research paper has investigated the role of both implicit and explicit processes in explaining moral behavior (i.e., Perugini & Leone, 2009). Specifically, they applied the IAT to measure moral self-concept and showed that while the implicit moral self-concept significantly predicted the actual moral behavior, the explicit measure did not.

The Development of an Implicit Moral Disengagement Measure

When developing the implicit MD measure, we acknowledged the limitations of using the IAT. Using, for instance, Cheating versus Not cheating as attributes and Self versus Others as targets would have resulted in the assessment of a form of self-identity, rather than MD. MD refers to the mechanisms that restructure the misconduct by focusing on those conditions that might legitimate it (e.g., cheating when everyone does it). An implicit measure of MD should assess the automatic processes associated with misconduct under certain circumstances. Moreover, as previously mentioned, the IAT is designed to capture automatic associations between concepts while ignoring the way in which those concepts are related (see Hughes et al., 2012, for a detailed overview). For instance, while both the propositional beliefs *I cheat* and *I legitimate cheating* involve a relationship between the categories *I* and *cheat*, they substantially differ in terms of the type of relationship, and IAT cannot differentiate between them.

To overcome these limitations, we used a different latency-based experimental paradigm, that is the RRT (De Houwer et al., 2015), which measures not mere automatic associations, but implicit beliefs, including the relational information between target categories and attributes (for an overview, see Barnes-Holmes et al., 2010). Propositions are different from automatic associations, as they contain relational information about how concepts are related with each other (e.g., De Houwer, 2014; Hughes & Barnes-Holmes, 2013).

One of the essential characteristics of the RRT is the requirement to respond in line with specific situational beliefs. The process involves the presentation of a series of statements in the middle of a computer screen, and participants are instructed to categorize them, as quickly as possible, *as if* they agree with certain statements and disagree with others. For instance, in the first block (see Figure 1, Panel A), participants are asked (a) to select *Right* when presented with statements that imply the moral views of the behavior (e.g., "It is serious to cheat even if no one is damaged") and (b) to select *Wrong* when presented with statements that imply the unmoral views of the behavior (e.g., "It is not serious to cheat if no one is damaged"). In the second block (see Figure 1, Panel B), they are asked to respond in the opposite manner: (a) select *Right* when presented with the unmoral statement and (b) select *Wrong* when presented with the moral statement. Reflective of the IAT scoring algorithm, the difference in the mean response latency between these two blocks of trials is assumed to provide a measure of the extent to which participants morally (dis)engage.

Similar to IAT, RRT requires to categorize stimuli as quickly and accurately as possible. As such, they capture automatic components, rather than participants' introspective processes (De Houwer & Moors, 2010). However, while RRT measures implicit propositional knowledge, IAT only

Panel A – Combined Compatible Block

RIGHT	Stimulus	WRONG
√	Licit	X
√	It is serious to cheat even if no one is damaged	X
X	Illicit	√
X	It is not serious to cheat if no one is damaged	√

Panel B – Combined Incompatible Block

RIGHT	Stimulus	WRONG
√	Licit	X
X	It is serious to cheat even if no one is damaged	√
X	Illicit	√
√	It is not serious to cheat if no one is damaged	X

Figure 1. Example of categorization of stimuli for the academic moral disengagement relational responding task (AMD-RRT).
Note. Symbols below RIGHT and WRONG columns shows the consistent (√) and inconsistent (X) response within each block. Panel A presents a block that is compatible with moral disengagement. Panel B presents a block that is incompatible with moral disengagement.

measures automatic associations. Hence, RRT and IAT measure different constructs that may or may not be reciprocally related. To date, different studies have used RRT to measure implicit beliefs, such as prejudice (De Houwer et al., 2015), parenting beliefs (Koning et al., 2016), desire to smoke (Tibboel et al., 2017), actual versus ideal body image (Heider et al., 2018), alcohol self-identity (Cummins et al., 2020), and self-esteem (Dentale et al., 2020). These studies provided evidences of RRT reliability and validity. A recent study (Dentale et al., 2020) demonstrated that, similar to the IAT, the RRT is consistently less prone to faking effects with respect to self-report measures. To the best of our knowledge, the RRT has never been used to study morality.

Study 1

The specific aim of this study was to present the development of an implicit MD measure and to initially test its psychometric properties. We focused on the academic context because MD has been recognized as an important factor predicting cheating behavior (Fida et al., 2018). This is a type of misconduct that can take different forms and is extremely pervasive and widespread (e.g., International Center for Academic Integrity, 2015; Whitley, 1998).

Method

Participants and procedure. We determined a priori the final sample size to achieve a minimum level of reliability of the implicit measures proposed in this preliminary study. Since IAT split-half coefficients generally range between .60 and .90 (see LeBel & Paunonen, 2011), we set this .60 as the minimum acceptable level of internal consistency and estimated the minimum sample size (for a two-tailed $\alpha = .05$ and with 80% of power) accordingly. Since the split-half correlation coefficient represents a special case of Cronbach's alpha (see Lord & Novick, 1968), we determined our sample size by using the formula devised by Bonett (2002, Equation 5). Results indicated that the minimum sample size was 52 participants.

Participants were psychology students. One of the authors presented the research project during a research method class. Interested students left their contact information and were later contacted by a research assistant. After signing the informed consent previously approved by the Ethical Review Board of the department to which the second author is affiliated, participants completed anonymously the two implicit measures described below. Students' participation was completely voluntary and was rewarded with course credits. Overall, 70 students (30 males and 40 females) with a mean age of 20.21 years ($SD = 2.30$) participated in the study. The final sample size was constant across different analyses. Sample size was determined before any data analysis.

Measures. The measures used in this study were collected along with others (personality traits, self-efficacy for self-regulated learning, Machiavellianism, academic citizenship behaviors) not relevant for this article. All measures, manipulations, and exclusions are reported.

Academic cheating behavior implicit association test (ACB-IAT). Participants performed both single and combined categorization tasks, using five stimuli-words for each category that were presented in a randomized order within each block of trials. The labels used in the target categorization task were as follows: Self (stimuli: me, my, I, self, mine) versus Others (stimuli: others, their, them, they, those), whereas the labels for the attribute categories were Respecting Rules (following, conforming, adhering, respecting, complying) versus Breaking Rules (deceiving, breaking, cheating, violating, tricking). The overall procedure consisted of seven blocks of trials: a single target categorization task (e.g., Self vs. Others, 20 trials), a single attribute categorization task (e.g., Respecting Rules vs. Breaking Rules, 20 trials), an initial combined categorization task (e.g., Self or Respecting Rules vs. Others or Breaking Rules; two sub-blocks of 20 and 40 trials respectively), a single target categorization task reversed (e.g., Others vs. Self, 40 trials), and a second combined categorization task (e.g., Others or Respecting Rules vs. Self or Breaking Rules; two sub-blocks of 20 and 40 trials, respectively). The order of the two combined blocks was counterbalanced across participants. Data from the combined blocks were used to compute the D scores, according to the built-in error penalty scoring procedure (Greenwald et al., 2003). More positive scores indicate a higher association between Self and Breaking Rules.

Implicit academic moral disengagement relational responding task (AMD-RRT). We developed two parallel sets of target

items drawing on previous work on academic MD (Farnese et al., 2011; Fida et al., 2018). Specifically, we first developed five statements related to MD in relation to a range of cheating behavior (e.g., copying, giving hints). Then, for each of them, we worded a corresponding statement of moral engagement (see Table 2 for the full set of items). The set of 10 attributes were developed as synonymous of "Right" (i.e., legitimate, licit, correct, acceptable, right) and "Wrong" (i.e., illegitimate, illicit, incorrect, inacceptable, wrong). The AMD-RRT consisted of seven blocks. During the first block (20 trials), participants were presented with the 10 synonymous of "Right" or "Wrong." Each of these words (hereafter referred to as "inducer words") was presented twice in a random order in an orange font. During the second block (20 trials), participants were presented with the 10 target statements (MD and moral engagement). Each of these statements was presented twice in a random order in a blue font. Participants were instructed to respond as quickly as possible to these statements in a manner that would reflect moral engagement (i.e., to judge moral engagement statements as "Right" and MD statements as "Wrong"). During the third (40 trials) and fourth (40 trials) blocks, all stimuli were presented twice, either in orange (i.e., attribute words) or in blue (i.e., target statements). Participants were asked to correctly categorize the attribute words and to respond to the target statements as if they endorsed moral engagement. During the fifth block (20 trials), each of the target statements was again presented twice in a blue font. Participants were now asked to respond to these statements in a manner consistent with MD. Finally, during the sixth (40 trials) and seventh (40 trials) blocks, all statements were again presented twice, either in orange (i.e., attribute words) or in blue (i.e., target statements). Participants were asked to correctly categorize the attribute words and to respond to the target statements as if they endorsed MD. During the administration of the AMD-RRT, the response labels "Wrong" and "Right" were presented at the top-left and top-right corner of the computer screen, respectively. All statements were presented in the middle of the computer screen until a response was registered. Incorrect responses resulted in the presentation of a red cross in the lower half of the computer screen until participants gave the appropriate response. The subsequent trial then began after an interval of 750 ms. Response latencies exceeding the cut-off value of 10,000 ms were thus excluded. Subjects with more than 10% of response latencies faster than 300 ms were deleted. The AMD-RRT data were scored using the D1 algorithm, after exclusion of all data stemming from practice and induction trials (see De Houwer et al., 2015). The final AMD-RRT scores were computed so that higher scores reflected higher levels of implicit MD.

Data analysis. We examined descriptive statistics and evaluated the reliability in terms of internal consistency (split-half Spearman–Brown corrected coefficients). We examined average latencies and error percentages of both implicit measures together with the zero-order correlation among ACB-IAT and AMD-RRT.

Results

Average latencies and error percentages are presented in Tables 1 and 2. Values observed for the ACB-IAT items were in line with those observed with the most commonly used IAT measures in the literature. Participants took approximately 5 min on average ($SD = 1.5$) to complete AMD-IAT. Although error percentages of AMD-RRT were in line with the values observed in other studies, latencies associated with its stimuli may be problematic. Indeed, their averages ranged between approximately 2,084 and 2,500 ms, much higher than those observed in the initial validation study conducted by De Houwer et al. (2015). This could have been the result of the excessive length of the AMD-RRT stimuli that has been also reported as an issue by several participants after the session.

In terms of internal consistency, split-half reliability coefficients for the ACB-IAT and the AMD-RRT were, respectively, .64 and .77. No significant association was found between implicit measures and their respective average latency and error percentage, and ABC-IAT and AMD-RRT total scores were substantially independent ($r = -.08$, $p > .05$).

Discussion

This preliminary study aimed to pilot and calibrate our implicit measures. The results showed that both measures were reliable, especially considering the level of internal consistency generally exhibited by implicit measures (Payne & Gawronski, 2010). However, AMD-RRT average latencies showed much higher values than those observed in other studies (e.g., De Houwer et al., 2015). This may indicate that, as indeed reported by several participants, sentences used as stimuli were too long, and high average latencies can be interpreted as a sign of excessive difficulty in completing the AMD-RRT. Hence, although the newly developed measure seemed quite promising, a revision of the stimuli was necessary to make them less verbose and easier to process.

Study 2

Drawing on the results of Study 1, the aim of Study 2 is twofold: (a) to revise the AMD-RRT measure by making the stimuli less verbose and complex (e.g., avoiding double negatives); (b) to examine the criterion and incremental validity of the revised AMD-RRT against its "explicit" counterpart (i.e., explicit academic MD), on both self-reported cheating behavior and the "actual" lie behavior. Consistent with previous literature (Perugini & Leone, 2009), we hypothesized a double dissociation pattern of association, with the implicit MD significantly associated only with the "actual" lie behavior and the explicit MD significantly associated only with the self-reported cheating behavior.

Table 1. Descriptive Statistics of Average Latencies and Error Percentages for the Academic Cheating Behavior Implicit Association Test (Study 1).

		Latencies (in ms)						Error percentages					
	Attributes	Min.	Max.	*M*	*SD*	Skewness	Kurtosis	Min.	Max.	*M*	*SD*	Skewness	Kurtosis
Breaking Rules	1. Deceiving	592.63	1,698.50	985.40	275.86	0.63	−0.53	0.00	71.40	7.49	11.17	3.04	14.53
	2. Breaking	616.50	2,194.88	977.26	304.48	1.80	3.88	0.00	62.50	8.23	13.79	2.53	7.17
	3. Cheating	600.13	1,661.50	960.49	239.81	1.02	0.82	0.00	37.50	6.13	9.49	1.60	2.15
	4. Violating	522.00	2,078.88	957.35	247.99	1.42	4.74	0.00	42.90	7.84	10.34	1.42	1.89
	5. Tricking	592.63	2,067.38	955.79	262.86	1.59	3.70	0.00	62.50	7.45	11.49	2.15	6.53
Respecting Rules	6. Following	605.83	2,420.63	1,014.95	311.82	2.13	6.39	0.00	37.50	7.19	11.11	1.44	1.11
	7. Conforming	602.63	2,046.75	1,046.84	310.91	1.04	0.91	0.00	83.30	8.62	13.03	3.07	14.50
	8. Adhering	580.00	1,750.50	965.90	274.36	1.04	0.71	0.00	33.30	5.87	8.72	1.30	0.76
	9. Respecting	605.25	2,048.38	1,008.70	263.84	1.12	2.16	0.00	57.10	8.46	11.51	1.70	3.71
	10. Complying	582.13	1,952.63	969.49	275.98	1.16	1.38	0.00	57.10	5.97	9.87	2.48	9.18

Method

Procedure and participants. Participants were university psychology students. Their participation was completely voluntary and was rewarded with course credits. The final sample comprised 65 participants (73.5% females) with a mean age of 21.8 years ($SD = 1.4$). Final sample size was constant for all analytic purposes, and no participants were added after the data collection phase. Sample size was determined before any data analysis. Participants anonymously completed both implicit and explicit measures in a laboratory setting. Prior to each session, they were informed of the general aims of the study by trained research assistants. Moreover, they signed the informed consent previously approved by the Ethical Review Board of the department to which the second author is affiliated. Consistent with Schmukle and Egloff (2005), implicit measures were administered first to reduce potential carry-over effects, since they require less conscious engagement than explicit ones.

Measures. The measures used in this study were collected along with others not relevant for this article. All measures, manipulations, and exclusions are reported.

Academic cheating behavior implicit association test (ACB-IAT). It was the same as the one described in Study 1.

Academic moral disengagement relational responding task (AMD-RRT). This task was the same as Study 1. However, to overcome the limitations highlighted in Study 1, its stimuli were revised to simplify the wording. The full list of the revised statements is presented in Table 3.

Explicit academic moral disengagement. This scale comprised 18 items (adapted from Farnese et al., 2011) assessing students' proneness to engage in different forms of academic misconduct. Participants reported their agreement with each item on a 5-point Likert-type scale (from 1 = *not agree at all* to 5 = *completely agree*). The items of this scale were split into two test halves by balancing their corrected item-total correlations for further analytic purposes (Little et al., 2013).

Explicit academic cheating behaviors. This scale comprised nine items (adapted from Farnese et al., 2011) describing different academic cheating behaviors. Participants reported the frequency of engaging in such behavior during their academic career on a 5-point Likert-type scale (from 1 = *never or almost never* to 5 = *most of the times or always*). This measure serves as the deliberative outcome of the present study.

Actual lie behavior. An *ad hoc* measure based on the overclaiming technique originally proposed by Phillips and Clancy (1972) was developed for the present study. Specifically, students were asked to report their familiarity on a 5-point Likert-type scale (from 0 = *I never heard of it* to 4 = *I know it very well*) with 35 notions equally divided into seven psychology sub-disciplines (five notions per group). Notions included popular authors, theories, topics, and scientific terms that students were supposed to know based on their past courses attended during their bachelor's degree. Within each sub-discipline, four notions were true (i.e., they actually exist) and one notion was completely invented. We opted for this task because we intended to elicit a spontaneous and not morally connotated behavior. Indeed, it was designed to avoid any constraints, effort, or premeditation: it was anonymous, there was no expected evaluation, there were no evident benefits to gain or lose, and participants were not explicitly asked to be honest. Overall, this task was designed to create a setting with low risks in terms of potential impact on participants' self-image.

The scoring of the task would imply the use of formulae devised by signal detection theory (Paulhus et al., 2003). However, the use of such a scoring method requires strong assumptions, including the mutual independence between

Table 2. Descriptive Statistics of Average Latencies and Error Percentages for the AMD-RRT (Study 1).

		Latencies (in ms)						Error Percentages					
	Stimuli	Min.	Max.	*M*	*SD*	Skewness	Kurtosis	Min.	Max.	*M*	*SD*	Skewness	Kurtosis
Moral disengagement	1. Cheating in exams is fine when everyone does it	1,188.50	4,171.38	2,094.45	737.48	0.99	0.43	0.00	62.50	7.47	14.18	2.17	4.24
	2. Copying is as serious as system corruption	1,119.88	4,264.50	2,110.99	762.38	1.05	0.51	0.00	75.00	10.42	17.43	1.91	3.28
	3. It is not serious to cheat if no one is damaged	1,077.00	5,451.75	2,499.19	982.55	0.88	0.42	0.00	75.00	9.75	15.31	2.09	4.80
	4. Compared to the corruption of the system, copying is not serious	1,223.13	5,759.50	2,483.88	891.29	1.02	1.32	0.00	50.00	10.50	13.29	1.55	2.29
	5. If students copy, the fault is of those who do not supervise	1,149.50	5,426.00	2,179.83	717.91	1.72	4.76	0.00	62.50	10.79	16.70	1.72	2.09
Moral engagement	6. Cheating is always wrong even when many do it	1,120.50	4,752.88	2,423.81	976.52	0.90	−0.19	0.00	62.50	10.54	16.12	1.72	2.22
	7. Copying is wrong even if there is no control	1,037.00	6,322.13	2,461.43	1,085.82	1.39	2.10	0.00	62.50	10.99	16.95	1.52	1.42
	8. It is serious to cheat even if no one is damaged	995.00	5,020.25	2,473.61	956.65	0.98	0.45	0.00	87.50	12.55	18.29	1.82	3.53
	9. It is legitimate to give hints if it helps a friend	1,044.38	5,561.38	2,083.99	875.60	1.72	3.75	0.00	37.50	4.89	9.53	2.15	4.22
	10. Giving hints is wrong even if it is to help a friend	1,023.25	4,982.50	2,452.31	948.39	0.80	0.04	0.00	62.50	14.09	17.79	1.34	0.98

Note. Stimuli 1–5: moral disengagement; Stimuli 6–10: moral engagement. AMD-RRT = academic moral disengagement relational responding task.

Table 3. Descriptive Statistics of Average Latencies and Error Percentages for the Revised AMD-RRT (Study 2).

		Latencies (in ms)						Error percentages					
	Stimuli	Min.	Max.	*M*	*SD*	Skewness	Kurtosis	Min.	Max.	*M*	*SD*	Skewness	Kurtosis
Moral disengagement	1. Misbehaving in a corrupt system	781.88	2,969.13	1,460.17	473.52	0.84	0.67	0.00	50.00	6.92	11.06	1.72	2.92
	2. Copying when no one is checking	809.13	3,209.38	1,572.84	489.12	0.68	0.64	0.00	50.00	10.19	12.28	1.10	0.67
	3. Giving hints to help a friend	774.50	2,575.13	1,502.37	425.68	0.50	−0.42	0.00	62.50	12.31	14.06	1.19	1.26
	4. Cheating if no one is damaged	771.88	3,676.75	1,733.75	591.68	0.90	1.20	0.00	50.00	6.35	9.91	1.91	4.76
	5. Breaking the rules when everyone does it	818.75	4,314.38	1,699.10	576.78	1.61	5.26	0.00	37.50	6.92	9.64	1.40	1.60
Moral engagement	6. Behaving in the right way in a corrupted system	881.88	3,061.38	1,605.61	418.89	0.82	1.15	0.00	62.50	10.58	14.70	1.62	2.33
	7. Not copying when no one is checking	750.75	3,271.63	1,814.92	674.12	0.45	−0.81	0.00	75.00	18.08	18.09	0.96	0.51
	8. Not giving hints to help a friend	770.50	3,375.50	1,757.89	577.83	0.52	0.03	0.00	50.00	12.89	15.14	0.97	−0.12
	9. Not cheating even when it could be advantageous	652.63	3,230.25	1,814.43	597.85	0.21	−0.40	0.00	75.00	12.31	15.23	1.69	3.76
	10. Respecting the rules when everyone breaks them	770.25	2,800.63	1,720.15	536.73	0.27	−0.94	0.00	62.50	11.73	16.52	1.71	2.27

Note. AMD-RRT = academic moral disengagement relational responding task.

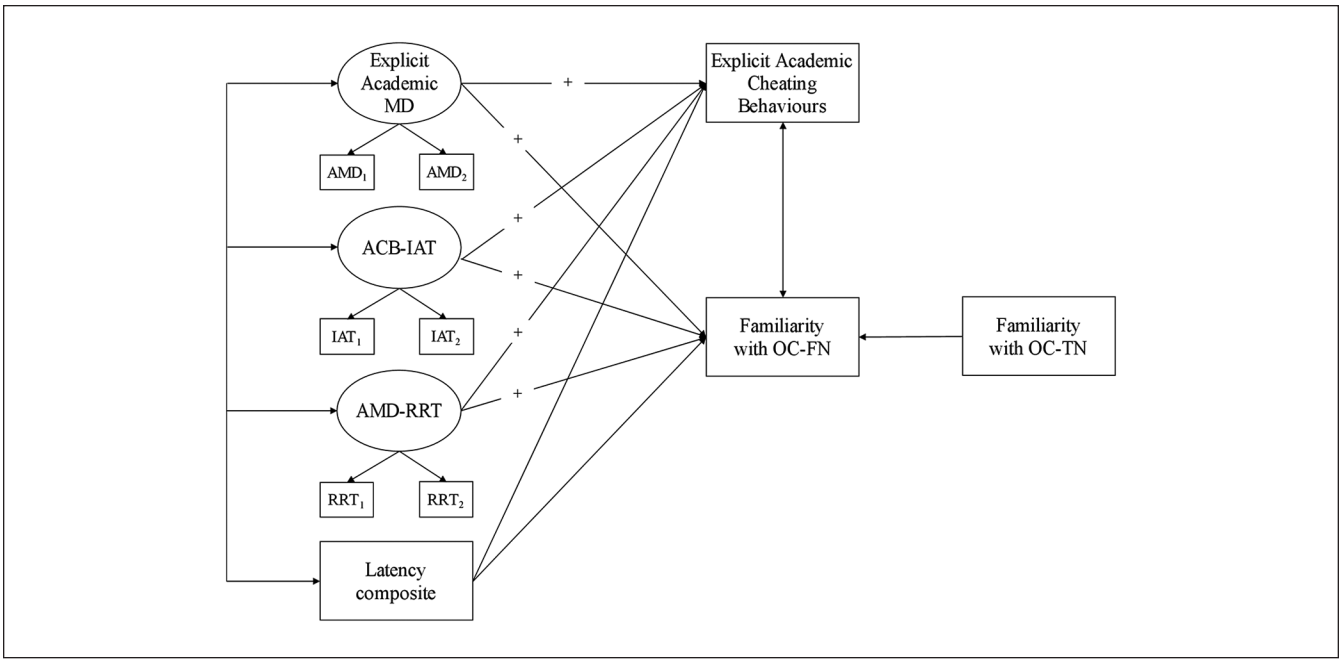

Figure 2. The posited SEM model (Study 2).

Note. The expected direction of substantive effects is placed above the arrows. Variance terms were not depicted to avoid clutter. SEM = structural equation model; MD = moral disengagement (AMD_1 and AMD_2 are the corresponding parcels); AMD = academic moral disengagement; ACB-IAT = academic cheating behavior implicit association test (IAT_1 and IAT_2 are the corresponding parcels); AMD-RRT = academic moral disengagement relational responding task (RRT_1 and RRT_2 are the corresponding parcels); latency composite = linear composite for latencies in ACB-IAT and AMD-RRT responses; OC-FN = continuous scores in the over-claiming measures attributed to false notions; OC-TN = continuous scores in the over-claiming measures attributed to true notions.

familiarity ratings with signal trials (i.e., true notions) and noise trials (i.e., false notions), as well as their univariate normality that are hardly met (on this topic, see Goecke et al., 2020). In our case, it is in fact plausible to expect that students may overclaim their familiarity with both true and false notions by activating motivated response biases (e.g., impression management). To overcome these limits, we preferred to adopt a two-phased strategy aimed to operationalize the substantive spontaneous criterion. First, we calculated two continuous scores: (a) overall familiarity with the true notions (i.e., the number of *hits*, Paulhus, 2012); (b) overall familiarity with the false notions (i.e., the number of false alarms, Paulhus, 2012). Second, we control the familiarity with the false notions for the familiarity with the true notions with a model-based strategy (see below). By doing so, we removed from the former the variability attributable to motivated response biases shared with the latter. Thus, we interpreted our substantive criterion as the actual lie behavior partialled out from possible motivated response biases.

Data analysis. As a first step, descriptive statistics and reliability of the study variables were investigated. The reliability was assessed with Spearman–Brown coefficient (Eisinga et al., 2013) for ACB-IAT and AMD-RRT (not being measured by multiple items), and with Cronbach's alpha for the other measures.

In line with our hypotheses, a structural equation model (SEM) was tested (see Figure 2). Following Gawronski and Bodenhausen (2006) and Perugini et al. (2010), we specified an additive pattern comprising direct effects of explicit (i.e., academic MD) and implicit measures (i.e., ACB-IAT and AMD-RRT) on both self-report (i.e., cheating behaviors) and actual cheating behaviors (i.e., overall familiarity with false notions of the over-claiming measure partialled out from the overall familiarity with true notions).

We defined three exogenous latent variables by two test halves each (i.e., explicit academic MD, ACB-IAT, and AMD-RRT). To favor the model identification, residual variances among test halves of the same latent construct were constrained to equality while all factor loadings were fixed to unity. In addition, a latency composite variable was defined by creating a single weighted linear component (by means of principal component analysis) based on the reaction times in both ACB-IAT and AMD-RRT (e.g., Möcks, 1986). This composite score was used as a control variable for both criteria, and it was specified as oblique with respect to all exogenous latent variables of the model.

Overall model fit was evaluated with multiple indices: (a) χ^2 test; (b) root mean square error of approximation (RMSEA); (c) comparative fit index (CFI); (d) Tucker–Lewis or non-normed fit index (TLI or NNFI); and (e) standardized root mean squared residual (SRMR). In line with

Table 4. Descriptive Statistics, Reliability Coefficients, and Zero-Order Correlations Among the Study Variables (Study 2).

	Descriptive statistics				Zero-order correlations						
Study variables	*M*	*SD*	Skewness	Kurtosis	1.	2.	3.	4.	5.	6.	7.
1. ACB-IAT	−0.56	0.33	0.53	0.29	.67						
2. AMD-RRT	−0.43	0.45	0.06	−0.52	−.15	.83					
3. Explicit academic moral disengagement	34.71	8.45	0.37	−0.31	−.01	−.04	.82				
4. Latency composite	0.00	1.00	0.02	−1.07	−.02	−.25*	−.08	—			
5. Explicit academic cheating behaviors	15.94	4.44	1.13	0.94	−.07	−.10	.61**	−.03	.79		
6. Familiarity with OC-FN	9.82	5.85	0.80	0.33	.12	.13	−.01	.08	.19	.81	
7. Familiarity with OC-TN	86.17	19.18	−1.31	1.47	−.19	−.01	.02	.04	.06	.50**	.92

Note. Reliability coefficients are reported along the principal diagonal as Spearman–Brown coefficients for implicit measures and latency composite, while they are Cronbach's alpha for the other study variables. ACB-IAT = academic cheating behavior implicit association test; AMD-RRT = academic moral disengagement relational responding task; OC-FN and OC-TN = continuous scores in the over-claiming measures attributed to false (F) and true (T) notions.
*$p < .05$. **$p < .01$.

commonly accepted cut-offs (e.g., Kline, 2016), models with satisfying fit should have RMSEA ≤ .08, CFI and TLI ≥ .90, and SRMR ≤ .08.

Finally, due to the novelty of this study, no prior knowledge regarding the effect sizes in terms of the impact of our implicit and explicit measures on the criterion variables was available. For this reason, we conducted a *sensitivity power analysis* with the software G*Power (Faul et al., 2009) to calculate the minimum detectable effect size (MED) for each structural regression coefficient. We adopted $1 - \beta = .80$ as power criterion (for a one-tailed level of $\alpha = .05$).

Results

Table 3 presents the average latencies and error percentages of the revised AMD-RRT stimuli. As expected, average latencies were much lower than those observed in Study 1, suggesting that trials were perceived as easier. This result may indicate that the AMD-RRT used in Study 2 reduces the possibility of a flattening effect between critical blocks due to the prevalence of a cautious style of response, increasing the validity of the task.

No missing data were detected in any study variables. Table 4 shows the descriptive statistics, while those pertaining to test halves defining the exogenous latent variables are provided in Supplemental Table S1. Explicit academic cheating behaviors showed a slight positive skewness, whereas the familiarity with truly existent notions from the over-claiming measure showed a similar skewness in the opposite direction of the frequency distribution. For these reasons, parameters of the further SEM were estimated using robust maximum likelihood (MLR). Explicit academic cheating behaviors were significantly and positively correlated with academic MD, and a similar result was found among the two scores of the over-claiming measure. Moreover, latency composite and AMD-RRT were negatively correlated, albeit this association was weak. All reliability coefficients were at least acceptable. Focusing on implicit measures, their reliability coefficients suggested that (especially for AMD-RRT) these measures comprised a substantive proportion of reliable variance, and their values were higher than what is commonly observed within the empirical literature regarding implicit attitudes (Fazio & Olson, 2003).

Results from the sensitivity power analysis indicated that the MED for each structural regression coefficient on the actual lie behavior was $f^2 = .089$. The final empirical model yielded the following fit: $\chi^2_{(n = 65,\, df = 29)} = 24.773$, $p = .690$, RMSEA = .000, CFI = 1.000, TLI = 1.050, SRMR = .069.[2]

Figure 3 presents the completely standardized model estimates. As factor loadings of test halves were fixed to unity and their residual terms were constrained to equality within each factor, they have the same standardized value. Explicit academic MD was positively associated with self-reported academic cheating behaviors (i.e., deliberative behavioral criterion). ACB-IAT and AMD-RRT showed a significant and positive association with the actual lie behavior. Overall, the 44% of the variability of the self-reported academic cheating behavior and the 41% of the variability of the actual lie behavior were explained by the independent variables, while the unique incremental contribution of implicit measures above and beyond the other independent variables on the familiarity with OC-TN scores was approximately the 17% of the criterion variability. In both cases, the effect size associated with ACB-IAT and AMD-RRT structural regression coefficients on actual lie behavior overcome the MED highlighted by the sensitivity power analysis. Finally, the actual lie behavior criterion was significantly and positively associated with self-reported cheating behavior.

Discussion

The results of Study 2 highlighted two important findings. First, changes made on the initial version of AMD-RRT stimuli proposed in Study 1 were effective. Indeed, the average

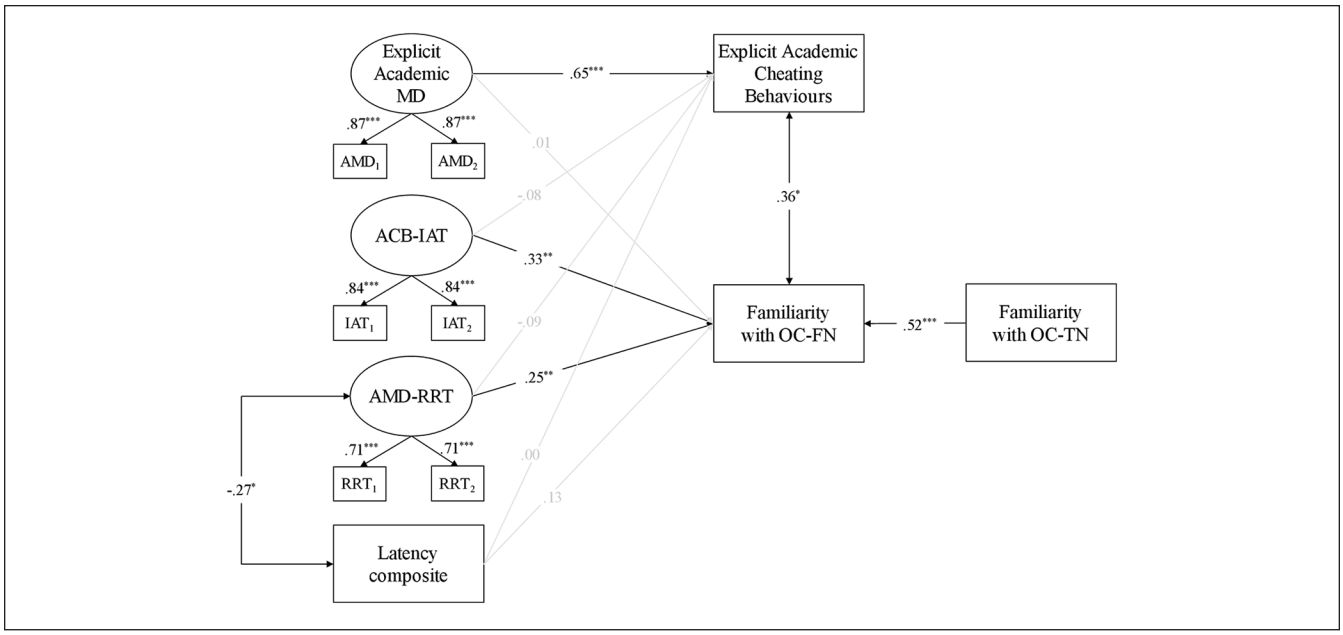

Figure 3. Completely standardized estimates of the text model (Study 2).
Note. The expected direction of substantive effects is placed above the arrows. Variance terms were not depicted to avoid clutter. Non-significant correlations among exogenous variables were not depicted. Non-significant direct effects are displayed in gray. MD = moral disengagement (AMD_1 and AMD_2 are the corresponding parcels); AMD = academic moral disengagement; ACB-IAT = academic cheating behavior implicit association test (IAT_1 and IAT_2 are the corresponding parcels); AMD-RRT = academic moral disengagement relational responding task (RRT_1 and RRT_2 are the corresponding parcels); latency composite = linear composite for latencies in ACB-IAT and AMD-RRT responses; OC-FN = continuous scores in the over-claiming measures attributed to false notions. OC-TN = continuous scores in the over-claiming measures attributed to true notions.
*$p < .05$. **$p < .01$. ***$p < .001$.

latencies associated with the revised stimuli were consistently lower than those observed in Study 1, and they were in line (as well as average error percentages) with others observed in other studies (e.g., De Houwer et al., 2015). Second, we found, in line with the double dissociation model (Perugini, 2005), that the explicit measure of MD was only associated with self-reported cheating behaviors, whereas the implicit MD was only associated with actual lie behavior above and beyond the relationship with the IAT measure of cheating behaviors. Findings from the final empirical model supported the mutual independence of the unique contributions of implicit measures, with respect to that of self-report academic MD. Both ACB-IAT and AMD-RRT yielded unique significant contributions in explaining the actual lie behavior. The implicit measures–spontaneous criterion relationships were not significant when considering the zero-order correlations, but significant in the SEM. This is mostly because in the SEM, the relationships are controlled for the measurement error (e.g., Meissner et al., 2019).

Results also showed that the ACB-IAT and the AMD-RRT were completely independent. This finding is not surprising, since these measures were designed to tap distinct constructs via different measurement paradigms. Specifically, as clarified in the introduction, the ACB-IAT is designed to assess the automatic associations between the self and misbehaviors, while the AMD-RRT is designed to assess relational information capturing the tendency to legitimate specific forms of academic misconducts. However, the specificity of the single criterion selected for Study 2 may lead to premature conclusions regarding the validity of the proposed implicit measures. Hence, we designed an additional study to replicate these results using a different spontaneous behavioral criterion.

Study 3

In this study, we aimed to examine the role of the newly developed measure of implicit MD in relation to a spontaneous behavioral criterion assessed following an approach already validated (Vohs & Schooler, 2008; von Hippel et al., 2005). Similar to Study 2, we expected that while the implicit MD would be associated with cheating behaviors in the task, the explicit MD would be associated only with the self-reported misconduct. In addition, we expected ACB-IAT to be also associated with cheating behaviors in the task.

Method

Procedure and participants. Due to the COVID-19 pandemic, it was not possible to conduct the study in a lab environment, and arrangements were made to implement it online. Participants were recruited in July 2020 using the web platform

Prolific Academic (ProA, http://www.prolific.ac). To ensure consistency with Studies 1 and 2, participants were required to be based in Italy, to be either full- or part-time students, and to be fluent in Italian. In addition, participants were required to complete the tasks by using a laptop or desktop. Sample size was determined a priori to ensure an acceptable likelihood to detect the expected effects (i.e., $1 - \beta = .90$ for a two-tailed level of $\alpha = .01$). We relied on the estimates obtained in the Study 2 model and determined the minimum sample size using the procedure developed by Satorra and Saris (1985) that recommended at least 110 participants. Both implicit and explicit measures were administered through the Inquisit 5 Web platform (Millisecond Software, 2020).

The initial sample included 123 participants, however five were excluded (one was not a student, three used a devise other than laptop or desktop, and one failed all the attention checks). The final sample comprised 118 participants (43.2% females) with a mean age of 22.7 years (SD = 3.3). The majority of the participants were born in Italy (94.1%) and were White Caucasians (87.3%). Most of them were full-time students (78.8%), enrolled in an undergraduate course (61.9%). Final sample size was constant for all analytic purposes, and no subjects were added after the data collection phase.

Participants anonymously completed both implicit and explicit measures as well as the behavioral task and were compensated £4.50 for their time. On average, participants needed about 39 min to complete the tasks. Before starting, participants were informed of the general aims of the study and were asked to provide their informed consent. The study was approved by the Ethical Review Board of the department to which the corresponding author is affiliated.

Measures. All the measures collected were included in the present study. All measures, manipulations, and exclusions are reported. Measures were presented in the following order: (a) actual cheating behavior, (b) implicit measures, and (c) explicit measures.

Actual cheating behavior. This was measured by following the computer-based mental-arithmetic task originally developed by von Hippel and colleagues (2005; Vohs & Schooler, 2008). Specifically, participants were presented with a sequence of 19 sums (e.g., 9 + 3 + 9 – 19 + 2 = ?), one after the other, with the first 4 aiming to familiarize participants with this task. Participants were told that they had 10 s to complete each task. They were also informed that, due to a "computer bug," the solution of each of the math problems would have appeared after 6 s unless they have pressed the spacebar. While further underlining in the instructions the importance of pressing the bar to prevent the solution to appear, participants were also told that the researcher would have not been able to know whether the bar was pressed or not. According to the previous studies (Vohs & Schooler, 2008), the criterion was defined as the average number of times participants pressed the spacebar across the 15 sums. To ensure that higher scores reflected higher cheating, the computed variable was multiplied by −1, ranging from −1.00 to 0.00. As per Study 2, also in this case, we can consider this criterion as reflecting spontaneous behavior, since the number of spacebar presses should not be inflated by explicit components of self-knowledge (von Hippel et al., 2005) and the task was designed to minimize constraints, effort, or premeditation.

Academic cheating behavior implicit association test (ACB-IAT). This was the same described and used in Studies 1 and 2.

Academic moral disengagement relational responding task (AMD-RRT). This task was the same described and used in Study 2.

Explicit academic moral disengagement. This was the same described and used in Study 2.

Explicit academic cheating behaviors. This measure was the same described and used in Study 2. As per Study 2, this measure served as the deliberative criterion.

Data analysis. The analytical strategy was the same as Study 2. First, descriptive statistics and reliability of the study variables were investigated. Then, the posited SEM was estimated and evaluated (see Figure 4) using the same approach as described in Study 2.

Results

Table 5 shows the descriptive statistics for the study variables. Given the slight departure from univariate normality of one variable, SEM was estimated using MLR as in Study 2. The explicit academic cheating behavior was significantly and positively correlated with the explicit academic MD. As expected, while AMD-RRT was significantly associated with the behavioral criterion, the ACB-IAT was not. Moreover, latency composite and AMD-IAT were positively correlated, albeit this association was weak. All reliability coefficients were at least acceptable.

The final empirical model yielded the following fit: $\chi^2_{(n = 118,\, df = 29)} = 40.064$, $p = .007$, RMSEA = .088, CFI = .942, TLI = .900, SRMR = .084. Figure 5 presents the model standardized estimates. As factor loadings of test halves were fixed to unity and their residual terms were constrained to equality within each factor, they have the same standardized value. Similar to Study 2, explicit academic MD was positively associated with self-reported academic cheating behaviors (i.e., deliberative behavioral criterion). While AMD-RRT was significantly and positively associated with the spontaneous behavioral criterion, ACB-IAT was not significant.

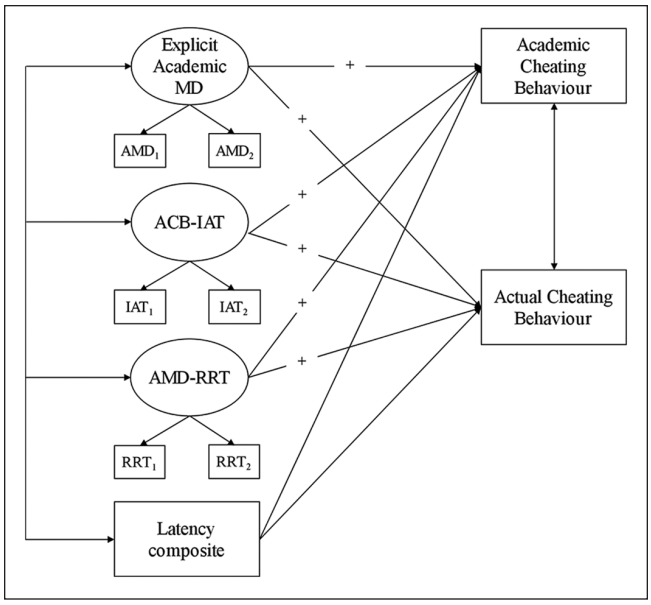

Figure 4. The posited SEM model (Study 3).
Note. The expected direction of substantive effects is placed above the arrows. Variance terms were not depicted to avoid clutter. SEM = structural equation model; MD = moral disengagement (AMD_1 and AMD_2 are the corresponding parcels); AMD = academic moral disengagement; ACB-IAT = academic cheating behavior implicit association test (IAT_1 and IAT_2 are the corresponding parcels); AMD-RRT = academic moral disengagement relational responding task (RRT_1 and RRT_2 are the corresponding parcels).

Overall, the 52% of the variability of the self-reported academic misbehavior and the 7% of the variability of the behavioral criterion were explained by the independent variables of the model. In the latter case, this result was fully attributable to the effect of AMD-RRT.

Discussion

Results from Study 3 confirmed the findings from Study 2 of a double dissociation pattern (Perugini, 2005; Perugini & Leone, 2009). Specifically, results showed that while the self-reported academic MD was associated only with the self-reported academic misconduct, the AMD-RRT was significantly associated only with the behavioral criterion. Results also showed a not significant relationship between the self-reported and actual cheating behavior, possibly reflecting the common intention-behavior gap (Sheeran, 2002).

Different to Study 2, the ACB-IAT was not associated with the spontaneous behavioral criterion. This finding might be attributed to the possible frame-of-reference effects (e.g., Schmit et al., 1995). Specifically, since the ACB-IAT measures implicit associations concerned with breaking versus respecting rules within the academic context, it might fail to capture aspects of spontaneous behavior which are not strictly rooted within this context (as it was for the spontaneous behavior measured in Study 2). As in Study 2, ACB-IAT and AMD-RRT were fully independent.

General Discussion

The results of this research support the theorization of an implicit MD as well as its assessment through a newly developed implicit measurement procedure. Consistent with double dissociation pattern, results from both Study 2 and Study 3 showed that only the implicit MD was associated with actual cheating behavior in situations in which one's own self-interest is not clearly at stake, there is no apparent external evaluation, and social desirability is minimized. In line with the literature on bounded ethicality, these results suggest that even when people know what is "the right thing to do," they may still behave otherwise (e.g., Bazerman & Gino, 2012; Chugh et al., 2005; Sezer et al., 2015). This might be the result of the automatic activation of traces of memories in which misconduct has been legitimized. In other words, the implicit MD represents the automatic component of the mechanisms bypassing the self-regulatory system, influencing spontaneous behavior. With this research, we complement the understanding of MD functioning by introducing its implicit counterpart and provided the first evidence of the existence of possible automatic MD processes.

In our study, we found no significant correlations between the implicit and the explicit components of MD. Although this result needs to be further investigated in future studies, it is likely that these two picked two different levels of functioning of MD. While the implicit component would capture what might lead an individual to misbehave in situations implying on-the-spot decisions, the explicit measure would capture the propensity to adopt justification mechanisms in situations characterized by deliberative decisions and moral dilemmas. Indeed, as suggested by Nosek and colleagues (2011, p. 154), implicit measures "can reveal effects that are very different from explicit measurement of the same content."

With this research, we also contributed to the debate on the possible methods to operationalize MD. The exclusive use of self-report assessment only enables measurement of what individuals think about themselves and are willing or able to report. The adoption of an implicit measure allows researchers to overcome these limits and provides a key to access individuals' automatic "internal world." As suggested in the behavioral process model of personality (Back et al., 2009; Strack & Deutsch, 2004), individuals might have a positive image of themselves (explicit self-concept of personality) and might tend not to attribute to themselves the negative elements that might, on the contrary, be implicitly part of them.

In relation to the broader literature on moral psychology, we have offered a novel approach to address the exclusive use of self-report measures when assessing morality. This major issue has been highlighted in a recent meta-analysis (Ellemers et al., 2019) suggesting that explicit beliefs about one's own morality are not necessarily sufficient in ensuring engagement in moral behavior. In addition, as suggested by Chugh and Kern (2016, p. 88), "much of our unethical

Table 5. Descriptive Statistics, Reliability Coefficients, and Zero-Order Correlations Among the Study Variables (Study 3).

	Descriptive statistics				Zero-order correlations					
Study variables	*M*	*SD*	Skewness	Kurtosis	1.	2.	3.	4.	5.	6.
1. ACB-IAT	−0.41	0.34	0.31	0.16	.72					
2. AMD-RRT	−0.20	0.49	0.41	0.17	.05	.73				
3. Explicit academic moral disengagement	4.80	11.68	0.55	−0.09	−.05	.05	.89			
4. Latency composite	0.00	1.00	0.82	1.75	.24*	−.02	.05	—		
5. Explicit academic cheating behaviors	16.01	4.84	0.90	0.67	−.12	−.01	.67**	−.09	.78	—
6. Actual cheating behavior	−.68	0.45	−0.08	−0.31	.01	.21*	−.06	.00	−.12	.95

Note. Coefficients are reported along the principal diagonal reliability (i.e., Spearman–Brown coefficients for ACB-IAT and AMD-RRT, and Cronbach's alpha for the remainders). ACB-IAT = academic cheating behavior implicit association test; AMD-RRT = academic moral disengagement relational responding task.
*$p < .05$. **$p < .01$.

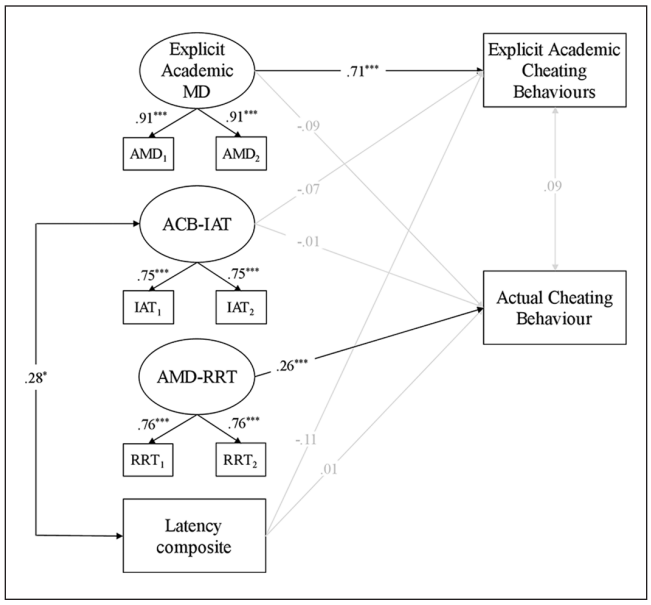

Figure 5. Completely standardized estimates of the text model (Study 3).
Note. Variance terms were not depicted to avoid clutter. Non-significant correlations among exogenous variables were not depicted. Non-significant direct effects are displayed in gray. MD = moral disengagement (AMD_1 and AMD_2 are the corresponding parcels); AMD = academic moral disengagement; ACB-IAT = academic cheating behavior implicit association test (IAT_1 and IAT_2 are the corresponding parcels); AMD-RRT = academic moral disengagement relational responding task (RRT_1 and RRT_2 are the corresponding parcels).
*$p < .05$. **$p < .01$. ***$p < .001$.

behaviour takes place outside of our awareness," and when considering individuals' moral conduct, the "self-view is a more forceful and more automatic influence than self-interest on ethical decision-making."

Moreover, results of the present research provide further support for the validity of the RRT as an implicit measure of psychological constructs and also for a relational conception of implicit social-cognitions. From this point of view, a relational implicit measure like the RRT is particularly suited to measure complex constructs like MD that cannot be assessed with implicit associative measures.

In our study, there was no significant correlation between the IAT and RRT measures. This result suggests that individuals' implicit MD is independent from their implicit moral self-concept. This could possibly explain the incongruence between the way individuals represent themselves and their tendency to legitimize wrongdoing, hence, why "otherwise considerate people to commit transgressive acts without experiencing personal distress" (Bandura et al., 2000, p. 58). This result could be also due to the different assumptions of these two methodologies: while the IAT is based on implicit associative models, the RRT is based on an implicit propositional theoretical framework.

Notwithstanding the innovative contribution of our findings, we are aware of some limitations that future studies should address. First, the newly developed implicit measure of MD should be tested in different contexts overcoming the possible self-selection bias (e.g., the recruitment of students enrolled within a research methods class or in an online research platform) that might have affected our research. However, results have been cross-validated using two different behavioral criteria. Second, it would be useful to further test the concurrent and predictive validity of the implicit MD, considering a wider range of misbehaviors, implying a different level of risk for participants' self-image. It is likely that when considering actual misbehaviors characterized by greater cognitive costs, for instance, in terms of planning or of moral dilemmas, and for which there are potentially serious consequences for social and moral self-image, the self-reported MD might have a more important role. Third, in line with the literature on moral psychology (e.g., Ellemers et al., 2019) underlining the need to concurrently examine the role of cognitions and emotions when studying transgressive and deviant behavior, it would be relevant to investigate whether and how implicit MD is influenced by emotions when misbehaving. Fourth, future studies should investigate the incremental validity of the implicit measure of MD, considering

also the role of moral standards and norms in situations where these are more or less salient and shared. Finally, it would be relevant to investigate the association between implicit and explicit MD. In our study, there was no significant correlation; however, this does not exclude the fact that it could be possible to identify different configurations of individuals characterized by a range of combination of implicit and explicit MD.

Conclusion

This research contributes to the broader debate of bounded ethicality (e.g., Bazerman & Gino, 2012; Bazerman & Sezer, 2016; Chugh et al., 2005) by postulating an implicit MD, presenting a valid and reliable strategy to assess it, and providing evidence of its association with actual misconduct (with two different behavioral spontaneous criteria). Overall, where the self-reported MD was only associated with self-reported cheating behavior, the implicit MD was only associated with the actual cheating behavior.

Misconduct is still a challenge in the educational context and more in general in our society. The results of this research have important implications for the design of training aimed at increasing individuals' moral regulation. In particular, preventing programs should include sessions that help individuals to learn about the role of implicit processes and self-reflective in the moral domain. This would in turn allow them to recognize their implicit justification mechanisms that redefine the misbehavior itself, alter the perception of its consequences, obscure the individual's agentic role, and hold the victim responsible.

Acknowledgments

The authors would like to thank the anonymous reviewer Prof. Jan De Houwer and the Action Editor Professor Leonel Garcia-Marques for their insightful and valuable comments which substantially improved the article. They would also like to thank Prof. Olga Tregraskis and Prof. Ana Sanz Vergel for their comments on a preliminary version of this paper.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

R. Fida https://orcid.org/0000-0001-6733-461X

Supplemental Material

Supplemental material is available online with this article.

Notes

1. In this article, the terms "implicit" and "automatic" are used as synonyms to encompass unintentional and "less accessible" processes (e.g., Moors & De Houwer, 2006).
2. In line with the signal detection theory (SDT; Macmillan & Creelman, 1991), an additional model was tested by operationalizing our substantive criterion as the average between the proportion of hits and false alarms (i.e., the common-sense approach). This scoring procedure provides an index of knowledge exaggeration roughly overlapping the criterion location *c* (Paulhus et al., 2003). Although this model reached an excellent fit to the data, $\chi^2_{(n = 65,\ df = 21)} = 15.956$, $p = .772$, RMSEA = .000, CFI = 1.000, TLI = 1.050, SRMR = .060, the effects of the implicit measures on the criterion (scored with the *common-sense*) were not significant. This result could be due to the fact that the SDT formula, by pooling together the proportion of hits and false alarms, does not remove the variance component attributable to the proportion of hits from the one related to false alarms.

References

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology*, *97*(3), 533–548.

Bandura, A. (1991). Social-cognitive theory of moral thought and action. In W. M. Kurtines & J. L. Gewirtz (Eds.), *Handbook of moral behavior and development* (Vol. 1, pp. 45–103). Lawrence Erlbaum.

Bandura, A. (1999). A social-cognitive theory of personality. In L. Pervin & O. John (Eds.), *Handbook of personality* (2nd ed., pp. 154–196). Guilford Publications. (Reprinted from D. Cervone & Y. Shoda [Eds.], The coherence of personality. Guilford Press)

Bandura, A. (2008). The reconstrual of "free will" from the agentic perspective of social-cognitive theory. In J. Baer, J. C. Kaufman, & R. F. Baumeister (Eds.), *Are we free* (pp. 86–127). Oxford University Press.

Bandura, A. (2016). *Moral disengagement: How people do harm and live with themselves*. Worth Publishers.

Bandura, A., Caprara, G. V., & Zsolnai, L. (2000). Corporate transgressions through moral disengagement. *Journal of Human Values*, *6*(1), 57–64.

Barnes-Holmes, D., Barnes-Holmes, Y., Stewart, I., & Boles, S. (2010). A sketch of the Implicit Relational Assessment Procedure (IRAP) and the Relational Elaboration and Coherence (REC) model. *The Psychological Record*, *60*(3), 527–542.

Bazerman, M. H., & Gino, F. (2012). Behavioral ethics: Toward a deeper understanding of moral judgment and dishonesty. *Annual Review of Law and Social Science*, *8*, 85–104.

Bazerman, M. H., & Sezer, O. (2016). Bounded awareness: Implications for ethical decision making. *Organizational Behavior and Human Decision Processes*, *136*, 95–105.

Bonett, D. G. (2002). Sample size requirements for testing and estimating coefficient alpha. *Journal of Educational and Behavioral Statistics*, 27(4), 335–340.

Chugh, D., Bazerman, M. H., & Banaji, M. R. (2005). Bounded ethicality as a psychological barrier to recognizing conflicts of

interest. In D. A. Moore, D. M. Cain, G. Loewenstein, & M. H. Bazerman (Eds.), *Conflicts of interest: Challenges and solutions in business, law, medicine, and public policy* (pp. 74–95). Cambridge University Press.

Chugh, D., & Kern, M. C. (2016). A dynamic and cyclical model of bounded ethicality. *Research in Organizational Behavior*, *36*, 85–100.

Cummins, J., Lindgren, K. P., & De Houwer, J. (2020). On the role of (implicit) drinking self-identity in alcohol use and problematic drinking: A comparison of five measures. *Psychology of Addictive Behaviors: Journal of the Society of Psychologists in Addictive Behaviors*. Advance online publication. https://doi.org/10.1037/adb0000643

De Houwer, J. (2014). A propositional model of implicit evaluation. *Social and Personality Psychology Compass*, *8*(7), 342–353.

De Houwer, J., Heider, N., Spruyt, A., Roets, A., & Hughes, S. (2015). The relational responding task: Toward a new implicit measure of beliefs. *Frontiers in Psychology*, *6*, Article 319.

De Houwer, J., & Moors, A. (2010). Implicit measures: Similarities and differences. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social-cognition: Measurement, theory, and applications* (pp. 176–193). Guilford Press.

Dentale, F., Vecchione, M., Ghezzi, V., Spagnolo, G., Szemenyei, E., & Barbaranelli, C. (2020). Beyond an associative conception of automatic self-evaluations: Applying the relational responding task to measure self-esteem. *The Psychological Record*, *70*, 227–242.

Eisinga, R., Te Grotenhuis, M., & Pelzer, B. (2013). The reliability of a two-item scale: Pearson, Cronbach, or Spearman-Brown? *International Journal of Public Health*, *58*(4), 637–642.

Ellemers, N., van der Toorn, J., Paunov, Y., & van Leeuwen, T. (2019). The psychology of morality: A review and analysis of empirical studies published from 1940 through 2017. *Personality and Social Psychology Review*, *23*(4), 332–366.

Farnese, M. L., Tramontano, C., Fida, R., & Paciello, M. (2011). Cheating behaviors in academic context: Does academic moral disengagement matter? *Procedia: Social and Behavioral Sciences*, *29*, 356–365.

Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*, 1149–1160.

Fazio, R. H., & Olson, M. A. (2003). Implicit measures in social-cognition research: Their meaning and uses. *Annual Review of Psychology*, *54*, 297–327.

Fida, R., Tramontano, C., Paciello, M., Ghezzi, V., & Barbaranelli, C. (2018). Understanding the interplay among regulatory self-efficacy, moral disengagement, and academic cheating behaviour during vocational education: A three-wave study. *Journal of Business Ethics*, *153*(3), 725–740.

Gawronski, B., & Bodenhausen, G. V. (2006). Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. *Psychological Bulletin*, *132*(5), 692–731.

Gawronski, B., & Payne, B. K. (2010). *Handbook of implicit social-cognition: Measurement, theory, and applications*. Guilford Press.

Goecke, B., Weiss, S., Steger, D., Schroeders, U., & Wilhelm, O. (2020). Testing competing claims about overclaiming. *Intelligence*, *81*, 101470.

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464–1480.

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003). Understanding and using the implicit association test: I. An improved scoring algorithm. *Journal of Personality and Social Psychology*, *85*(2), 197–216.

Heider, N., Spruyt, A., & De Houwer, J. (2015). Implicit beliefs about ideal body image predict body image dissatisfaction. *Frontiers in Psychology*, *6*, Article 1402.

Heider, N., Spruyt, A., & De Houwer, J. (2018). Body dissatisfaction revisited: On the importance of implicit beliefs about actual and ideal body image. *Psychologica Belgica*, *57*(4), 158–173.

Hughes, S., & Barnes-Holmes, D. (2013). A functional approach to the study of implicit cognition: The IRAP and the REC model. In B. Roche & S. Dymond (Eds.), *Advances in relational frame theory & contextual behavioural science: Research & applications* (pp. 97–126). New Harbinger.

Hughes, S., Barnes-Holmes, D., & Vahey, N. (2012). Holding on to our functional roots when exploring new intellectual islands: A voyage through implicit cognition research. *Journal of Contextual Behavioral Science*, *1*(1–2), 17–38.

Hyde, L. W., Shaw, D. S., & Moilanen, K. L. (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.

International Center for Academic Integrity. (2015). *Overview*. https://www.academicintegrity.org/statistics/

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.

Koning, I. M., Spruyt, A., Doornwaard, S. M., Turrisi, R., Heider, N., & De Houwer, J. (2016). A different view on parenting: Automatic and explicit parenting cognitions in adolescents' drinking behavior. *Journal of Substance Use*, *22*(1), 96–101.

Lapsley, D. K., & Hill, P. L. (2008). On dual processing and heuristic approaches to moral cognition. *Journal of Moral Education*, *37*(3), 313–332.

LeBel, E. P., & Paunonen, S. V. (2011). Sexy but often unreliable: The impact of unreliability on the replicability of experimental findings with implicit measures. *Personality and Social Psychology Bulletin*, *37*(4), 570–583.

Little, T. D., Rhemtulla, M., Gibson, K., & Schoemann, A. M. (2013). Why the items versus parcels controversy needn't be one. *Psychological Methods*, *18*(3), 285–300.

Lord, F. M., & Novick, R. (1968). *Statistical theories of mental test scores*. Addison-Wesley.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge University Press.

Marquardt, N. (2010). Implicit mental processes in ethical management behavior. *Ethics & Behavior*, *20*(2), 128–148.

Marquardt, N., & Hoeger, R. (2009). The effect of implicit moral attitudes on managerial decision-making: An implicit social-cognition approach. *Journal of Business Ethics*, *85*(2), 157–171.

Meissner, F., Grigutsch, L. A., Koranyi, N., Müller, F., & Rothermund, K. (2019). Predicting behavior with implicit measures: Disillusioning findings, reasonable explanations, and sophisticated solutions. *Frontiers in Psychology*, *10*, Article 2483.

Millisecond Software. (2020). *Inquisit 5* [Computer software]. https://www.millisecond.com

Möcks, J. (1986). The influence of latency jitter in principal component analysis of event-related potentials. *Psychophysiology*, *23*(4), 480–484.

Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*, *132*(2), 297–326. https://doi.org/10.1037/0033-2909.132.2.297

Newman, A., Le, H., North-Samardzic, A., & Cohen, M. (2020). Moral disengagement at work: A review and research agenda. *Journal of Business Ethics*, *167*, 535–570.

Nosek, B. A., Hawkins, C. B., & Frazier, R. S. (2011). Implicit social-cognition: From measures to mechanisms. *Trends in Cognitive Sciences*, *15*(4), 152–159.

Paciello, M., Fida, R., Tramontano, C., Lupinetti, C., & Caprara, G. V. (2008). Stability and change of moral disengagement and its impact on aggression and violence in late adolescence. *Child Development*, *79*(5), 1288–1309.

Paulhus, D. L. (2012). Overclaiming on personality questionnaires. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 151–164). Oxford University Press.

Paulhus, D. L., Harms, P. D., Bruce, M. N., & Lysy, D. C. (2003). The over-claiming technique: Measuring self-enhancement independent of ability. *Journal of Personality and Social Psychology*, *84*(4), 890–904.

Payne, B., & Gawronski, B. (2010). A history of implicit social cognition. In B. Gawronski & B. K. Payne (Eds.), *Handbook of implicit social-cognition: Measurement, theory, and applications* (pp. 1–15). Guilford Press.

Perugini, M. (2005). Predictive models of implicit and explicit attitudes. *British Journal of Social Psychology*, *44*(1), 29–45.

Perugini, M., & Leone, L. (2009). Implicit self-concept and moral action. *Journal of Research in Personality*, *43*(5), 747–754.

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. 255–278). Guilford Press.

Phillips, D. L., & Clancy, K. J. (1972). Some effects of "social desirability" in survey studies. *American Journal of Sociology*, *77*(5), 921–940.

Reynolds, S. J., Leavitt, K., & DeCelles, K. A. (2010). Automatic ethics: The effects of implicit assumptions and contextual cues on moral behavior. *Journal of Applied Psychology*, *95*(4), 752–760.

Satorra, A., & Saris, W. E. (1985). The power of the likelihood ratio test in covariance structure analysis. *Psychometrika*, *50*, 83–90.

Schmit, M. J., Ryan, A. M., Stierwalt, S. L., & Powell, A. B. (1995). Frame-of-reference effects on personality scale scores and criterion-related validity. *Journal of Applied Psychology*, *80*(5), 607–620.

Schmukle, S. C., & Egloff, B. (2005). A latent state-trait analysis of implicit and explicit personality measures. *European Journal of Psychological Assessment*, *21*(2), 100–107.

Sezer, O., Gino, F., & Bazerman, M. H. (2015). Ethical blind spots: Explaining unintentional unethical behavior. *Current Opinion in Psychology*, *6*, 77–81.

Sheeran, P. (2002). Intention–behavior relations: A conceptual and empirical review. In W. Strobe & M. Hewstone (Eds.), *European review of social psychology* (Vol. 12, pp. 1–30). Wiley.

Strack, F., & Deutsch, R. (2004). Reflective and impulsive determinants of social behavior. *Personality and Social Psychology Review*, *8*(3), 220–247.

Tibboel, H., De Houwer, J., Dirix, N., & Spruyt, A. (2017). Beyond associations: Do implicit beliefs play a role in smoking addiction? *Journal of Psychopharmacology*, *31*(1), 43–53.

Vohs, K. D., & Schooler, J. W. (2008). The value of believing in free will: Encouraging a belief in determinism increases cheating. *Psychological Science*, *19*(1), 49–54.

von Hippel, W., Lakin, J. L., & Shakarchi, R. J. (2005). Individual differences in motivated social cognition: The case of self-serving information processing. *Personality and Social Psychology Bulletin*, *31*(10), 1347–1357.

Whitley, B. E. (1998). Factors associated with cheating among college students: A review. *Research in Higher Education*, *39*(3), 235–274.