


RESEARCH ARTICLE

# A double-edged sword: when does identity threat affect unethical behavior?

Jinyi Zhou<sup>1\*</sup> , Jifang Dou<sup>2</sup> and Xiaoye (May) Wang<sup>3</sup>

<sup>1</sup>Department of Business Administration, Donlinks School of Economics and Management, University of Science and Technology Beijing, Beijing, China, <sup>2</sup>The PBC School of Finance, Tsinghua University, China and the Teaching Center for Writing and Communication, Tsinghua University, China and <sup>3</sup>Department of Leadership and Organization Management, School of Economics and Management, Tsinghua University, Beijing, China

\*Corresponding author. Email: [jinyi@ustb.edu.cn](mailto:jinyi@ustb.edu.cn)

(Received 22 June 2018; accepted 8 November 2019)

## Abstract

Although individuals have different kinds of defensive strategies towards identity threat, the relationship between identity threat and unethical behavior is still unclear. In the current study, according to identity threat and self-affirmation theory, we propose and test the role of publicness of identity threat in determining whether identity threat will lead to unethical behavior. One online experiment with 197 participants (mixed design) and one laboratory experiment with 86 participants (between-subject design) are used to test our hypotheses. Our findings reveal that when individuals' identity threat is from the public sphere, it will increase their unethical behavior, but when such a threat is from the private sphere, it will reduce their unethical behavior. Theoretical and practical implications are discussed.

**Keywords:** identity threat; unethical behavior; publicness; self-affirmation

## Introduction

Identity threat, as a kind of environmental challenge to the adequacy of individuals' identity, is defined as 'experiences appraised as indicating potential harm to the value, meanings, or enactment of an identity' (Petriglieri, 2011: 644) or 'any overt action by another party that challenges, calls into question, or diminishes a person's sense of competence, dignity, or self-worth' (Aquino & Douglas, 2003: 196). It includes destructive behavior, verbal harassment, harsh criticism of one's competency, and public humiliation from environments (Aquino & Douglas, 2003; Bies, 2001).

Identity threat is a common occurrence in the workplace (Baron & Neuman, 1996) that may arouse individuals' essential motives to defend themselves. Prior studies find that threats to individuals' social identity or personal identity can increase some kinds of unethical behavior. For example, scholars find that identity threat can lead to aggressive workplace behaviors, such as stealing, cheating, and lying (Belmi, Barragan, Neale, & Cohen, 2015), as social interactionist theories indicate that employees will use retaliatory responses against the perceived source of threat by the threat-recipient (Tedeschi & Felson, 1994). Others observe that individuals' displaced aggression is a robust and valid phenomenon in organizations (Marcus-Newhall, Pedersen, Carlson, & Miller, 2000) and that individuals' defensive behaviors, such as unethical behavior, can serve as value-expressive tools to vent negative emotions towards any available targets (Petriglieri, 2011). Additionally, threat-recipients in organizations can direct such aggressive behaviors against both perpetrators and nonperpetrators (Aquino & Douglas, 2003).

Prior studies have noted the critical role of identity in morality literature (Aquino & Reed, 2002; Treviño, Weaver, & Reynolds, 2006). Individuals are always present behind different

kinds of roles (Ashforth & Mael, 1989; Frone, Russell, & Cooper, 1995; Leavitt, Reynolds, Barnes, Schilpzand, & Hannah, 2012; Rozuel, 2011; Song, Yu, Zhang, & Jiang, 2015). Given that activating different personal identities simultaneously activates different self-concepts, motives, and behavior scripts (Brewer & Gardner, 1996; Markus, 1977; Markus & Kunda, 1986), disparate types of identity threats may also lead to individuals' various kinds of defensive intentions (Leavitt et al., 2012; Mayer, Greenbaum, Kuenzi, & Shteynberg, 2009). In the current study, we focus on one specific kind of unethical behavior that can be aroused to protect individuals' identity, such as cheating to maintain the status of one's identity (Schoderbek & Deshpande, 1996). Moreover, we try to test the dichotomous impacts of public or private aspects of individuals' identity on their unethical behaviors.

Our study is based on theories about individuals' different responses to identity threats (Aquino & Douglas, 2003; Petriglieri, 2011). We propose that the effect of identity threat on individuals' unethical behavior may depend on the publicness of such threat. Specifically, the publicness of identity threat, which taps into a more generally sensitive self towards others' responses, is one of the key factors determining whether individuals can use unethical behavior as a defensive response to defend their identities (Hollenbeck, Williams, & Klein, 1989; Petriglieri, 2011; Salancik, 1977). Generally, individuals have a basic need to construct their desired identity image (Leary & Baumeister, 2000; Sherman & Cohen, 2006). Unethical behavior serves as an identity-protection response that can be used to conceal individuals' actual identity status in public spheres due to individuals' basic need to meet public standards (i.e., construct a desirable identity image; Leary & Baumeister, 2000) or bifurcate the effects of identity threat when believing it cannot be perceived by others (i.e., privateness).

Self-affirmation, as an effective psychological intervention that can reduce the influences of identity threat (Cohen & Sherman, 2014; Sherman & Cohen, 2006; Steele, 1988), is used in our study as the control group. Self-affirmation theory (Steele, 1988; Steele & Liu, 1983) and related empirical studies support the idea that affirming individuals' global self-worth can reduce the amount of their identity-defensive behaviors (Cohen & Sherman, 2014). In the current study, we suggest that self-affirmation can also serve as an identity-defensive response (Petriglieri, 2011). Affirming identities other than the currently threatened ones provide suitable control groups for us to investigate the effects of identity threat (Fast, Burris, & Bartel, 2014). Ultimately, in this study, we use self-affirmation as our control group in which identity threat does not exist.

In exploring the above issues, we seek to make at least three important contributions. First, we seek to contribute to the identity threat theory by investigating the potential influences of the public or private aspect of identity threat. Second, by examining the relationship between identity threat and unethical behaviors in public and private situations, we seek to enrich unethical behavior literature by incorporating identity threat as one possible predictor and differentiating the effects of private or public identity threat simultaneously. Third, we provide more empirical evidence of the 'dark side' of self-affirmation, which is often considered as a positive accelerator of mitigatory attitudes and behaviors (Cohen & Sherman, 2014). One online experiment and one laboratory experiment are conducted to test these effects in this study.

## Theoretical Development and Hypotheses

The resistance to challenge the idea that threats to currently salient identity forces individuals to behave defensively (Aquino & Douglas, 2003; Petriglieri, 2011; Steele, 1988). Based on this premise, Petriglieri (2011) observes two basic coping responses individuals may use to deal with identity threats: the identity-protection response, which targets the source of identity threat, and the identity-restructuring response, which targets the importance of the threatened identity. The identity-protection strategy encompasses concealing the current identity status, and the identity-restructuring strategy encompasses the importance-changing response towards the threatened identity. In Petriglieri's view, individuals will evaluate these two responses to identity threat

and choose one based on the inherent characteristics of the identity threat. For example, unethical behavior can be used to express fake impressions to deal with potential harm to one's identity image (Treviño, den Nieuwenboer, & Kish-Gephart, 2014). However, as previous studies suggest, not all kinds of identity threats can be concealed by unethical behaviors (Baumeister, 2010; Mar, De Young, Higgins, & Peterson, 2006; Trivers, 2000). If the self-importance of identity threat is increasing, individuals will hardly use any concealment strategy (Petriglieri, 2011).

According to James (1950), individuals' personal identity consists of moral sensibility; conscience; and desire for achievement, mastery, and competence (cf. Mayer et al., 2009). In this study, we posit that the effect of individuals' identity threat on unethical behavior may be influenced by the publicness of such threat. People hold generally positive self-concepts of themselves as rational, intelligent, responsible, and healthy (Crocker, Niiya, & Mischkowski, 2008; Steele, 1988) and have a basic need to maintain positive identity image in both their own and others' eyes (Baumeister & Leary, 1995; Sherman & Cohen, 2006). Therefore, we shed light on one important categorized characteristic: the *publicness* of a specific identity threat that affects an individual's decisions under the threat. In fact, some scholars have already attempted to differentiate the public aspects of the self from the private aspects (Baumeister, 1986; Tetlock & Manstead, 1985). For instance, individuals have a basic need to maintain desirable identity images by publicly claiming attributes consistent with those identities (Gollwitzer, 1986). Baumeister and Leary (1995) claim that individuals want to maintain the identity status in both their own and others' eyes. In this study, we posit that when identity threat comes from the public sphere, unethical behavior can serve as an effective defensive method in a specific domain because of individuals' inherent need to satisfy publicly acceptable requirements. In contrast, when identity threat is from the private sphere, it increases the self-importance of such identity. The need for accuracy requires them to have an accurate self-evaluation. Individuals will be less likely to use a concealment response (i.e., unethical behavior) as their identity defensive strategy (Petriglieri, 2011).

In addition, self-affirmation theory articulates that most individuals' basic motive is to maintain global integrity of the self, 'which can be defined as the sense that, on the whole, one is a good and appropriate person' rather than perceived worth in specific domains or particular situations (Sherman & Cohen, 2006; Steele, 1988). Self-affirmation provides individuals with useful psychological interventions to deal with identity threat. That is, individuals can draw on alternative positive traits, abilities, or values irrelevant to the activated threat to maintain their global perceived integrity (Cohen & Sherman, 2014). Their flexible self-systems enable them to define success in a way that highlights their idiosyncratic strengths. Thus, self-affirmation may eliminate drastic attitudes or behaviors and promote adaptations to an array of challenges faced over one's lifespan. In summary, affirming alternative identities unrelated to current identity threat will decrease the salience and self-importance of identity threat at stake (Cohen, Aronson, & Steele, 2000; Correll, Spencer, & Zanna, 2004; Fein & Spencer, 1997; Reed & Aspinwall, 1998; Sherman, Nelson, & Steele, 2000). In the current study, to investigate the influences of different kinds of identity threats on individuals' unethical behaviors, the self-affirmation method is also adopted as the control group.

### ***Identity threat from the public sphere***

The needs to appear consistent with others and to project a positive identity image in others' eyes stem from humans' basic need for social adaptation and evolutionary advantage (Colvin, Block, & Funder, 1995; Salancik, 1977; Taylor & Brown, 1988; Trivers, 2000) while other-concern and self-concern are two basic motives for individuals' behaviors (Meglino & Korsgaard, 2004). Therefore, impression management-based identity-defensive responses must be considered when we investigate identity threat (Schoderbek & Deshpande, 1996). Impression management refers to the process through which individuals try to maintain their images in others' eyes (Bolino, 1999).

Individuals tend to resist changing an established course of action that may make their competency look inconsistent with public standards (Hollenbeck, Williams, & Klein, 1989; Salancik, 1977). For example, Dweck and Gilliard (1975) find that children persist longer at an insoluble puzzle when they have made public statements regarding their expected success. Moreover, according to impression management studies (Jones & Pittman, 1982; Leary & Kowalski, 1990), when individuals confront threats of potential harm to their valued identity image in other individuals' eyes, they will use any strategy to maintain the original conditions. Prior studies also find that individuals may set even higher performance goals to show their own identities can still be protected or even enhanced (Taylor & Brown, 1988). In fact, individuals' behavior can result from the motive to verify an important facet of the self through appraisals of others (Winterich, Aquino, Mittal, & Swartz, 2013). According to moral disengagement theory, this strong extrinsic incentive to protect their public identity from threat may also motivate individuals to dehumanize and objectify the unethical instruments they use to pursue these defensive goals (Bandura, 1999).

Moreover, in individuals' goal system, scholars find that the extrinsic goals of *image*, *popularity*, and *financial success* are inherently incompatible with intrinsic goals such as *community feeling* or *affiliation* (Grouzet et al., 2005). These extrinsic goals increase the likelihood of setting higher performance goals and draw individuals' attention to attaining such goals (Van Yperen, Hamstra, & Van der Klauw, 2011). Also from the social value system (Schwartz & Boehnke, 2004), *image* and *achievement values*, which reflect self-enhancement, are diametrically opposed to the *benevolence* value, which reflects self-transcendence (Aquino, Freeman, Reed, Lim, & Felps, 2009). Based on the above studies, we propose that when identity threat is from the public sphere, individuals' main motive underlying their threatened identity is to protect their identity images. Using unethical behavior to conceal their actual identity status can be one such type of strategy (Petriglieri, 2011).

However, when individuals are self-affirmed, they have little need to protect their public identity image in a given domain. This intervention makes the defensive process milder and appears in the form of more rumination (Koole, Smeets, Van Knippenberg, & Dijksterhuis, 1999) and use of various paradigms (Blanton, Pelham, DeHart, & Carvallo, 2001). Thus, self-affirmation lessens the stress in encountering ethical dilemmas and concerns about making a good impression under identity threat. In the current study, we use self-affirmation as the comparative condition opposite to identity threat. Therefore, we posit the following hypothesis:

**Hypothesis 1:** Individuals will engage in more unethical behavior when identity threat is from the public sphere than those without such threat (i.e., after self-affirmation).

### ***Identity threat from the private sphere***

The story differs with regard to the private sphere. Identity threat from the private sphere can establish a sense of high personal salience or self-importance of such identity threat. As Bandura's moral disengagement model suggests individuals need motivations for unethical behavior to disengage from their self-condemnation system (Bandura, 1999; Detert, Treviño, & Sweitzer, 2008; Mazar, Amir, & Ariely, 2008). We suggest that when individuals' private identity is threatened, the self-importance of such identity is actually strengthened. Additionally, individuals tend to use their own sense of the importance of the threatened identity as a reference when evaluating others'

Moreover, as long as identity threat cannot be perceived by significant others, the beneficiaries of the defensive identity will always be the individuals themselves (Leary & Kowalski, 1990). As a traditional Chinese allegory suggests, ‘people cannot just plug their own ears while stealing a bell.’ It is individuals’ own need to accurately evaluate their certain identity status that makes threats to the target identity impossible to defend by unethical behavior (Cialdini & Goldstein, 2004). Therefore, unethical behavior, such as cheating or lying, may become unacceptable in the threatened domain when an individual’s identity is threatened privately.

However, self-affirmation reverses this effect. As self-affirmation theory demonstrates, self-affirmation broadens individuals’ perspectives (Critcher & Dunning, 2015) and perceived sources of self-integrity (Sherman & Cohen, 2006). Individuals who are affirmed by alternative positive traits, abilities, or values other than the currently threatened identity are less likely to be influenced by such threats. Consequently, the restraining effect of private identity threat on unethical behavior will also be decreased in individuals’ global self-worth systems after self-affirmation (Simon, Greenberg, & Brehm, 1995). Individuals in such circumstances may become more intent on engaging in unethical behavior when encountering unethical temptation (e.g., cheating for money). In conclusion, self-affirmation will show its ‘dark side’: impairing self-importance in the threatened identity and increasing the likelihood of engaging in unethical behavior at this time. Thus, we also use self-affirmation as the comparative condition opposite to identity threat from the private sphere. Following this logic, we posit the following hypothesis:

**Hypothesis 2:** Individuals will engage in less unethical behavior when identity threat is from the private sphere than those without such threat (i.e., after self-affirmation).

**Hypothesis 3:** Publicness of identity threat and self-affirmation will interactively predict unethical behavior.

## Overview of Experiments

Two laboratory experiments were conducted to test these hypotheses. Study 1 examined whether individuals’ intention to engage in unethical behavior decreased or increased when public or private competent identities were threatened. Study 2 replicated this result by conducting the experiment in an actual situation and tested whether participants would cheat more to gain money when their competent identities from the private sphere or from the public sphere were threatened. Study 2 also demonstrated that individuals with higher risk preference would be more eas-





**Table 1.** Means and comparison tests for the variables assessed by condition (study 1)

Variables	Conditions	Mean	SD	<i>t</i>	<i>p</i> -Value
Unethical intention	Private				
	Self-affirmation	5.89	1.51		
	Non-affirmation	4.35	1.81	2.36	.02*
	Public				
	Self-affirmation	4.43	1.49		
	Non-affirmation	4.61	1.50	.78	.44

Note: Private: private identity threat, Public: public identity threat.  
\**p* ≤ .05.

on unethical behavior intentions. However, this study may have some limitations. For example, single-item measurement of unethical intention may impede the reliability of actual effects. Therefore, to replicate the causal inferences of the interactive effect by examining the real unethical behavior in experimental settings and to identify the boundary conditions, we conducted the following laboratory experiment.

**Study 2: Laboratory Experiment**

**Participants and design**

To further demonstrate the causal effect of interaction between self-affirmation and identity threat on unethical behavior, we conducted this experiment in the laboratory. The experiment featured a 2 by 2 between-subject design with self-affirmation condition (i.e., affirmation and non-affirmation) and identity threat condition (i.e., public and private) manipulated by a between-subject design. Eighty-eight participants were recruited by one experiment recruitment system, with two of them ultimately excluded due to incomplete results. Thus, 86 participants' results were recorded, consisting of 38 males and 48 females with an average age of 22.17 (*Mean age* = 22.17 years; *SD* = 2.47). These participants were undergraduate or graduate students from almost 20 different schools in one major university in China. They were told that they would receive an average of 40 RMB (almost \$6) for their participation and were randomly assigned to four conditions by one experiment system (i.e., Qualtrics). Our entire experiment was directed automatically by this system to avoid any personal disturbance.

**Procedures**

A cover story was provided on the computer screens to each participant that explained that the study sought to investigate humans' subjective memory (memories of personal experiences) and objective memory (memories of others' experiences). After signing the consent forms and being assured of the confidentiality of their responses, the participants were asked to take a deep breath and empty their minds and were then automatically and randomly assigned to one of four conditions by one computer system.

Following the instructions on the screen, half of the participants first completed the affirmation manipulation by ranking the values and characteristics according to personal importance (or recalled and wrote down the foods they had eaten and drunk in last 48 hr in the non-affirmation condition) following traditional methods (Cohen, Aronson, & Steele, 2000). Then, they received the basic information about the main task and identity threat information on the screens (i.e., not successfully completing the task would mean they were *a person* who did not have high intelligence and would have problems in daily study) to threaten their competency





Figure 1. The unethical intention in different conditions (Study 1)

identity and then told to write a story about campus life without awareness of our purpose. After that, they were told that their rewards would be based on their numbers of correctly answered items in the main task. They were asked to report their results in the main matrix task to obtain the monetary reward and completed some questions about their demographic information. The participants were also told that after they completed all the tasks, we would pay their fees based on their own reported results on the main task (3 RMB per item; i.e., almost \$.4 per item). However, at the end of the experiments, we debriefed them on the aims of our experiment and actually paid them equally.

In our study, instead of directly giving participants negative feedback on an IQ test (Fein & Spencer, 1997; Koole et al., 1999), we chose to set an impossible goal as the identity threat as suggested by goal-setting theory (Locke & Latham, 2002). We told the participants on the computer screen to complete the main task (i.e., the problem-solving matrix) and that ‘this task has a strong predictive power to measure your IQ level, which has been demonstrated to relate to your career success’. The basic standard was successfully completing 15 items within a limited time. While ‘not successfully completing the task means you are a *person* who does not have high intelligence and would have problems in daily study’ was used to activate identity threat. In both conditions, both halves of the participants were directed to complete a problem-solving matrix task (main task), which they had been told would measure their IQ level; the only difference was that in the public identity threat condition, one complementary procedure was added to announce the results to all the participants.

### Manipulation of publicness of identity threat

In the private identity threat condition, to ensure a totally private environment that half of the participants were directed to write down their results on an answer sheet without any obvious mark to identify them. To provide a more confidential workspace, their screens were also covered by some cardboard on the sides to avoid any peeping behavior. After they completed the tasks, they sealed their results in some envelopes and returned them directly to us. In the public identity threat condition, that half of the participants were given unique paper labels with different shapes to mark their unique identities and told that their results (i.e., whether more or less than 15 items) would be announced by sticking their labels on different parts of a whiteboard in the front of the laboratory (i.e., more than 15 items or less than 15 items).

### **Measurement of unethical behavior**

Participants' unethical behavior was measured by a traditional problem-solving matrix adapted from Wiltermuth (2011) and Kouchaki and Wareham (2015) except that we told the participants this task would be used to measure their IQ levels as noted above. After reading the identity threat, the participants had an opportunity for 1 min pretest practice. After the practice, they were presented with 20 matrices of 12 three-digit numbers (e.g., 3.18) on 20 pages (i.e., one matrix per page) and asked to indicate whether they had found the matching pair of numbers that could be summed to 10 within 15 s per matrix. Half of the matrices could be solved (i.e., contained two numbers summing to 10), while the other half could not (i.e., did not contain two numbers summing to 10, unbeknownst to the participants). We counted the number of items for which the participants indicated that they could find the solution but which could not actually be solved as the unethical behavior measurement.

### **Other variables**

Prior studies have demonstrated self-affirmation affect people's attitude towards risky information assimilation (Cohen et al., 2006). As risk preference represents an attitude orientation that anchors whether people would take high risks in order to receive more rewards. Individuals with high risk preference might hold more objective attitudes towards uncertainty and embrace more courage to deal with risky decision. Thus, individuals with high risk preference might be easy to assimilate more perspectives and thus become more easy to be self-affirmed. Therefore, we believe that risk preference can also act as one kind of affirming resources and accelerate the effects of self-affirmation on unethical behavior.

To test the influences of individuals' differences in their decisions, we also measured the participants' risk preference. After the participants completed all the experiment procedures, the computer system asked them to make six decisions about lottery outcomes to elicit their risk preference developed by Holt and Laury (2002). As shown in Table 2, each decision consisted of choosing either a relatively safe option, A, or a relatively risky option, B, and the possibilities of each lottery outcome were carefully manipulated so that the risky choice involved progressively higher expected earnings than the safe choices. The switching points of each participant were recorded and reversed as the participants' risk preference measurement. We also conducted ANOVA tests to ensure that this measurement of risk preference was not influenced by self-affirmation ( $F(1, 85) = .23, n.s.$ ) and different kinds of identity threats ( $F(1, 85) = 1.31, n.s.$ ).

## **Results**

### **Manipulation check**

We first asked the participants whether they believed that their results would be confidential or publicized to check the manipulation of publicness of identity threat. All the participants passed this check. We used the same method to check the self-affirmation manipulation as in study 1. Half of the participants were told to write a story about their campus lives following the instruction, 'Now, close your eyes and picture something in your campus lives, tell a story about it' as a relaxing task immediately after the self-affirmation manipulation. Following the previous process, we used the number of positive affect-related words to ensure manipulations. As expected, half of the participants in the self-affirmation condition tended to have more positive affect ( $M = 4.23, SD = 2.75$ ) than those in the non-affirmation condition ( $M = 2.93, SD = 2.18$ ), ( $t(84) = -2.43, p < .05$ ).

### **Unethical behavior**

A 2 (self-affirmation or non-affirmation) by 2 (identity threat or identity threat) ANOVA revealed a significant interactive effect of self-affirmation and different kinds of identity threats



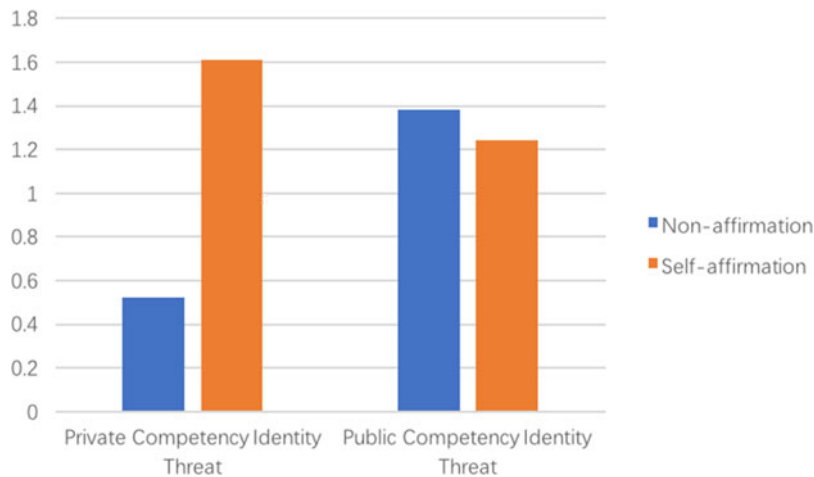


Figure 2. The unethical behavior in different conditions (Study 2)

Table 3. Means and comparison tests for the variables assessed by condition (Study 2)

Variables	Conditions	Mean	SD	<i>t</i>	<i>p</i> -Value
Unethical behavior	<i>Private</i>				
	Risk high				
	Self-affirmation	2.89	2.89		
	Non-affirmation	.29	.49	3.27**	<i>p</i> < .01
	Risk low				
	Self-affirmation	.79	1.25		
	Non-affirmation	.64	.93	.24	n.s.
	<i>Public</i>				
	Risk high				
	Self-affirmation	.33	.82		
	Non-affirmation	1.43	1.40	−1.24	n.s.
	Risk low				
	Self-affirmation	1.07	1.33		
	Non-affirmation	1.36	2.06	−.49	n.s.

Note: Private: private identity threat, Public: public identity threat; Risk high: high risk preference, Risk low: low risk preference.  
\*\**p* ≤ .01.

Discussion

In this study, we extended our study 1 and validated the contingencies of the relationships between identity threat and unethical behavior. We found a significant interactive effect of identity threat and self-affirmation on the participants’ unethical behavior, while this effect was contingent on the publicness of the identity threat. We also demonstrated a three-way interaction of the participants’ risk preference in facilitating the positive and negative relationships between identity threat and unethical behavior, which further verified the effect of individual differences.

## General Discussion

In the current study, we articulate a different story by declaring identity threat a double-edged sword for unethical behavior based on the theory of identity threat and self-affirmation theory with two studies. The results of our study contribute to the literature on both identity threat and self-affirmation theory.

### ***Theoretical contributions***

First, the current study proposes and validates that identity threat acts as a double-edged sword that can both increase individuals' unethical behavior as defensive responses to public identity threat and reduce their unethical behavior when their threatened identity is inherently private. Our findings combine previous theoretical streams by arguing that although cheating can act as a deceptive strategy to buffer temporal threats to individuals' identity in others' eyes, increasing their own competency is the only way for them to maintain identity in their own eyes (Cohen & Sherman, 2014; Leary & Baumeister, 2000; Steele, 1988). More theoretical evidence supporting our arguments can also be derived from research on the effects of other morally incompatible identity (Galperin, Bennett, & Aquino, 2011), while we provide empirical evidence to confirm this effect in experimental settings. Our study raises serious concerns about the actual effects of different kinds of identity threats and provides supplementary supporting evidence of individuals' value or goal system underlying a global identity system in an experimental design. Therefore, our findings provide a new perspective on identity threat theory by incorporating public and private aspects of individuals' identity systems (Aquino & Reed, 2002).

Second, the current study provides new insights into self-affirmation theory by suggesting that self-affirmation can also backfire by identifying negative influences of affirming other kinds of identities, which may not be in accordance with prior findings on the theory of self-affirmation (Steele, 1988; Sherman & O'Leary, 1998; Sherman & O'Leary, 2000; Sherman & O'Leary, 2006; Sherman & O'Leary, 2007; Sherman & O'Leary, 2008; Sherman & O'Leary, 2009; Sherman & O'Leary, 2010; Sherman & O'Leary, 2011; Sherman & O'Leary, 2012; Sherman & O'Leary, 2013; Sherman & O'Leary, 2014; Sherman & O'Leary, 2015; Sherman & O'Leary, 2016; Sherman & O'Leary, 2017; Sherman & O'Leary, 2018; Sherman & O'Leary, 2019; Sherman & O'Leary, 2020; Sherman & O'Leary, 2021; Sherman & O'Leary, 2022).

than alternative ones could actually enhance the escalation of commitment because attempts at self-affirmation at this time will magnify the perceived severity of the failures and strengthen individuals' motives to justify their previous bad decisions (Arndt & Greenberg, 1999; Aronson, Blanton, & Cooper, 1995; Sivanathan, Molden, Galinsky, & Ku, 2008). Our study suggests that even affirming positive traits, abilities, or values in task-irrelevant domains can backfire. We demonstrate that when encountering identity threat, individuals' unethical intent or actual unethical behavior is contingent on the type of publicness of such threat. Specifically, when the motive activated by the threatened identity is from the private sphere, self-affirmation tends to show its 'dark side' by eliminating the positive effect of that specific identity.

Third, our results illustrate that personal and situational factors additively influence individuals' affirmation processes and actual workplace behaviors. This finding of the moderating role of risk preference also contributes to self-affirmation theory. Our study identifies a critical individual difference other than self-esteem (Sivanathan et al., 2008) in self-affirmation processes. As previous studies propose that affirming global self-integrity broadens individuals' perspectives, one possible explanation regarding our results might be that individuals with high risk preference tend to have a broader field of vision, which makes them more easily affirmed. Risk preference represents an attitude orientation that anchors whether individuals will take high risks to receive more rewards (Sitkin & Pablo, 1992). Individuals with high risk preference hold more objective attitudes towards uncertainty and have more courage to make risky decisions. Thus, those with high risk preference may find it easier to assimilate more perspectives. Additionally, the current study may create an opportunity for future research to investigate the function of individuals'

### Limitations and future directions

Some limitations are inevitable in our research. First, the intermediate mechanisms should also be emphasized in future research. In fact, scholars have tried to investigate whether expectation, trivialization, positive mood, moral disengagement, internal or external attribution, and regulatory focus could function as potential mediators in the relationship between self-affirmation and unethical behavior based on established theory (Koole et al., 1999; Sherman & Cohen, 2006). However, all the variables we measured appeared to decline in the current study. A recent study on self-affirmation suggests that individuals who are affirmed tend to broaden their construal towards the focal events and prevent daily adversity as threat (Sherman et al., 2013); we call for more research to investigate whether these psychological factors could help explain the mechanisms through which self-affirmation impacts their unethical behavior. Specifically, more publicness-relevant mediators should be considered. For example, impression management motives can be aroused when public identity is threatened, while self-enhancement motives will be activated when private identity is threatened. Thus, more publicness-relevant mediators may mediate the relationship between identity threat and some kinds of unethical behavior.

Second, Aquino and his colleagues propose that moral identity can function through different angles: moral internalization, which refers to the degree to which moral traits are central to individuals' self-concept, and symbolization, which is the degree to which moral traits are reflected in individuals' actions in the world (Aquino & Reed, 2002). As we used the announcement of the participants' experimental results as a threat to their public competent identity, this manipulation may also have aroused the participants' identity in their own eyes at the same time we aroused their public competent identity. According to our arguments, these two sources of identity threat should be clearly differentiated and manipulated. Future studies could focus on using purer manipulations to identify the different effects of these two identity threats.

Third, there is an alternative explanation for the effect of self-affirmation on individuals' unethical behavior. The moral licensing view (Klotz & Bolino, 2013; Sachdeva, Iliev, & Medin, 2009) refers to a theoretical perspective indicating that when people engage in morally praiseworthy behaviors, they grant themselves a moral license to behave immorally in the future. In our study, self-affirmation may make individuals feel positive about themselves and grant them a license to behave immorally after affirmation. This view may also explain why we do not find differences between the self-affirmation and non-affirmation conditions under public identity threat. That is, the moral licensing effect may neutralize the reducing effect of self-affirmation under the public condition.

Finally, since we conducted this study in China, cultural differences should be considered when testing hypotheses. For example, in China, the 'losing face' problem may also influence our results. Scholars have indicated that the Chinese face consists of a moral face, which refers to society's focus on the integrity of an individual's moral character (Zhang, Cao, & Grigoriou, 2011). Therefore, future studies should integrate or at least control for 'losing face' when investigating impression management and unethical behavior.

### Conclusion

The current study indicates that publicness of identity threat determines whether such threat will lead to more unethical behaviors. We posit this argument because not all kinds of identity threats are detrimental for individuals, especially when they are underlain with inherently moral motives. In conclusion, identity threat may serve as a double-edged sword that can both increase and reduce individuals' unethical behavior in the workplace. The 'dark side' of self-affirmation is also found in our study. When self-affirmation reduces the importance of maintaining a morally compatible identity, individuals can more easily rationalize and justify their unethical behavior.

**Financial support.** The authors acknowledge funding support from the National Natural Science Foundation of China (71232003 and 71573147).

## References

- Aguinis, H., & Bradley, K. J. (2014). Best practice recommendations for designing and implementing experimental vignette methodology studies. *Organizational Research Methods*, 17(4), 351–371.
- Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*. Thousand Oaks, CA: Sage.
- Aquino, K., & Douglas, S. (2003). Identity threat and antisocial behavior in organizations: The moderating effects of individual differences, aggressive modeling, and hierarchical status. *Organizational Behavior and Human Decision Processes*, 90(1), 195–208.
- Aquino, K., Freeman, D., Reed, A., Lim, V. K., & Felps, W. (2009). Testing a social-cognitive model of moral behavior: The interactive influence of situations and moral identity centrality. *Journal of Personality and Social Psychology*, 97(1), 23–141.
- Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83(6), 1423–1440.
- Arndt, J., & Greenberg, J. (1999). The effects of a self-esteem boost and mortality salience on responses to boost relevant and irrelevant worldview threats. *Personality and Social Psychology Bulletin*, 25(11), 1331–1341.
- Aronson, J., Blanton, H., & Cooper, J. (1995). From dissonance to disidentification: Selectivity in the self-affirmation process. *Journal of Personality and Social Psychology*, 68(6), 986–996.
- Ashforth, B. E., & Mael, F. (1989). Social identity theory and the organization. *Academy of Management Review*, 14(1), 20–39.
- Atzmüller, C., & Steiner, P. M. (2010). Experimental vignette studies in survey research. *Methodology*, 6(3), 128–138.
- Bandura, A. (1999). Moral disengagement in the perpetration of inhumanities. *Personality and Social Psychology Review*, 3(3), 193–209.
- Baron, R. A., & Neuman, J. H. (1996). Workplace violence and workplace aggression: Evidence on their relative frequency and potential causes. *Aggressive Behavior*, 22(3), 161–173.
- Baumeister, R. F. (1986). *Identity: Cultural change and the struggle for self*. New York, NY: Oxford University Press.
- Baumeister, R. F. (2010). The self. In R. F. Baumeister & E. J. Finkel (Eds.), *Advanced social psychology: The state of the science* (pp. 139–175). New York, NY: Oxford University Press.
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497–529.
- Belmi, R., Barragan, R. C., Neale, M. A., & Cohen, G. L. (2015). Threats to social identity can trigger social deviance. *Personality and Social Psychology Bulletin*, 41(4), 467–484.
- Bies, R. J. (2001). Interactional injustice: The sacred and the profane. In J. Greenberg & R. Cropanzano (Eds.), *Advances in organizational on(al)*,



- Detert, J. R., Treviño, L. K., & Sweitzer, V. L. (2008). Moral disengagement in ethical decision making: A study of antecedents and outcomes. *Journal of Applied Psychology*, 93(2), 374–391.
- Dweck, C. S., & Gilliard, D. (1975). Expectancy statements as determinants of reactions to failure: Sex differences in persistence and expectancy change. *Journal of Personality and Social Psychology*, 32(6), 1077–1084.
- Erikson, E. (1964). *Insight and responsibility*. New York, NY: Norton.
- Fast, N. J., Burris, E. R., & Bartel, C. A. (2014). Managing to stay in the dark: Managerial self-efficacy, ego defensiveness, and the aversion to individual voice. *Academy of Management Journal*, 57(4), 1013–1034.
- Fein, S., & Spencer, S. J. (1997). Prejudice as self-image maintenance: Affirming the self through derogating others. *Journal of Personality and Social Psychology*, 73(1), 31–44.
- Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology*, 43(4), 522–527.
- Frone, M. R., Russell, M., & Cooper, M. L. (1995). Job stressors, job involvement and employee health: A test of identity theory. *Journal of Occupational & Organizational Psychology*, 68(1), 1–11.
- Galperin, B. L., Bennett, R. J., & Aquino, K. (2011). Status differentiation and the protean self: A social-cognitive model of unethical behavior in organizations. *Journal of Business Ethics*, 98(3), 407–424.
- Gollwitzer, P. M. (1986). Striving for specific identities: The social reality of self-symbolizing. In R. F. Baumeister (Ed.), *Public self, & private self* (pp. 143–159). New York, NY: Springer.
- Griffith, K. H., & Hebl, M. R. (2002). The disclosure dilemma for gay men and lesbians: ‘coming out’ at work. *Journal of Applied Psychology*, 87(6), 1191–1199.
- Grouzet, F. M., Kasser, T., Ahuvia, A., Dols, J. M. F., Kim, Y., Lau, S., ... Sheldon, K. M. (2005). The structure of goal contents across 15 cultures. *Journal of Personality and Social Psychology*, 89(5), 800–816.
- Harber, K. (1995). Sources of validation scale. *Unpublished Scale*.
- Hollenbeck, J. R., Williams, C. R., & Klein, H. J. (1989). An empirical examination of the antecedents of commitment to difficult goals. *Journal of Applied Psychology*, 74(1), 18–23.
- Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, 92(5), 1644–1655.
- James, W. (1950). *The principles of psychology*. New York, NY: Dover.
- Jones, E. E., & Pittman, T. S. (1982). Toward a general theory of strategic self-presentation. In J. Suls (Ed.), *Psychological perspectives on the self* (pp. 231–263). Hillsdale, NJ: Erlbaum.
- Klotz, A. C., & Bolino, M. C. (2013). Citizenship and counterproductive work behavior: A moral licensing view. *Academy of Management Review*, 38(2), 292–306.
- Koole, S. L., Smeets, K., Van Knippenberg, A., & Dijksterhuis, A. (1999). The cessation of rumination through self-affirmation. *Journal of Personality and Social Psychology*, 77(1), 111–125.
- Kouchaki, M., & Wareham, J. (2015). Excluded and behaving unethically: Social exclusion, physiological responses, and unethical behavior. *Journal of Applied Psychology*, 100(2), 547–556.
- Leary, M. R., & Baumeister, R. F. (2000). The nature and function of self-esteem: Sociometer theory. *Advances in Experimental Social Psychology*, 32, 1–62.
- Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, 107(1), 34–47.
- Leavitt, K., Reynolds, S. J., Barnes, C. M., Schilpzand, P., & Hannah, S. T. (2012). Different hats, different obligations: Plural occupational identities and situated moral judgments. *Academy of Management Journal*, 55(6), 1316–1333.
- Locke, E. A., & Latham, G. P. (2002). Building a practically useful theory of goal setting and task motivation: A 35-year odyssey. *American Psychologist*, 57(9), 705–717.
- Mar, R. A., De Young, C. G., Higgins, D. M., & Peterson, J. B. (2006). Self-liking and self-competence separate self-evaluation from self-deception: Associations with personality, ability, and achievement. *Journal of Personality*, 74(4), 1047–1078.
- Marcus-Newhall, A., Pedersen, W. C., Carlson, M., & Miller, N. (2000). Displaced aggression is alive and well: A meta-analytic review. *Journal of Personality and Social Psychology*, 78(4), 670–689.
- Markus, H. (1977). Self-schemata and processing information about the self. *Journal of Personality and Social Psychology*, 35(2), 63–78.
- Markus, H., & Kunda, Z. (1986). Stability and malleability of the self-concept. *Journal of Personality and Social Psychology*, 51(4), 858–866.
- Mayer, D. M., Greenbaum, R. L., Kuenzi, M., & Shteynberg, G. (2009). When do fair procedures not matter? A test of the identity violation effect. *Journal of Applied Psychology*, 94(1), 142–161.
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6), 633–644.
- McQueen, A., & Klein, W. M. (2006). Experimental manipulations of self-affirmation: A systematic review. *Self and Identity*, 5(4), 289–354.
- Meglino, B. M., & Korsgaard, A. (2004). Considering rational self-interest as a disposition: Organizational implications of other orientation. *Journal of Applied Psychology*, 89(6), 946–959.

- Pennebaker, J. W., Chung, C. K., Ireland, M., Gonzales, A., & Booth, R. J. (2007). *The development and psychometric properties of LIWC2007 [LIWC manual]*. Austin, TX: LIWC.net.
- Petriglieri, J. L. (2011). Under threat: Responses to and the consequences of threats to individuals' identities. *Academy of Management Review*, 36(4), 641–662.
- Reed, M. B., & Aspinwall, L. G. (1998). Self-affirmation reduces biased processing of health-risk information. *Motivation and Emotion*, 22(2), 99–132.
- Rozuel, C. (2011). The moral threat of compartmentalization: Self, roles and responsibility. *Journal of Business Ethics*, 102(4), 685–697.
- Sachdeva, S., Iliev, R., & Medin, D. L. (2009). Sinning saints and saintly sinners the paradox of moral self-regulation. *Psychological Science*, 20(4), 523–528.
- Salancik, G. R. (1977). Commitment and the control of organizational behavior and belief. In B. M. Staw & G. R. Salancik (Eds.), *New directions in organizational behavior* (pp. 1–54). Chicago, IL: St. Clair Press.
- Schlenker, B. R., & Weigold, M. F. (1992). Interpersonal processes involving impression regulation and management. *Annual Review of Psychology*, 43(1), 133–168.
- Schoderbek, P. P., & Deshpande, S. P. (1996). Impression management, overclaiming, and perceived unethical conduct: The role of male and female managers. *Journal of Business Ethics*, 15(4), 409–414.
- Schwartz, S. H., & Boehnke, K. (2004). Evaluating the structure of human values with confirmatory factor analysis. *Journal of Research in Personality*, 38(3), 230–255.
- Sherman, D. A., Nelson, L. D., & Steele, C. M. (2000). Do messages about health risks threaten the self? Increasing the acceptance of threatening health messages via self-affirmation. *Personality and Social Psychology Bulletin*, 26(9), 1046–1058.
- Sherman, D. K., Bunyan, D. P., Creswell, J. D., & Jaremka, L. M. (2009). Psychological vulnerability and stress: The effects of self-affirmation on sympathetic nervous system responses to naturalistic stressors. *Health Psychology*, 28(5), 554–562.
- Sherman, D. K., & Cohen, G. L. (2006). The psychology of self-defense: Self-affirmation theory. *Advances in Experimental Social Psychology*, 38, 183–242.
- Sherman, D. K., Hartson, K. A., Binning, K. R., Purdie-Vaughns, V., Garcia, J., Taborsky-Barba, S., ... Cohen, G. L. (2013). Deflecting the trajectory and changing the narrative: How self-affirmation affects academic performance and motivation under identity threat. *Journal of Personality and Social Psychology*, 104(4), 591–618.
- Simon, L., Greenberg, J., & Brehm, J. (1995). Trivialization: The forgotten mode of dissonance reduction. *Journal of Personality and Social Psychology*, 68(2), 247–260.
- Sitkin, S. B., & Pablo, A. L. (1992). Reconceptualizing the determinants of risk behavior. *Academy of Management Review*, 17(1), 9–38.
- Sivanathan, N., Molden, D. C., Galinsky, A. D., & Ku, G. (2008). The promise and peril of self-affirmation in de-escalation of commitment. *Organizational Behavior and Human Decision Processes*, 107(1), 1–14.
- Song, W., Yu, H., Zhang, Y., & Jiang, W. (2015). Goal orientation and employee creativity: The mediating role of creative role identity. *Journal of Management & Organization*, 21(1), 82–97.
- Steele, C. M. (1988). The psychology of self-affirmation: Sustaining the integrity of the self. *Advances in Experimental Social Psychology*, 21(2), 261–302.
- Steele, C. M., & Liu, T. J. (1983). Dissonance processes as self-affirmation. *Journal of Personality and Social Psychology*, 45(1), 5–19.
- Tausczik, Y. R., & Pennebaker, J. W. (2010). The psychological meaning of words: LIWC and computerized text analysis methods. *Journal of Language and Social Psychology*, 29(1), 24–54.
- Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin*, 103(2), 193–210.
- Tedeschi, J. T., & Felson, R. B. (1994). *Violence, aggression, and coercive actions*. Washington, DC: American Psychological Association.
- Tetlock, P. E., & Manstead, A. S. (1985). Impression management versus intrapsychic explanations in social psychology: A useful dichotomy? *Psychological Review*, 92(1), 59–77.
- Treviño, L. K., den Nieuwenboer, N. A., & Kish-Gephart, J. J. (2014). (Un) ethical behavior in organizations. *Annual Review of Psychology*, 65, 635–660.
- Treviño, L. K., Weaver, G. R., & Reynolds, S. J. (2006). Behavioral ethics in organizations: A review. *Journal of Management*, 32(6), 951–990.
- Trivers, R. (2000). The elements of a scientific theory of self-deception. *Annals of the New York Academy of Sciences*, 907(1), 114–131.
- Van Yperen, N. W., Hamstra, M. R., & Van der Klauw, M. (2011). To win, or not to lose, at any cost: The impact of achievement goals on cheating. *British Journal of Management*, 22(1), 5–15.
- Vohs, K. D., Park, J. K., & Schmeichel, B. J. (2013). Self-affirmation can enable goal disengagement. *Journal of Personality and Social Psychology*, 104(1), 14–27.
- Willemuth, S. S. (2011). Cheating more when the spoils are split. *Organizational Behavior and Human Decision Processes*, 115(2), 157–168.

- Winterich, K. P., Aquino, K., Mittal, V., & Swartz, R. (2013). When moral identity symbolization motivates prosocial behavior: The role of recognition and moral identity internalization. *Journal of Applied Psychology*, 98(5), 759–770.
- Zhang, X. A., Cao, Q., & Grigoriou, N. (2011). Consciousness of social face: The development and validation of a scale measuring desire to gain face versus fear of losing face. *The Journal of Social Psychology*, 151(2), 129–149.