TESIS DOCTORAL

2018

Identity Fusion and Metacognition: Nature, antecedents and metacognitive consequences of identity fusion

Borja Paredes Sansinenea

PROGRAMA DE DOCTORADO EN Psicología de la Salud

Ángel Gómez Jiménez, Universidad Nacional de Educación a Distancia

Pablo Briñol Turnes, Universidad Autónoma de Madrid

ACKNOWLEDGEMENTS

At this stage of my life, I cannot help but think that there are some people who are directly or indirectly responsible of me successfully finishing this doctoral dissertation. Allow me to thank them.

First and foremost, I would like to thank my two advisors, Ángel Gómez and Pablo Briñol, for their constant guidance and support throughout these last four years. Words cannot express how lucky I feel for having mentors that are not only willing to provide absolutely brilliant professional and academic advice, but also to share a good time if needed be. Throughout this years, I have genuinely felt that they have my best interests in mind at all times, and that is truly priceless. For this and all the years to come, thank you.

I feel I also owe thanks to Richard E. Petty and William B. Swann for setting up the foundations of what would end up being two ground-breaking lines of research. Together, they merge and nurture this dissertation with rich theoretical background, and I am very much thankful for that.

I would like to extend my gratitude to all the members of the Department of Social and Organizational Psychology at UNED. They have been outstanding partners and professionals.

I also want to thank my lab mates both at UAM and at UNED, for surrounding me with brilliant minds I could learn from, and with friends I could count on. Alexandra, Ana, Blanca, Darío, David, Lucía, María, Mercedes, Miguel Ángel, thank you. Special thanks to Joshua Guyer who, on top of everything else, helped me in thoroughly reviewing most of language quality in this dissertation.

I would also like to thank Alberto Becerra for many things, but specially for being the first one who lighted the spark of academia in me. Without you, none of this would have happened. Thank you.

Thank you to my family, especially to my parents, who always believed in me and supported me, no matter what the circumstances were. A lifetime of appreciation would not be enough.

And finally, thanks to Leticia, my accomplice, my girlfriend, my dream come true. Thank you for waking up with me every morning. Sharing this with you is the best achievement of all.

ABSTRACT

Identity fusion is a relatively novel approach to alignment with groups (Swann, Seyle, Gómez, Morales & Huici, 2009; Swann, Jetten, Gómez, Whitehouse & Bastian, 2012). It entails a visceral feeling of oneness with the group, which is associated with unusually porous, highly permeable borders between the personal and social self. These porous borders encourage people to channel their personal agency into group behavior, raising the possibility that the personal and social self will combine synergistically to motivate pro-group behavior.

On the other hand, metacognition refers to thoughts about our thoughts or judgments (Petty, Briñol & DeMarree, 2007). Primary thoughts are those that occur at a direct level of cognition and involve our initial associations of some object with some attribute. Following a primary thought, people can also generate other thoughts that occur at a second level which involve reflections on the first level thoughts. Meta-cognition refers to these second order thoughts. One of the most essential dimensions of meta-cognitive thought consists of the degree of confidence people place in their thoughts, ranging from extreme certainty to extreme doubt in their validity. Thus, two people might have the same thought, but one person might have considerably greater confidence in that thought than the other, and the greater confidence in the thought, the greater its impact on judgment. This idea is referred to as the self-validation hypothesis (Petty, Briñol & Tormala, 2002).

The goal of this dissertation is to merge these two separate lines of research and explore how confidence on several types of judgments may affect perceived levels of identity fusion and/or attitudes towards extreme pro-group behavior. Chapter 1 introduces these ideas and sets up the several lines of research resulting from them.

Chapter 2 explores how confidence on the impact that one´s actions have on the group drives changes in perceived identity fusion with such group. Chapter 3 tests how strongly-fused individuals´ confidence in their willingness to wait may predict a temporary attenuation in their willingness to engage

in extreme pro-group behavior. Chapter 4 investigates how high thought confidence may stem from matching high levels of pro-group behavior with high levels of fusion, or low levels of pro-group behavior with low levels of fusion.

Finally, Chapter 5 concludes the dissertation by summarizing results and bringing up some limitations to the different lines of research.

RESUMEN

La fusión de la identidad es una aproximación relativamente novedosa a la vinculación con el grupo. Implica un sentimiento visceral de unidad con el grupo, el cual está asociado barreras inusualmente porosas y altamente permeables entre el yo personal y el yo social. Estas barreras porosas alientan a las personas a canalizar su agencia personal en comportamiento grupal, aumentando la posibilidad de que el yo personal y social se combinen sinergísticamente para motivar comportamiento pro-grupal.

Por otro lado, la meta-cognición se refiere los pensamientos sobre nuestros pensamientos o juicios (Petty, Briñol & DeMarree, 2007). Los pensamientos primarios son aquellos que ocurren a un nivel directo de cognición e involucran nuestras asociaciones iniciales de algún objeto con algún atributo. Siguiendo a un pensamiento primario, la gente también puede generar otros pensamientos que ocurren en un segundo nivel el cual involucra reflexiones sobre los pensamientos de primer nivel. La meta-cognición se refiere a estos pensamientos de segundo orden. Uno de las dimensiones más esenciales del pensamiento meta-cognitivo consiste en el grado de confianza que la gente coloca en sus pensamientos, yendo de la certeza extrema a la duda extrema en su validez. Por tanto, dos personas pueden tener el mismo pensamiento, pero una persona puede tener considerablemente mayor confianza en ese pensamiento que otra, y cuanto mayor es la confianza en ese pensamiento, mayor es su impacto en el juicio. Esta idea es referida como la hipótesis de auto-validación (Petty, Briñol & Tormala, 2002).

El objetivo de esta Tesis Doctoral es unir dos líneas de investigación separadas y explorar cómo la confianza en distintos tipos de juicios puede afectar los niveles percibidos de fusión de la identidad y/o las actitudes hacia comportamientos pro-grupales extremos. El Capítulo 1 introduce estas ideas y prepara las diferentes líneas de investigación resultantes de las mismas.

El Capítulo 2 explora cómo la confianza en el impacto que las acciones propias tienen en el grupo lleva a cambios en la fusión de identidad percibida con dicho grupo. El Capítulo 3 prueba cómo

la confianza que los altamente fusionados tienen en su disposición a esperar puede predecir una atenuación temporal en su disposición a realizar conductas pro-grupales extremas. El Capítulo 4 investiga cómo la alta confianza en los pensamientos puede ser consecuencia del encaje entre altos niveles de comportamiento pro-grupal y altos niveles de fusión, o bajos niveles de comportamiento pro-grupal y bajos niveles de fusión.

Finalmente, el Capítulo 5 concluye la Tesis Doctoral resumiendo los resultados obtenidos y mencionando algunas limitaciones de las distintas líneas de investigación.

TABLE OF CONTENTS/ÍNDICE

LIST OF FIGURES

CHAPTER 1

Theoretical review

Introduction

Identity fusion is a visceral feeling of oneness with the group that is particularly predictive of extreme pro-group behaviors. Identity fusion has had a remarkable amount of research success over the last decade. Building on other forms of bonding with a group, identity fusion has managed to research already familiar topics such as pro-group behavior and shed new light on its underlying processes and potential antecedents. Despite having explored a wide arrange of pro-group-related behaviors, identity fusion has yet to examine its relation to attitude change and persuasion. The present doctoral thesis aims to enrich identity fusion literature from a novel perspective: The construct of confidence. From self-efficacy (Bandura, 2006) to attitude certainty (Petty & Krosnick, 1995) to thought confidence (Briñol, Petty & Tormala, 2003) to matching (Evans & Clark, 2012), confidence has had a major impact on the literature of attitudes and persuasion in the last few years, the basic idea being that people are more likely to act on thoughts and/or attitudes which they are more confident about. Basically, our proposal builds on this construct to paint a different portray of the strongly-fused individual. Past literature has provided evidence of strongly-fused individuals having a gut, automatic reaction to engage in extreme pro-group behavior (Gómez & Vázquez., 2015; Swann & Buhrmester, 2015 for a review). Capitalizing on the construct of confidence, this dissertation intends to show a more paused, thoughtful, selective, side of strongly-fused individuals when engaging in pro-group behavior.

One of the main contributions of the current thesis is the merger of two lines of research separated so far: Self-validation and Identity fusion. It also extends knowledge on both fields in many ways. Regarding the Self-Validation hypothesis, this thesis provides new insight on the effects of confidence in one´s evaluations. First, Chapter 2 shows that increasing confidence on the impact one´s actions have on one´s group can have a significant impact on one´s own levels of identity fusion with the group. Chapter 3 shows how confidence in one´s willingness to engage in pro-group behavior can lead to pause rather than action when others have similar intentions and motivations than oneself.

Previous research shows that thought confidence is a predictor of the relationship between thoughts and attitudes, which in turn lead to behavior (Petty, Briñol & Tormala, 2002; Briñol, Petty & Tormala, 2003). Chapter 3 shows that, when it comes to strongly-fused individuals, high levels of confidence can lead to delaying action in order to serve the group´s interests. Finally, Chapter 4 shows that strongly-fused individuals´ confidence in their thoughts is not independent of the situation they are facing. Strongly-fused individuals hold high levels of confidence on actions that match their levels of fusion. That is, extreme sacrifices for the group and/or its members. However, weakly-fused persons hold relatively higher confidence in non-extreme type of sacrifices. As we will see later in this chapter, this matching hypothesis (Huntsinger, 2013; Evans & Clark, 2012) has been applied to other individual-difference scales before, but never to identity fusion.

In regards to identity fusion, this doctoral dissertation provides a novel approach to it, with an overarching theme that is very clear, particularly in Chapters 3 and 4. Strongly-fused persons´ sacrifices are not thoughtless ones. They may decide to attenuate their willingness to self-sacrifice if others are as willing as them to do so and with similar moral motivations, and they may hold lower confidence in sacrifices that are not extreme enough for their degree of bonding and oneness with the group. Hence, they are selective as to what type of sacrifice to engage in, and when to act on their judgments. Chapter 2 tells a somewhat different story, albeit still related to the rest. Chapter 2´s take-home message is that identity fusion, unlike previous literature suggested, can be a type of bond that is sensitive to meaningful manipulations, however incidental these may be. Increasing confidence in one´s actions by learning that they have an impact on the group leads to those actions having an effect on one´s perceived level of identity fusion. This doctoral thesis then shows that identity fusion is sensitive to the impact of one´s actions on the group, the type of sacrifice one is facing, and whether delaying such sacrifice may be more beneficial to the group or not. Additionally, this is the first empirical work that explores the role

of meta-cognition on Identity fusion, as well as being the first empirical work that examines the role of Identity fusion on attitude change and persuasion.

Throughout this introductory chapter we present the theoretical framework of the thesis. In the first section, a brief review of the construct of identity fusion. In the second section, we summarize the main contributions of the matching hypothesis and thought confidence applied to individual-difference variables. In the third and last section, we provide an overview of the thesis.

Identity Fusion

Identity fusion is a visceral feeling of 'oneness' with the group wherein the personal self (characteristics of individuals that make them unique) joins with the social self (characteristics of individuals that align them with a group) and the borders between the two become porous. The result is a potent feeling of connectedness to the group category whereby the integrity of either the personal or social self is not diminished. This allows fused individuals to experience a high sense of personal agency and derive reciprocal strength from group membership. For fused persons, strong relational ties among group members are likely to develop because members are valued by their membership but also due to their idiosyncratic personal qualities. Relational ties are also reinforced by the fact that fused individuals believe they share 'essence' with other group members. For those who are fused, the connection with others and the group tends to be maintained over time and contextual factors are less likely to affect their fusion levels. Together, this reasoning gives rise to the four principles of identity fusion: the agentic-personal-self, identity synergy, relational ties and irrevocability (Swann, Jetten, Gómez, Whitehouse, & Bastian, 2012). The agentic personal-self principle suggests that the personal self can motivate pro-group behavior by channeling personal agency into pro-group action (see also Haggard & Tsakiris, 2009, Swann et al., 2012). To test the idea that heightened arousal fosters feelings of personal agency and thus increases endorsement of extreme pro-group behaviors, researchers experimentally increased physiological arousal through physical exercise. Consistent with their predictions, increased

arousal bolstered endorsement of extreme pro-group behaviors (e.g., sacrificing one's life for the group) among strongly fused individuals but not among weakly fused or highly identified ones (Gómez et al., 2011; Swann, Gómez, Huici, et al., 2010; Swann et al., 2009). Furthermore, in several studies, researchers have assessed participants' self-reported feelings of group-directed agency (e.g., "I am responsible for my group's actions"). Perceptions of personal agency mediated the links between fusion and pro-group behavior (Gómez et al., 2011; Swann et al., 2009). Such findings offer converging evidence for the causal role of the personal self in the pro-group actions of strongly fused persons.

The identity-synergy principle suggests that the personal and social identities of highly fused persons may combine synergistically to motivate extreme pro-group behavior. If so, it should be possible to amplify the pro-group behavior of highly fused persons by activating either their personal or their social self-views. Consistent with this prediction, activating highly fused persons' personal selves (by asking them how they would react to a threat to their personal well-being) or their social selves (by asking them how they would react to a threat to their group) increased their subsequent endorsement of sacrifices for the group (Gómez et al., 2011; Swann et al., 2009). In contrast, highly identified participants displayed more pro-group behavior in response to activating their social selves but not their personal selves.

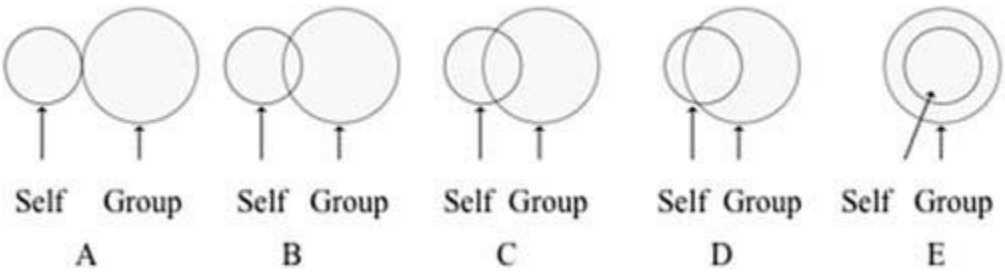
The relational-ties principle recognizes that strongly fused persons care about individual members of the group as well as the abstract collective. For this reason, strongly fused persons should be especially inclined to endorse sacrificing their own lives to save the lives of individual members of the group (e.g., when imperiled by a runaway trolley). In over a dozen studies, fusion predicted self-sacrifice, but identification did not (Swann, Gómez, et al., 2014; Swann, Gómez, Dovidio, et al., 2010). Further support for the relational-ties principle comes from two forms of evidence. First, when strongly fused participants learned that group members might be killed in a hypothetical trolley dilemma, they became upset, and these emotional reactions predicted subsequent endorsement of self-sacrifice for the

group (Swann, Gómez, et al., 2014). Second, self-reported feelings of familial connection to other group members statistically mediated links between fusion and pro-group outcomes (Buhrmester, Fraser, Lanman, Whitehouse, & Swann, 2015; Swann, Buhrmester, et al., 2014). Apparently, strongly fused persons view their group members as fictive family members, and these perceptions motivate them to take extreme actions on the behalf of these individuals.

The irrevocability principle indicates that once people become highly fused with a group, their feelings of fusion will be supported not only by their alignment with the collective but also by their personal selves and ties to other group members. As a result, once strongly fused, people will tend to remain fused. Support for this hypothesis comes from several studies in which researchers had participants complete the fusion scale developed by Gómez et al. (2011) once and then again up to 18 months later. Strongly fused participants (i.e., those scoring in the upper tertile) displayed stable rank orderings over time, whereas the scores of moderately or weakly fused participants fluctuated. This evidence of "irrevocability" among strongly fused persons puts them in sharp contrast to strongly identified persons, whose rank orderings vary with changes in the context.

Three measures of identity fusion have been developed and validated: a single pictorial item (Swann, Gómez, Seyle, Morales, & Huici, 2009), a computer-based variant of the pictorial item (the Dynamic Identity fusion Index; DIFI, Jiménez et al., 2015) and a seven-item verbal scale (Gómez, Brooks et al., 2011). To assess identity fusion, Swann et al. (2009) turned to a pictorial measure, (see Figure 1.1). Based on an instrument originally developed to assess closeness in interpersonal relationships (Aron, Aron, & Smollan, 1992), this pictorial measure depicts the self and the group as separate entities (e.g., two circles) and asks respondents to indicate to what extent the two entities overlap. Swann et al. (2009) opted for a modified version of the measure developed by Schubert and Otten (2002). Participants were asked to indicate which picture best represented the way they perceived their 'relationship with the group'.

Figure 1. *Pictorial Identity Fusion Scale.*



In the pictorial measure of identity fusion, the self was represented with a small circle and the group with a big circle. Only one of the options (the E option) identified participants as fused with the group. In contrast, the A to D options identified non-fused participants. Swann et al. (2009) demonstrated that being fused with the country was unrelated to being fused with other groups, thus ruling out that fusion reflects a stable trait. In an attempt to measure fusion more comprehensively, Gómez, Brooks et al. (2011) developed and validated a verbal measure. Seven items captured the two key sentiments that underlie fusion: connectedness (the powerful union between the personal and social self) and reciprocal strength (the belief that oneself and the group each strengthen one another). In a series of studies, the scale displayed discriminant validity (i.e., it was independent of individual differences in self-concept clarity, self-efficacy, etc. and it was related to, but distinct from, the standard measure of alignment with groups, group identification) as well as convergent validity (i.e., it co-varied with the pictorial item of identity fusion and was especially effective in capturing feelings of agency and invulnerability).

The verbal scale of fusion was a strong predictor of willingness to fight and die for the group and self-sacrifice in responses to intergroup versions of the trolley dilemma. Regarding nomological validity, several studies demonstrated that feelings of agency and invulnerability mediated the effects of fusion on willingness to fight and die for the group, and replicated findings from Swann et al. (2009)

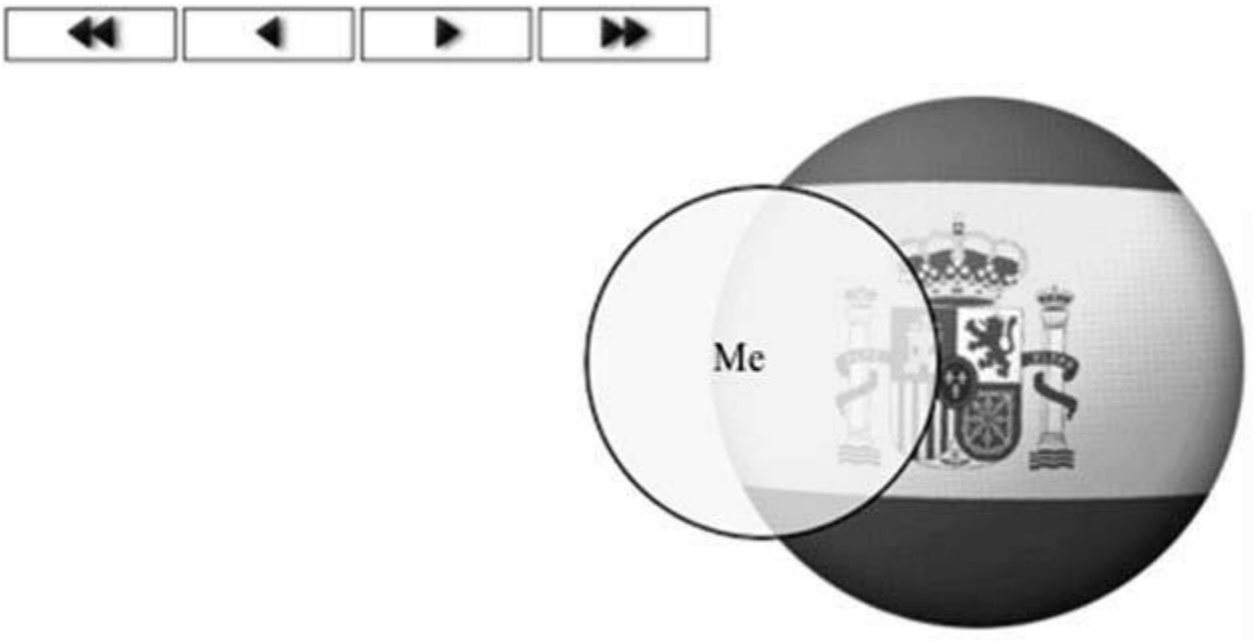
indicating that challenging the personal or the social self-view amplifies endorsement of fighting and dying for the group among fused but not among non-fused participants (See Table 1.1).

Table 1. *Identity Fusion Verbal Scale. (Gómez et al., 2011)*

1. I am one with my group
2. I feel immersed in my group
3. I have a deep emotional bond with my group
4. My group is me
5. I´ll do for my group more than any of the other group members would do
6. I am strong because of my group
7. I make my group strong

Recently, Jiménez et al. (2015) introduced the Dynamic Identity fusion Index (DIFI). The visual design of the DIFI was adapted from the original pictorial item of identity fusion and shows a figure formed by two circles of different sizes in the screen of the computer. The small circle represents 'the self' and it is initially positioned in the left quadrant of the screen. The big circle represents 'the group' and is fixed in the right margin of the screen (see Figure 1.2).

Figure 1. *The Dynamic Index of Identity Fusion, DIFI (Jiménez et al., 2015)*



The DIFI allows collecting responses in two independent fields — namely distance (the separation between the centers of the circles) and overlap (the degree to which the areas of the circles intersect). Overlap was a better indicator of identity fusion than distance. The authors also demonstrated the temporal stability of the DIFI, as well as convergent, discriminant and predictive validity. When the three measures of identity fusion were compared, the verbal scale was the best predictor of endorsement of pro-group behaviors. Interestingly, the DIFI was a better predictor than the pictorial item. Despite the three measures' presentational differences, all three strongly converge on the same construct — identity fusion (range of correlations between .65 and .93). Usage of each measure, then, depends on other criteria. For instance, the verbal measure is generally most recommended because it is the best predictor of pro-group behavior across many groups (Swann, Buhrmester et al., 2014). However, if we expect a binary distribution rather than normal, or if we want to mask the goal of what we are measuring, any of

the pictorial measures is more appropriate. The DIFI is obviously useful only for web-based questionnaires. In any case, the decision of choosing one measure or the other is pragmatic rather than theoretical.

Identity fusion was originally conceived as an explanation for why some individuals fight and even self-sacrifice for a group. Research has consistently demonstrated that identity fusion predicts willingness to fight and die for the group, sometimes also described as endorsement of extreme pro-group behavior (Gómez, Brooks et al., 2011; Gómez, Morales, Hart, Vázquez, & Swann, 2011; Swann, Buhrmester et al., 2014; Swann, Gómez, Huici, Morales, & Hixon, 2010; Swann et al., 2009). Identity fusion also predicts willingness to sacrifice for in-group members on different intergroup and intragroup versions of the trolley dilemma (Gómez, Brooks et al., 2011; Swann, Gómez et al., 2014; Swann, Gómez, Dovidio, Hart, & Jetten, 2010), refusing to leave the group after being ostracized (Gómez, Morales et al., 2011), readiness to deny group wrongdoing (Besta, Gómez, & Vázquez, 2014), quality of life expectations after in-group defeats (Buhrmester, Gómez, Brooks, Morales, Fernández & Swann, 2012) donating real currency to in-group members (Buhrmester, Fraser, Lanman, Whitehouse & Swann, 2014), fusion with transgender predicts surgery of primary sexual features (Swann, Gómez, Vázquez, Guillamón, Segovia & Carrillo, 2015), fusion with a sibling predicts willingness to forgive, to grant favors and to make sacrifices for a sibling (Vázquez, Gómez, Ordoñana, Swann & Whitehouse, 2017), and fusion with one´s romantic partner predicts making extreme sacrifices for such partner (Joo & Park, 2017). Also, sharing negative experiences has shown to predict an increase in levels of identity fusion (Jong, Whitehouse, Kavanagh & Lane, 2015) and increasing accessibility of historic threats to the group has shown to diminish identity fusion and its correlates (Vázquez, Gómez & Swann, 2017).

Of course, identity fusion is not the only construct that has dived into explaining the relationship between group membership and the relationship between thoughts, attitudes and behavior. The field of persuasion and attitude change has also shown interest in explaining these kinds of phenomena and what

mechanisms might explain them, being applied to group membership and racial attitudes (Nosek, Banaji & Greenwald, 2002; McConnell & Leibold, 2001) other forms of out-group derogation (Waldzus & Mummendey, 2004; Johnson, Rowatt & LaBouff, 2012) in-group bolstering (Brewer, 1979), and intergroup relations in general (Tajfel, 1982; Brewer & Kramer, 1985). Of course, although these are topics that have been thoroughly studied, they are not the only instances of research that combine group psychology and attitudes. Recent research has studied how certain group-related features of a source may affect how participants respond to a message advocated by such source. Specifically, Clark & Thiem (2015) examined how group message sources that differ in entitativity/cohesiveness can influence how confident people are in their thoughts about a communication. That is, participants were more confident and had attitudes that were more reflective of their message-related thoughts when source entitativity was high rather than low. Other research has also shown the polarizing effects that entitativity may have on attitudes (Dang, Liu, Ren & Gu, 2017). These findings are consistent with the Self-Validation hypothesis (Petty, Briñol & Tormala, 2002; Briñol, Petty & Tormala, 2003) which states that metacognitive confidence is a key aspect in persuasion and attitude change. Thoughts that are held with higher levels of confidence are more likely to impact attitudes and behavior than thoughts that are held with lower confidence. Despite how influential Self-Validation and thought confidence has shown to be in group-related phenomena (Clark & Thiem, 2015; Dang, Liu, Ren & Gu, 2017) and in attitude change and persuasion (Briñol & Petty, 2009; Briñol, DeMarree & Petty, 2010; Petty & Briñol, 2014, Briñol & Petty, 2015), no research to date has explored the link between Identity Fusion and attitudes in general, and metacognitive confidence in particular. In this thesis, the main focus is set on exploring the different roles that the perceived validity of one´s beliefs may have on levels of identity fusion with one´s group and on willingness to engage in pro-group behavior.

Next, we will dedicate the following session to describe how thought confidence can affect the extent to which thoughts impact attitude change (Self-Validation hypothesis; Briñol, Petty & Tormala,

2003), its overarching theoretical frame (the Elaboration Likelihood Model; Petty & Cacioppo, 1986; Petty & Briñol, 2012) and how it may apply to individual-difference variables in general, and to Identity fusion in particular.

Fundamental processes of attitude change, ELM and thought confidence

Over the past 50 years, researchers have developed numerous theories to account for the psychological processes underlying attitude change (for an historical review, see Briñol & Petty, 2012). Contemporary comprehensive theories of persuasion, such as the elaboration likelihood model (ELM; Petty & Cacioppo, 1986; Petty & Briñol, 2012), the heuristic-systematic model (HSM; Chaiken, Liberman, & Eagly, 1989), and the unimodel (Kruglanski & Thompson, 1999) have been generated to articulate the multiple ways in which variables can affect attitudes in different situations.

In this section we use the ELM to organize the processes of attitude change. Consistent with the ELM, the psychological processes mediating the effects of any given variable (regardless of whether related to the source, the recipient or the context) on attitude change can be placed into a finite set that operate at different points along an elaboration continuum. Specifically, under low thinking conditions, variables can influence attitudes (and other judgments such as a decision to engage in pro-group behavior) by operating as a simple judgment cue or heuristic (e.g., I'll help you because you´re an in-group member). When the likelihood of thinking is relatively high, variables can impact the extent of attitude change by more thoughtful means such as by affecting the direction (valence) of the thoughts that come to mind, serving as a piece of evidence (i.e., an argument) to be scrutinized, or affecting the confidence people have in the thoughts they generated and thus how much the thoughts are relied upon. When elaboration is not constrained to be very low or high, variables can influence attitudes by affecting the amount of thinking that occurs. Thus, the ELM describes several processes by which variables can affect persuasion in different situations.

In specifying the underlying processes of attitude change that we will use to understand the impact of individual differences in general and Identity fusion in particular, the ELM specifies the antecedents and consequences of each of these processes, making clear predictions about when and for whom each of these processes is more likely to operate in changing attitudes. As we have introduced in the previous section, in this doctoral dissertation, the main focus will be on metacognitive confidence, and how it may tie to identity fusion.

Although more processes by which individual differences in general and identity fusion in particular can affect attitude change have been mentioned, we will focus on describing metacognitive ones such as thought confidence, as it is one of the main goals of this doctoral thesis.

Identity fusion may influence attitude change by affecting what people think about their own thoughts, judgments or actions. That is, in addition to affecting the number of thoughts, and valence of thoughts when the amount of thinking is set to be high, variables can also affect meta-cognitive features of the thoughts that are generated such as how much confidence people have in them, how much they like them, or how biasing they are perceived to be. Confidence in thoughts is important because when thoughts are held with greater confidence, people are more likely to use them in forming their judgments (Petty, Briñol, & Tormala, 2002). On the other hand, if people doubt the validity of their thoughts, the thoughts will not have an impact on judgments. Previous literature has shown that individual differences in attitude and thought confidence do exist and have been identified (DeMarree et al., 2017). We anticipate that high scores in identity fusion are associated with high levels of confidence on one´s judgments about the group, possibly stemming from a heightened sense of invulnerability (Gómez, Brooks et al., 2011). Depending on how this confidence is measured and what it is applied to, different effects might be obtained, such as the ones that are shown in Chapters 2, 3 and 4.

It is important to note that identity fusion can be studied not only in isolation but also in combination with other variables. In particular, recipient variables can interact with source, message,

and context variables to produce unique effects. For example, in one study, Haddock, Maio, Arnold, and Huskinson (2008) found that higher levels of the need for affect predicted greater persuasion in response to an affective but not a cognitive based appeal, whereas higher levels of need for cognition predicted greater persuasion in response to a cognitive but not an affective appeal. This suggests that these scales can be used to tap into individual differences in the affective versus cognitive bases of attitudes (i.e., individual difference in attitude basis matched to a message type; see Maio & Haddock, 2015, for a review).

There are many examples in the literature of matching to individual-difference variables. For example, in one study, messages were matched to the Big 5 personality dimensions and it produced an increase in persuasion. Specifically, Hirsh, Kang, and Bodenhausen (2012) personalized persuasive appeals and found more attitude change for matching arguments to each of the Big 5 dimensions than mismatching (e.g., using words related to openness was more persuasive for those with higher scores on openness to experience; using words related to extroversion was more persuasive for extroverts than introverts, and so forth).

In addition to the Big 5, the individual difference variables that have been matched to particular messages or sources in persuasion paradigms include: need for cognition (See, Petty, & Evans, 2009), sensation seeking (Palmgreen, Stephenson, Evertt, Baseheart, & Francies, 2002), independent versus interdependent self-construals (Lee, Aaker, & Gardner, 2000), extraversion (Wheeler, Petty & Bizer, 2005), consideration of future consequences (Strathman, Gleicher, Boninger, & Edwards, 1994), ideal versus ought self-guides (Evans & Petty, 2003), chronic self-discrepancies (Tykocinskl, Higgins, & Chaiken, 1994), one's political ideology (Feinberg & Willer, 2015; Ludeke & Rasmussen, 2016), sensitization versus repression (DeBono & Snyder, 1992), locus of control (Williams-Piehota, Schneider, Pizarro, Mowad, & Salovey, 2004), rational vs. experiential cognitive style (Pacini &

Epstein, 1999), stage of change (Prochaska, DiClemente, & Norcross, 1992), and individual differences in construal level (Cesario, et al., 2004).

In this section, one illustrative example is provided of how matching individual differences to other aspects of the persuasive situation can influence attitudes via the mechanisms specified by the ELM. In considering research on matching, it is important to note that a source or message that matches with one aspect of a person might mismatch with another aspect (e.g., McCann, 2011). The effect on persuasion is likely to depend on which matching aspect has more relative accessibility, importance, and confidence. In any case, when variables match some aspect of oneself, according to the ELM, the match produces attitude change by one of several discrete processes that we have already articulated. That is, matching can influence attitudes by serving as a peripheral cue when elaboration is low, by biasing thoughts, serving as an argument, or affecting thought validation when elaboration is high, and by affecting the amount of information processing when elaboration is not constrained by other variables. As the key example of multiple roles for matching variables to individual differences, we focus on the personality trait of self-monitoring (Snyder, 1974). This particular individual difference has been studied with regard to more possible roles than any other individual difference measure.

However, of particular interest to the current thesis, we will solely focus on how matching can also affect validation processes when people are motivated and able to think about their thoughts and the matching becomes salient after thought generation. For example, Evans and Clark (2012) matched individuals who varied in their scores on the self-monitoring scale to whether a communication source was high in either expertise or attractiveness. Prior research on self-monitoring had shown that people high in self-monitoring are particularly interested in image-related information and thus are drawn to attractive sources whereas people low in this trait are particularly influenced by quality or merit information and are particularly drawn to expert sources (DeBono & Harnish, 1988). The critical result of the Evans and Clark study was that high self-monitors relied on their thoughts to the message more

when they learned after processing it that the source was attractive rather than expert, but low self-monitors relied on their thoughts to the message more when they learned the source was expert rather than attractive. Thus, when message recipients learned about the message source after they had processed a message, matching the source to the recipient type increased thought use. Chapter 3 of this dissertation uses a similar person by situation matching approach to identity fusion and self-sacrifice, predicting and finding that persons scoring high in identity fusion rely more on thoughts generated on extreme pro-group behaviors, whereas persons scoring low in identity fusion rely more on thoughts generated on non-extreme pro-group behavior.

In summary, the accumulated research suggests that matching of an information (e.g., message, self-generated thoughts) to some characteristic of the recipient can influence attitudes by serving multiple roles depending on the circumstances. Consistent with the ELM, the psychological processes mediating the effects of matching on attitude change fall into a finite set that operate at different points along an elaboration continuum, among which we are interested on thought confidence (for complete reviews on matching, see, e.g., Avnet, Laufer, & Higgins, 2013; Briñol & Petty, 2006; Maio & Haddock, 2015, Salovey & Wegener, 2003; Teeny, et al., 2017). Chapter 4 of this doctoral dissertation will focus on how identity fusion can match with certain features of the situation to increase thought confidence.

Overview of the present thesis

The present thesis aims to address some of the issues aforementioned through several studies encapsulated in three thematic packages, theoretically and methodologically independent, but traversed by a common concern: Exploring the role of Identity fusion on confidence and examining a novel aspect of Identity fusion´s relation with pro-group behavior, namely, a more calculated, selective, strategic route towards pro-group behavior through the several validating roles of confidence.

The first package (Chapter 2) seeks to investigate whether confidence in the impact of one´s actions on the group (e.g., personal agency) mediates the relationship between identity fusion and pro-

group behavior. In order to empirically test whether the perception of agency may be an antecedent of fusion, three studies were carried out. Across three studies, participants donated a fictional amount of money to an in-group member and then received information about what the recipient was planning on doing with the donation. Participants were then assigned either to a beneficial-to-the-group condition, in which they were told that the recipient spent the money to help other in-group members, or to a beneficial-to-the-person condition, in which they were told that the recipients had spent the money on themselves. Lastly, Identity fusion with the in-group was measured. Results show a significant two-way interaction between Donation and Donation Consequence on Identity fusion (Studies 1, 2 & 3), Willingness to fight and die (Study 2), and on confidence in the impact of one´s actions (e.g., Personal Agency) (Study 3). Specifically, for those assigned to the Group-beneficial condition, Identity fusion scores were higher for participants who donated larger amount of money. For those assigned to the Person-beneficial condition, there was no relation between donation and fusion. This effect emerged only when participants donated fictitious money to in-group members (Study 2). This line of research shows, for the first time, how Identity fusion with a group can be affected through the confidence on the impact of one´s control of the actions that strongly fused individuals have on their group (e.g., personal agency).

In the second package (Chapter 3), we examine how strongly-fused individuals react when they learn the motivations driving other strongly-fused individuals' pro-group behaviors. Three studies were conducted in two countries (Spain and USA). Results showed an attenuation effect in their willingness to fight and die for in-group members when strongly-fused individuals learned that other participants displaying their same level of visceral commitment to the group would self-sacrifice for morally relevant reasons, compared to pragmatic reasons. The final study suggests that the attenuation in willingness to fight and die observed in those strongly fused can be interpreted as a sign of confidence to delay their self-sacrifice rather than a need to be unique.

In the last package (Chapter 4) we examine whether matching Identity fusion with level of sacrifice influences thought confidence and in turn affects attitude change. These data revealed an interactive rather than additive effect of Identity fusion and level of sacrifice such that people relied on their thoughts more when they matched their nature than when they did not. The results showed that matching individuals' dispositions related to their visceral connection to the group (level of identity fusion) to the behaviors they are aligned with (level of self-sacrifice for in-group members) produced greater thought use compared to mismatching these variables. Specifically, strongly fused participants relied more on thoughts about relatively high sacrifices and weakly fused participants relied more on thoughts about relatively low sacrifices. In addition to identifying a new role for identify fusion, the present research provides evidence for the underlying process by which this matching effect occurred. Matching affected attitude change by influencing the confidence people had in the thoughts they generated. The present research has implications for Identity fusion, thought validation, matching, attitudes change and beyond.

In the final section (Chapter 5), we discuss the relevance and applicability of these findings for Identity fusion and Confidence, as well as the shortcomings and implications of this thesis for future research.

This doctoral dissertation has been structured by combining three different manuscripts in publication format. Although only one of them has been published (Chapter 3) and all three have a common narrative that link them together (identity fusion and metacognitive confidence) each one is written like a separate manuscript that could stand on its own. Furthermore, one chapter has already been submitted and accepted in an international journal (Self and Identity) as a stand-alone paper and the other two will be submitted for publication after this dissertation is defended. Hence, this may lead to some repetitiveness in initial construct descriptions between chapters, as each briefly describes identity fusion and the different forms of confidence that will be applied.

CHAPTER 2

Confidence on the impact of one´s actions on the group (e.g., Personal Agency) as an antecedent of Identity fusion

Abstract

The present work seeks to investigate whether agency mediates the relationship between identity fusion and pro-group behavior. In order to empirically test whether the perception of agency may be an antecedent of fusion, three studies were carried out. Across three studies, participants donated a fictional amount of money to an in-group member and then received information about what the recipient was planning on doing with the donation. Participants were then assigned either to a beneficial-to-the-group condition, in which they were told that the recipient spent the money to help other in-group members, or to a beneficial-to-the-person condition, in which they were told that the recipient had spent the money on themselves. Lastly, Identity Fusion with the in-group was measured. Results show a significant two-way interaction between Donation and Donation Consequence on Personal Agency (Study 3), Identity Fusion (Studies 1, 2 & 3), and Willingness to fight and die (Study 2). Specifically, for those assigned to the Group-beneficial condition, Identity Fusion scores were higher for participants who donated larger relative to smaller amount of money. For those assigned to the Person-beneficial condition, there was no relation between donation and fusion. This effect emerged only in interactions with in-group members (Study 2). This line of research shows for the first time how Identity Fusion with a group can be affected quasi-experimentally through personal agency.

Introduction

Identity Fusion is a remarkable form of group alignment that predicts a wide range of pro-group behaviors, from willingness to fight and die for the group (Swann, Gómez, Seyle, Morales & Huici, 2009) to self-sacrificing for other in-group members in simulated trolley dilemmas (Swann, Gómez, Dovidio, Hart & Jetten, 2010), denying the group´s wrongdoing (Besta, Gómez & Vázquez, 2014), donating personal funds to a needy in-group member (Swann, Gómez, Huici, Morales & Hixon, 2010) or even genital sex reassignment surgery if the person if strongly-fused with the cross-gender (Swann, Gómez, Vázquez, Guillamón, Segovia & Carrillo, 2015).

Identity fusion theory explains such high willingness to engage in sometimes extreme pro-group behavior by resting on four main theoretical principles (Swann, Jetten, Gómez, Whitehouse & Bastian, 2012). The irrevocability principle, the relational ties principle, the identity synergy principle, and the agentic-personal-self principle. The irrevocability principle states that once fused, people will tend to remain fused. This principle speaks to the strength, resistance and stability of identity fusion as a form of bond with the group. The relational ties principle states that strongly-fused individuals will value other in-group members not only by virtue of their representativeness of the group, but also for their unique personal characteristics. The identity synergy principle holds that the personal and social identities of strongly fused persons may combine synergistically to motivate pro-group behavior. Finally, the agentic-personal-self principle holds that strongly-fused individuals will tether their personal agency to their group´s agendas. The current research project will focus on offering a novel approach to personal agency as an explanatory mechanism of changes in identity fusion. Specifically, this research aims at impacting levels of identity fusion by virtue of affecting the extent to which participants are confident that their actions have an impact on the in-group. That is, by affecting participants´ sense of personal agency.

Although personal agency is generally associated with a sense of individual intentionality and influence on one´s own environment (Bandura, 2006), identity fusion capitalized on this concept to highlight conceptual differences between Social Identity theory (Tajfel & Turner, 1979) and identity fusion theory (Swann et al., 2009; Swann et al., 2012). Whereas Social Identity theory explained pro-group behavior by virtue of the higher accessibility of social identity over personal identity, identity fusion theory posits that people do not need to abdicate their personal selves to feel a strong bond with the group, nor to engage in pro-group behavior (Swann et al., 2012). Swann, Gómez, Huici, Morales & Hixon (2010) provided empirical evidence to this claim. They predicted that, if indeed identity fusion (controlling for identification) could lead to pro-group behavior through activating the personal rather than the social self, a heightened sense of confidence on the impact that one´s actions have onto the group (e.g., personal agency) should increase the relation between fusion and pro-group behavior. Indeed, across four experiments, they found that increasing physiological arousal through exercise increased personal agency, which in turn promoted endorsement of dying for the group and other pro-group actions among strongly-fused persons (Swann et al., 2010).

After this compelling case for the centrality of this form of confidence in the relationship between identity fusion and pro-group behavior, Swann et al (2012) described the aforementioned four principles of identity fusion theory, among which was the agentic-personal-self principle.

Swann et al., (2010) research, however, presented two questions whose answers may pose a valuable contribution for identity fusion theory, namely: First, in all experiments carried out by Swann et al (2010), identity fusion was measured before the agency manipulation (that was modified by increasing physiological arousal), leaving the question of whether identity fusion measures would be sensitive to such manipulations or not. Secondly, although agency was manipulated using three different inductions, all of them involved increasing physiological arousal through performing different physical exercises that had no relation with the rest of the measures participants were asked to complete, leaving

the question of whether personal group-related agency can be manipulated through more cognitive and group-related means.

Given that personal agency *induced through physiological means* can impact pro-group behavior if it is introduced *after measuring identity fusion,* the goal of the present research was to examine the extent to which personal agency *induced through non-physiological means* can also impact perceived oneness with the group if it is introduced *before measuring identity fusion*. Specifically, we aim at affecting participants´ feelings of confidence in the impact of one´s actions onto the group and, in turn, identity fusion by manipulating whether the consequences of participants´ pro-group behavior are beneficial to the group or not.

In trying to predict how this form of confidence (e.g., personal agency) might influence identity fusion, several outcomes are possible: First, increasing or decreasing agency could affect participants´ perceived levels of identity fusion with their in-group in a consistent manner. Given that the experimental manipulation is a novel one, a second possibility is that our manipulation may have an impact on fusion, but not necessarily on personal agency, or vice versa. A third possibility is that neither agency nor fusion may be affected. Indeed, identity fusion has shown to be a remarkably stable form of alignment with the group (Swann et al., 2009; Swann et al., 2012), therefore one can expect it not to be particularly sensitive to incidental laboratory manipulations.

The current set of experiments was designed to explore these different possibilities. Specifically, our primary goal was to examine whether an experimental manipulation of one´s positive or negative impact on the group matters for one´s own levels of identity fusion with such group. Assuming that one´s impact on the group plays a role, our second goal was to examine *under which conditions* (interacting with in-group vs out-group member) the effect was more likely to occur, whether those changes were predictive of relevant behavioral intentions, and the extent of its impact. Our final goal

was to potentially find an empirical link between confidence (e.g., personal agency) and the effect found in the two previous studies.

Study 1

Study 1 tested the effects of being informed about the consequences of one´s pro-group behavior on levels of identity fusion. Our goal was to provide initial evidence of whether providing information that participants´ pro-group behavior was benefitial (vs not benefitial) to the group had any effect on identity fusion with such group.

Method

Participants and Design.

Eighty-one students from Universidad Nacional de Educación a Distancia, UNED, participated online in exchange for course credit. Sample size was determined based on the number of participants collected during the week in which the study was posted. Participants were randomly assigned to one of two (Beneficial to the group vs. to the person) between-subjects conditions, with donation to an in-group member as a continuous predictor, and Identity fusion with the country as the dependent variable.

Procedure and Materials.

Participants were first told that they would be participating in a game that offered the possibility of earning fictitious money. This game was a conceptual adaptation of the Dictator Game (Cason & Mui, 1998). The game consisted of two players (Player A and Player B). Player A´s role was to first play the game and then give a fictitious donation to Player B. Player B´s role was to decide what to do with that donation from a predetermined set of options. Participants were told that they would randomly be assigned to play either Player A or Player B, although in actuality all participants were assigned the role of Player A.

Player A had 10 fictitious Euros at their disposal, and could donate any amount within those 10 euros to Player B, who would receive that donation multiplied by three. Player B was described as an in-group member (i.e., a Spanish citizen) and was simulated by the computer, though participants were led to believe that they were playing with a real person. After indicating the amount of money they wished to donate, participants were informed that Player B could decide how to spend that donation. Participants were randomly assigned to one of two conditions (Beneficial to the group vs. to the person use of donation, see below) in which they were told how Player B had decided to use their donation. Participants then completed the dependent measures and were debriefed.

Predictor Variable.

Donation. Participants indicated the amount of money they were willing to donate to Player B (minimum = 0 €, maximum = 10 €).

Independent Variable.

Donation Consequence. After indicating the amount of fictitious money donated, participants were assigned to either the Group-beneficial or Person-beneficial donation-consequence condition. In the Group-beneficial condition, participants were informed that the recipient had decided to spend the money on other in-group members (e.g., NGO´s, charity causes, and give money back to participant). In the Person-beneficialcondition, participants were informed that the recipient had decided to spend the money on themselves (e.g., personal image, consume alcoholic beverages and gambling).

Dependent Variable.

Identity fusion. Identity fusion with the country was measured using the seven-item verbal scale (Gómez, Brooks, et al., 2011). Reliability was $\alpha = .84$.

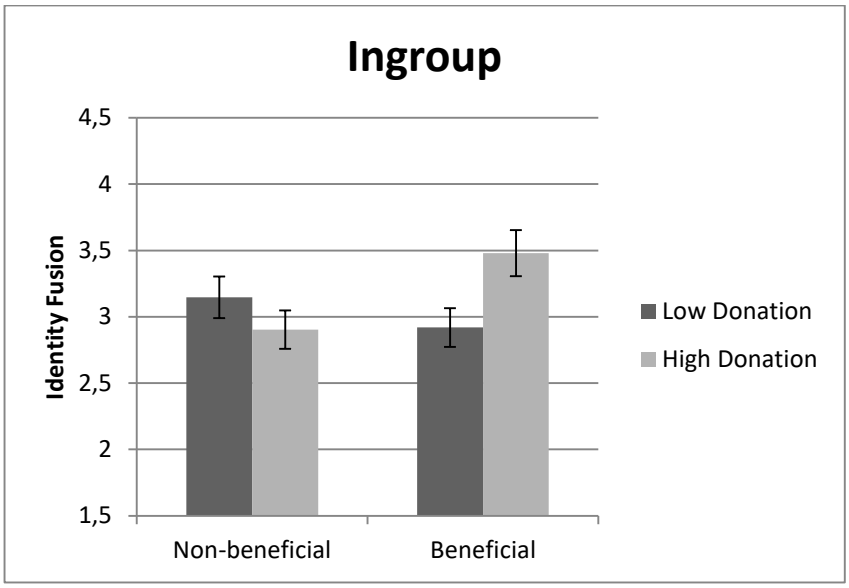
We report all measures, manipulations, and exclusions in these three studies. Furthermore, we report all the studies conducted in this line of research.

Results

Donation. The amount of money donated was submitted to a one-way ANOVA with Donation Consequence (Group-beneficial vs Person-beneficial) as the predictor. Donation amount did not change as a function of the Donation Consequence manipulation, $F(1, 79) = 1.124$, $p = 0.292$.

Identity fusion. Multiple regression was used to examine the impact of Amount of Donation (centered), Donation Consequence (effect coding: -1, 1), and the interaction of these variables on identity fusion with Spain. The regression revealed a significant two-way interaction between Donation and Donation Consequence, $B = -0.07$, $t(76) = 2.21$, $p = 0.029$, 95% CI: 0.0074, 0.1399. As seen in Figure 1, among participants assigned to the Group-beneficial condition, Identity fusion scores were higher for participants who donated larger relative to smaller amount of money, $B = 0.10$, $t(76) = 2.15$, $p = 0.034$, 95% CI: 0.0075, 0.1978. In contrast, among participants assigned to the Person-beneficial condition, the amount of money donated was not related to reported levels of Identity fusion. $B = -0.04$, $t(76) = -0.96$, $p = 0.337$, 95% CI: -0.1364, 0.0473. Put differently, among participants who donated more than average (+1SD), those assigned to the Group-beneficial condition reported significantly higher levels of identity fusion with Spain than those assigned to the Person-beneficial condition, $B = .28$, $t(76) = 2.26$, $p = 0.026$, 95% CI: 0.0342, 0.5426. However, among participants who donated less than average (-1SD), no difference in reported levels of identity fusion were reported between those assigned to the Group-beneficial vs. Person-beneficial condition, $B = -.11$, $t(76) = -0.88$, $p = 0.378$, 95% CI: -0.3687, 0.1415.

Figure 1. Identity fusion scores as a function of Amount of Donation to an in-group member and Donation Consequence (Study 1).



Discussion

Study 1 indicated that a higher amount of pro-group behavior was associated with higher levels of identity fusion among those whose pro-group behavior had Group-beneficial consequences for the group. In contrast, no difference in identity fusion was found among those whose pro-group behavior was non-beneficial for the group. That is, participants´ pro-group behavior affected their fusion levels only when their behavior turned out to be used for the benefit for the group. This might be initial evidence for the effect of personal agency on identity fusion.

One of the questions that arises when learning the results from Study 1 is whether the resulting identity fusion scores were as predictive of pro-group behavior as previous research has found (fusion could change but its correlates do not). It may be that differences in fusion scores found in Study 1 are merely an experimental artifact that are not as predictive of classic outcome measures of identity fusion, such as willingness to fight and die for the in-group.

When interpreting these results, it is important to note that all participants were interacting with an in-group member. Although identity fusion has shown to mostly explain intra-group rather than between-group processes (Swann et al., 2012; Gómez & Vázquez, 2015), it is unknown whether being informed of the consequences of aiding an out-group member can have consequences on the perceived bond one has with the in-group. If the effects found in Study 1 were replicated using out-group members in the fictitious donation game, we would have stronger evidence to conclude that personal agency with the in-group is not being the driving force of such effect, given that the interaction and its consequences would be taking place with a member of the out-group. In order to replicate the effect, to find out how predictive fusion was under these circumstances and to explore these potential boundary conditions, Study 2 used a similar design and included a condition in which the game was played with an out-group member, as well as a measure of willingness to fight and die for the group.

Study 2

Study 2 tested whether learning the Group-beneficial vs Person-beneficial consequences of one´s donation would impact one's own level of identity fusion. Of course, it is possible to make a donation to other than in-group members (e.g., out-group members). Thus, an important goal of study 2 was to compare the extent to which donations given to in-group vs out-group members might differentially influence levels of identity fusion with the in-group. Given the intra-group nature of identity fusion (Swann et al., 2012) we expect donations to out-group members not to impact identity fusion with the in-group. Study 2 also included a measure of willingness to fight and die for the in-group in order to find out to what extent experimentally manipulated identity fusion scores were predictive of classic outcome measures of pro-group behavior.

Method

Participants and Design.

Two-hundred and nine students from Universidad Nacional de Educación a Distancia, UNED, participated online in exchange for course credit. Sample size was determined based on the number of participants collected during the two-week period in which the study was posted. Participants were randomly assigned to a 2 (Donation Receiver´s Group: Ingroup vs Outgroup) x 2 (Donation Consequence: Group-beneficial vs Person-beneficial) between-subjects factorial design, with donation to as a continuous predictor, and identity fusion with the in-group and willingness to fight and die as the dependent variables.

Procedure and Materials.

The procedures were identical to Study 1, except for the addition of the receiver´s group independent variable. Half of the participants played the donation game with a (fictitious) player from their in-group, whereas the other half played the donation game with a (fictitious) player from an out-group.

Predictor Variable.

Donation. This variable was measured identically to Study 1.

Independent Variables.

Donation Receiver´s Group. Participants were assigned either to the in-group condition, in which they were told that Player B belonged to the in-group (i.e., Spain), or to the out-group condition, in which they were told that Player B was a foreigner.

Donation Consequence. This variable was manipulated identically to Study 1.

Dependent Variables.

Identity fusion. Identity fusion with Spain was measured using the seven-item verbal scale (Gómez, Brooks, et al., 2011),$\alpha = .88$.

Willingness to fight and die. All participants completed the willingness to fight and die scale (e.g. I would fight someone physically threatening another member of my country", "I would sacrifice

my life if it saved another group member´s life"), taken from Swann et al., (2009). Responses were recorded on a 1 (completely disagree) to 7 (completely agree), α = .85.

Results

Donation. The amount of money donated was submitted to a two-way ANOVA with Donation Consequence (Group-beneficial vs Person-beneficial) and Donation Receiver´s Group (In-group vs Out-group) and their interaction as the predictors. Donation amount did not change as a function of the Donation Consequence manipulation, the Donation Receiver´s Group manipulation, or their interaction, $F(1, 217) < 2.414$, $ps > 0.122$.

Identity fusion. Multiple regression was used to examine the impact of Donation (centered), Donation Receiver´s Group (effect coding: -1, 1) and Donation Consequence (effect coding: -1, 1), as well as the interaction of these three variables on identity fusion with Spain. Central to our predictions, a significant three-way interaction was found $B = 0.16$, $t(201) = 2.16$, $p = 0.031$, 95% CI: 0.0143, 0.3060, such that identity fusion was only affected for those participants who were assigned to play with a fictitious in-group member. As Figure 2B shows, no effect on identity fusion emerged for participants assigned to play with a fictitious out-group member, $Bs < 0.007$, $ts(201) < 0.101$, $ps > 0.91$.

Replicating the interaction found in Study 1, the regression revealed a significant two-way interaction between Donation and Donation Consequence for participants assigned to the in-group Donation Receiver condition, $B = 0.30$, $t(201) = 3.24$, $p = 0.001$, 95% CI: 0.1206, 0.4941. As seen in Figure 2A,among participants assigned to the Group-beneficial condition, identity fusion scores were higher for participants who donated larger relative to smaller amount of money, $B = 0.13$, $t(201) = 2.15$, $p = 0.032$, 95% CI: 0.0115, 0.2570. In contrast, among participants assigned to the Person-beneficial condition, identity fusion scores were lower for participants who donated larger relative to smaller amount of money. $B = -0.17$, $t(201) = -2.42$, $p = 0.016$, 95% CI: -0.3138, -0.0323. Put differently, with regard to participants assigned to the in-group receiver condition, among those who donated more than

average (+1SD), those assigned to the Group-beneficial condition reported significantly higher levels of identity fusion with Spain than those assigned to the Person-beneficial condition, B = .99, $t(201)$ = 2.90, p = 0.004, 95% CI: 0.3213, 1.6772. However, among participants who donated less than average (-1SD), no difference in reported levels of identity fusion were reported between those assigned to the Group-beneficial vs. Person-beneficial condition, B = -.54, $t(201)$ = -1.58, p = 0.115, 95% CI: -1.2177, 0.1340.

Figure 2A. *Identity fusion scores as a function of Amount of Donation to an in-group member and Donation Consequence (Study 2).*

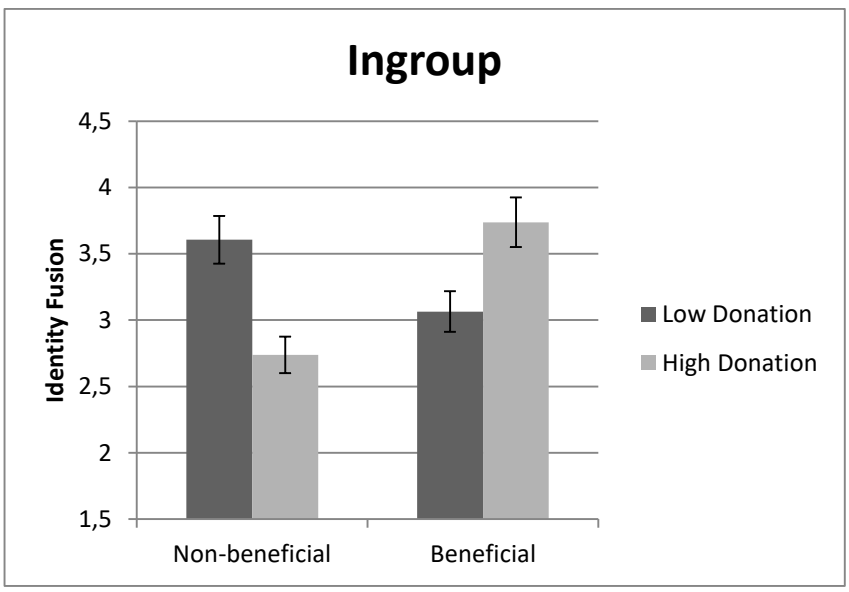
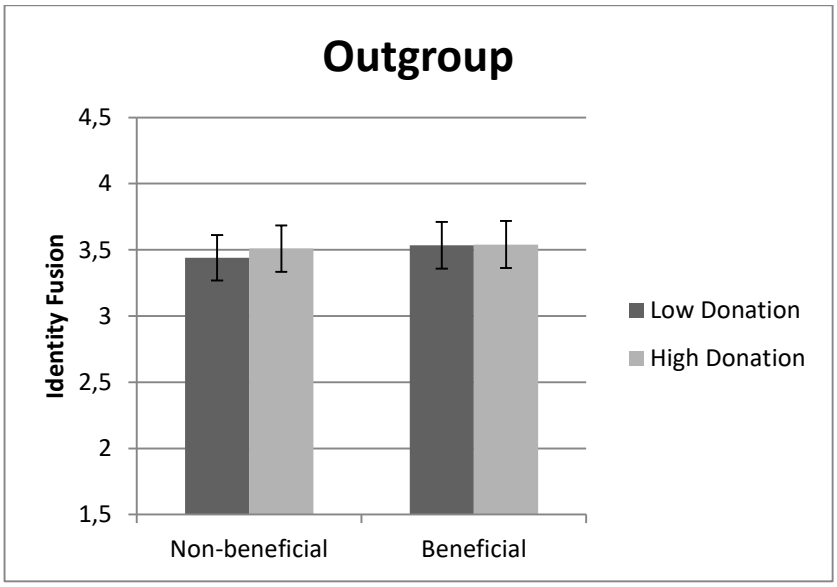


Figure 2B. *Identity fusion scores as a function of Amount of Donation to an out-group member and Donation Consequence (Study 2).*



Willingness to fight and die. Multiple regression was used to examine the impact of Donation (centered), Donation Receiver´s Group (effect coding: -1, 1) and Donation Consequence (effect coding: -1, 1), as well as the interaction of these three variables on willingness to fight and die for Spain. Central to our predictions, a marginally significant three-way interaction was found $B = 0.11$, $t(201) = 1.79$, $p = 0.074$, 95% CI: -0.0113, 0.2420, such that willingness to fight and die was only affected for those participants who were assigned to play with a fictitious in-group member. As Figure 3B shows, no effect on willingness to fight and die emerged for participants assigned to play with a fictitious out-group member, $Bs < 0.007$, $ts(201) < 0.075$, $ps > 0.93$.

Similar to the interaction found in Study 1 on identity fusion, the regression revealed a significant two-way interaction between Donation and Donation Consequence for participants assigned to the in-group Donation Receiver condition, $B = 0.23$, $t(201) = 2.89$, $p = 0.004$, 95% CI: 0.0760, 0.4005. As seen in Figure 3A, among participants assigned to the Group-beneficial condition, there was no relation between amount of donated money and willingness to fight and die, $B = 0.06$, $t(201) = 1.14$, $p = 0.252$,

95% CI: -0.0445, 0.1686. In contrast, among participants assigned to the Person-beneficial condition, willingness to fight and die scores were lower for participants who donated larger relative to smaller amount of money. *B* = -0.17, *t*(201) = -2.84, *p* = 0.005, 95% CI: -0.2985, -0.0539. Put differently, with regard to participants assigned to the in-group receiver condition, among those who donated more than average (+1SD), those assigned to the Group-beneficial condition reported significantly higher levels of willingness to fight and die for Spain than those assigned to the Person-beneficial condition, *B* = .70, *t*(201) = 2.34, *p* = 0.020, 95% CI: 0.1117, 1.2965. However, among participants who donated less than average (-1SD), those assigned to the Group-beneficial condition reported marginally lower levels of willingness to fight and die for Spain than those assigned to the Person-beneficial condition, *B* = -.49, *t*(201) = -1.65, *p* = 0.099, 95% CI: -1.0818, 0.0951.

Figure 3A. *Willingness to fight and die scores as a function of Amount of Donation to an in-group member and Donation Consequence (Study 2).*

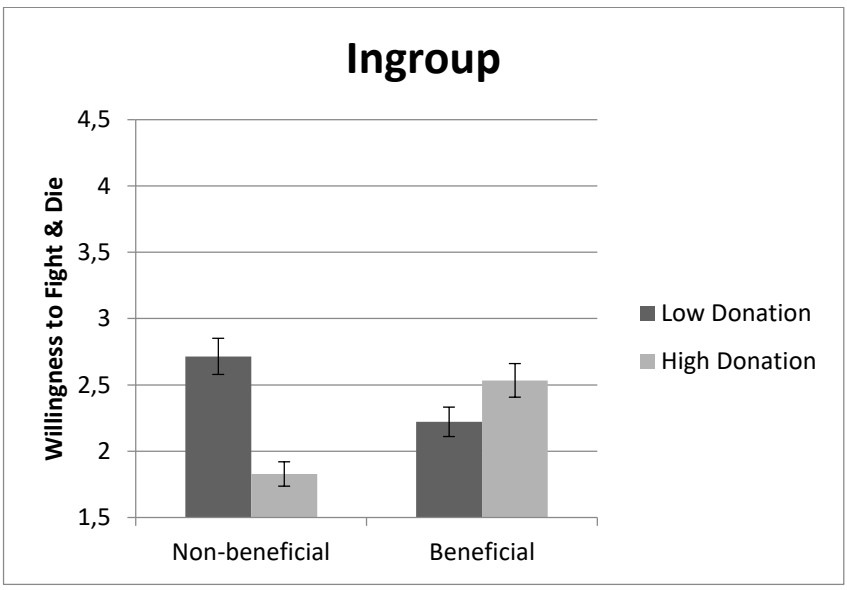
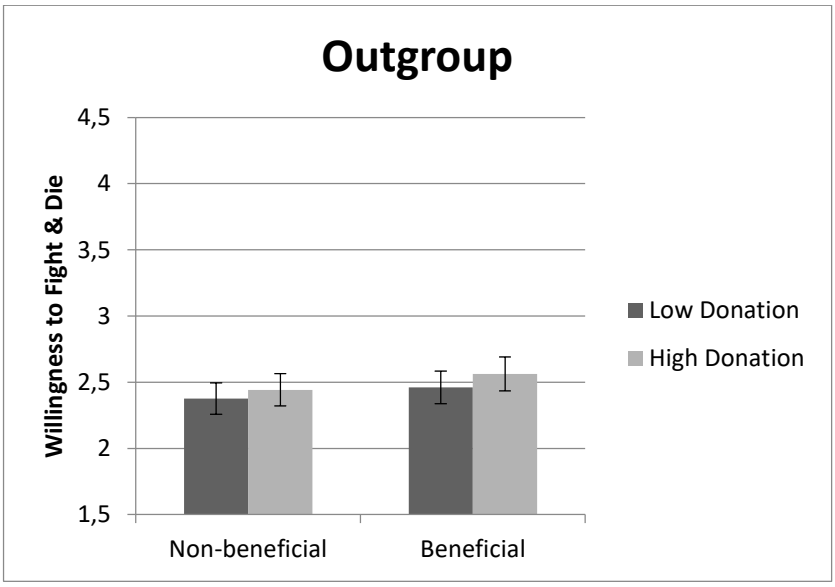


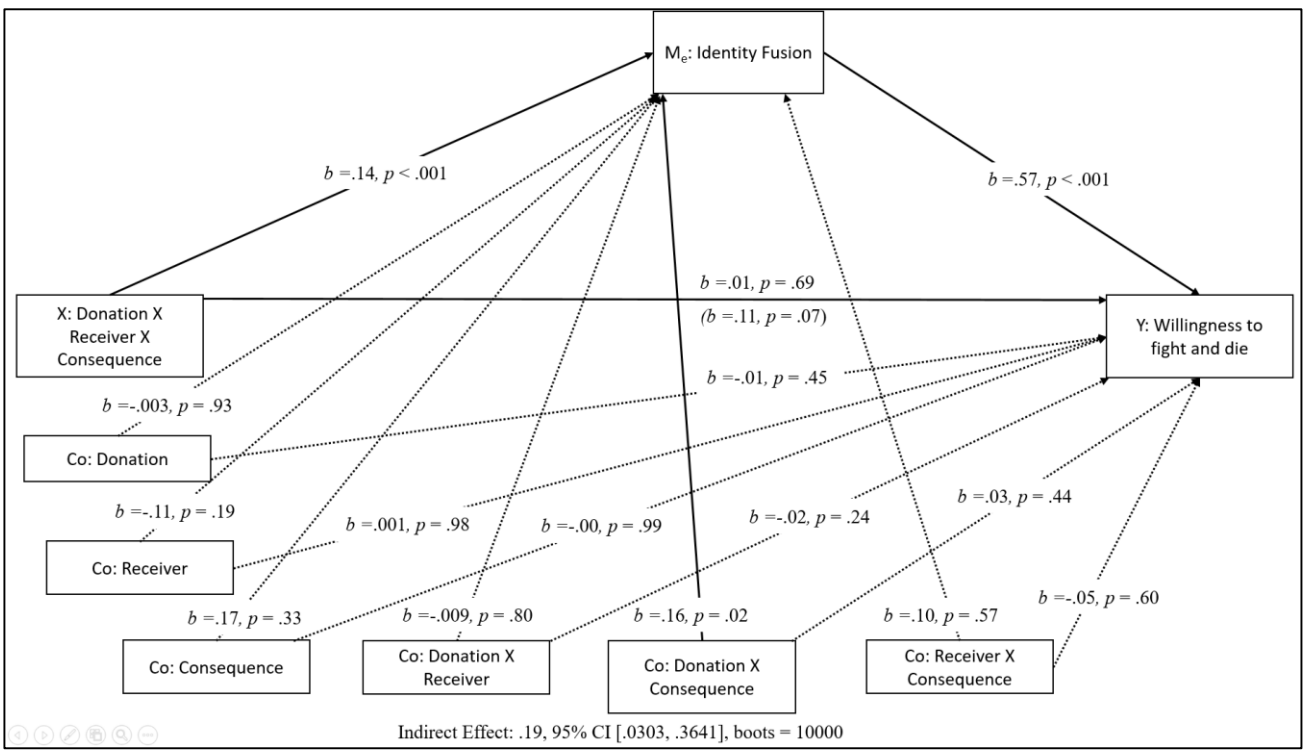
Figure 3B. *Willingness to fight and die scores as a function of Amount of Donation to an out-group member and Donation Consequence (Study 2).*



Identity fusion mediates the interactive effect of Donation and Consequence of Donation on Willingness to fight and die. Our next step was to examine whether identity fusion mediated the interactive effect of Donation, Donation Receiver and Use of Donation on willingness to fight and die. In particular, following Hayes (2017), we conducted a bootstrapping test (n boots = 10,000) using the PROCESS macro for SPSS to test whether the effects of Donation, Donation Receiver and Use of Donation on willingness to fight and die were mediated by identity fusion both on the In-group receiver and the Out-group receiver condition (see Hayes, 2017, Model 12). As seen in Figure 4, this analysis confirmed that the indirect effect of the interaction between Donation, Donation Receiver and Donation Consequence on willingness to fight and die through identity fusion was significant, IE = 0.1930, 95% CI: 0.0326, 0.3670. Additionally, PROCESS provided an index of conditional indirect effects Donation and Donation Consequence on willingness to fight and die in the two experimental conditions in the Donation Receiver independent variable, revealing that the indirect path was significant through identity fusion in the In-group condition, IE = 0.1853, 95% CI: 0.0665, 0.3234, but not in the Out-group

condition, IE = -0.0077, 95% CI: -0.1122, 0.1031. That is, identity fusion impacted willingness to fight and die only when participant´s donations were being delivered to an in-group member.

Figure 4. *Identity fusion with Spain mediates the interactive effects of Donation x Donation Receiver x Donation Consequence on Willingness to fight and die for Spain. CI = Confidence Interval (Study 2).*



Discussion

Study 2 replicated the pattern of effects that emerged in Study 1 and specified under which conditions the effect was more likely to occur. Specifically, a higher amount of pro-group behavior was associated with higher levels of identity fusion among those whose pro-group behavior had Group-beneficial consequences for the group. In contrast, no difference in identity fusion was found among those whose pro-group behavior was non-beneficial for the group. This two-way interaction only emerged when participants' behavior was directed towards an in-group member. No effect was found when participants interacted with a member of the out-group. Once again, engaging in pro-social

behavior directed towards an in-group member affected identity fusion only if such behavior turned out to be used for the benefit of the group. The first replication of the potential role of agency on identity fusion.

Additionally, Study 2 built up on the findings of Study 1 by providing empirical evidence of identity fusion´s predictive power of one of its classic outcome measures (e.g., willingness to fight and die).

Having replicated the basic pattern of effects that emerged in Study 1, the final study had several goals aimed at unanswered questions from the previous studies. First, the primary goal of the final study was to provide additional evidence for the replication of the effects found in Studies 1 & 2.

The second goal was to control for identification measures in our findings. Knowing that there is some statistical overlap between some identification and fusion measures (Swann et al., 2009; Gómez et al., 2011; Jiménez, Gómez, Buhrmester, Vázquez, Whitehouse & Swann, 2015) and identification measures have shown to be sensitive to experimental manipulations (e.g. Ellemers, Spears & Dooske, 1997; Jetten, Branscombe, Schmitt, Spears, 2001; Vugt & Cremer, 1999), one could especulate that the variance found in identity fusion measures in the first two studies was mostly overlapping with social identity. It was then an open question whether the results would hold significant when controlling for identification.

An additional question before running the final study was wether the experimental conditions introduced in studies 1 & 2 were being perceived as Group-beneficial vs Person-beneficial . Therefore, the third goal of this final study was to provide with manipulation checks informing us of whether the intended effect on the intended variable was taking place or not.

The final goal for the third Study is central not only to the current study, but also to the entire manuscript. Once we have shown the robustness of the effect by replicating it, Study 3 aimed at

examining whether personal agency can play a mediating role in the effects found in studies 1 & 2, providing then direct empirical support for the idea of agency being a cause of fusion.

Study 3

There were several goals to be accomplished in Study 3. First. the main goal was to provide empirical evidence for the mediating role of personal agency (Swann et al., 2010) in the changes found in indentity fusion. Second, a measure of identification (Mael & Ashforth, 1992) was included in Study 3 in order to test whether the effect found in previous studies could remain significant when controlling for other conceptually different measure of bonding with the group. Third, Study 3 included a manipulation check in order to test whether the independent variable had the intended effect. Lastly, Study 3 aimed at providing one more replication of the effect found in Studies 1 & 2.

Method

Participants and Design.

Sixty-eight students from Universidad Nacional de Ecucación a Distancia, UNED, participated online in exchange for course credit. Sample size was determined based on the number of participants collected during the week in which the study was posted. Participants were randomly assigned to one of two (Group-beneficial vs. Person-beneficial) between-subjects conditions, with donation to an in-group member as a continuous predictor, identification as a control variable and identity fusion and personal agency as the dependent variables.

Procedure and Materials.

The procedure in this study was identical to that of Study 1, with the exception of including the Identification measure (Mael & Ashforth, 1992), the Personal Agency scale (Swann et al., 2010) and a manipulation check after the use of donation Independent Variable.

Predictor Variable.

Donation. This variable was measured identically to Study 1.

Independent Variable.

Donation Consequence. This variable was manipulated identically to Study 1

Dependent Variables.

Identity fusion. Identity fusion with Spain was measured using the seven-item verbal scale (Gómez, Brooks, et al., 2011). Reliability was $\alpha = .86$.

Identification. Identification with Spain was measured using the Mael & Ashforth (1992) six-item scale. Reliability was $\alpha = .87$. All subsequent analysis in this study controlled for this particular identification scale.

Confidence on the impact of one´s actions on the in-group (e.g., Personal Agency). This form of confidence was measured using five items based on Haggard & Tsakiri´s (2009) discussion on the construct of Agency. This scale has been used on Identity fusion research before (Gómez, Brooks et al., 2011; Swann et al., 2010). Participants answered to these items using Spain as a reference ($\alpha = .89$).

Manipulation check. Participants answered to the following two seven-point Likert scales (0 = Completely Disagree; 6 = Completely Agree): "I consider Player B´s use of my donation very valid for our group" "I consider Player B´s use of my donation very positive for our group". The two items were strongly correlated (r=.87) so they were averaged to form an index.

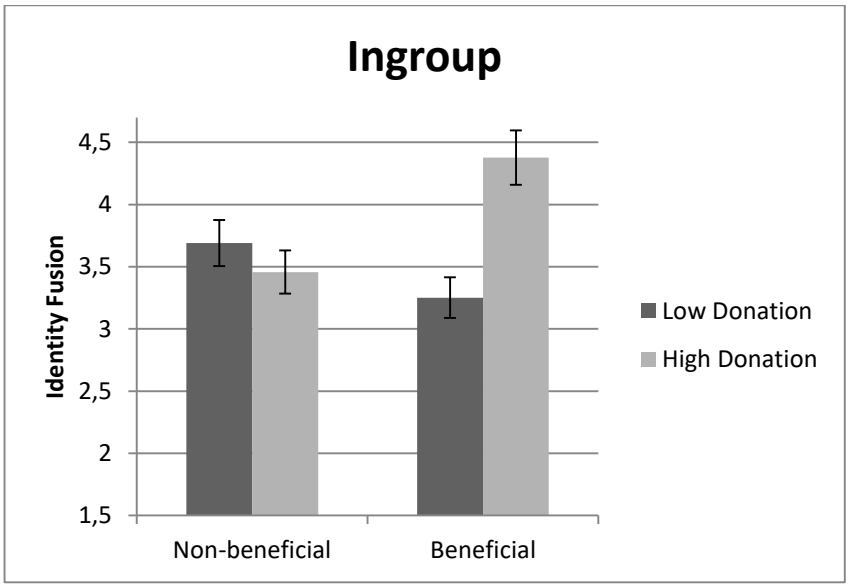
Results

Manipulation check. Multiple regression was used to examine the impact of Donation (centered), Donation Consequence (effect coding: -1, 1), and the interaction of these variables on the Manipulation check. The regression revealed a main effect of the Donation Consequence independent variable, B = 1.91, $t(64)$ = 12.59, $p < 0.001$, 95% CI: 1.6120, 2.2201, meaning that participants in the Group-beneficial condition perceived the consequence of their donation to be more positive and valid to the group than participants in the Person-beneficial condition. No other effects or interactions emerged, $B < .04$, $t(64) < 0.51$, $p > 0.608$.

Donation. The amount of money donated was submitted to a one-way ANOVA with Donation Consequence (Group-beneficial vs Person-beneficial) as the predictor. Donation amount did not change as a function of the Donation Consequence manipulation, $F(1, 69) = 0.391$, $p = 0.534$.

Identity fusion. Multiple regression was used to examine the impact of Donation (centered), Donation Consequence (effect coding: -1, 1), and the interaction of these variables on identity fusion with Spain. The regression revealed a significant two-way interaction between identity Donation and Donation Consequence, $B = 0.10$, $t(64) = 2.52$, $p = 0.013$, 95% CI: 0.0212, 0.1805. As seen in Figure 5, among participants assigned to the Group-beneficial condition, identity fusion scores were higher for participants who donated larger relative to smaller amount of money, $B = 0.16$, $t(64) = 2.36$, $p = 0.021$, 95% CI: 0.0251, 0.2995. In contrast, among participants assigned to the Person-beneficial condition, the amount of money donated was not related to reported levels of identity fusion. $B = -0.04$, $t(64) = -0.97$, $p = 0.333$, 95% CI: -0.1201, 0.0413. Put differently, among participants who donated more than average (+1SD), those assigned to the Group-beneficial condition did not differ in their reported levels of identity fusion compared to those who were assigned to the Person-beneficial condition, $B = .20$, $t(64) = 1.52$, $p = 0.131$, 95% CI: -0.0630, 0.4712. However, among participants who donated less than average (-1SD), those assigned to the Group-beneficial condition reported significantly lower levels of identity fusion with Spain than those assigned to the Person-beneficial condition, $B = -.30$, $t(64) = -2.18$, $p = 0.032$, 95% CI: -0.5831,-0.0261.

Figure 5. *Identity fusion scores as a function of Amount of Donation to an in-group member and Donation Consequence (Study 3).*



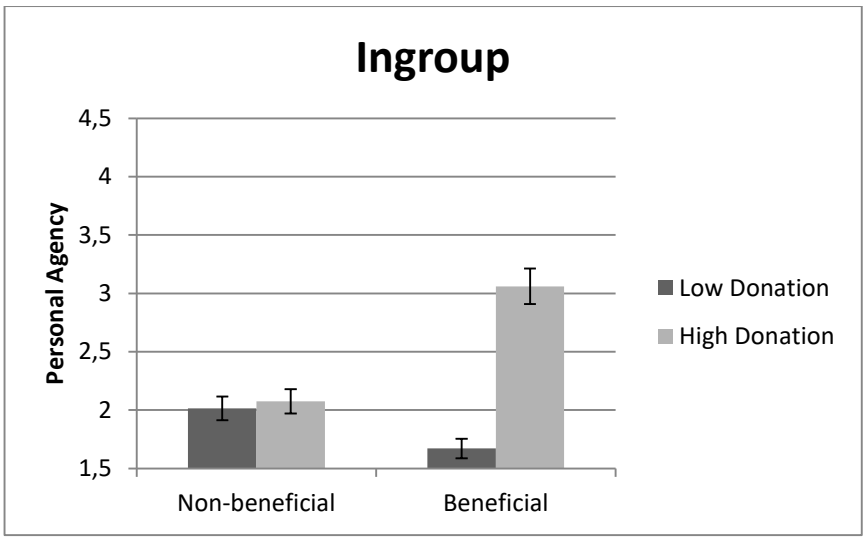
Identification. No effects were significant, $B < 0.05$, $t(64) < 0.612$, $p > 0.404$.

Personal Agency. Multiple regression was used to examine the impact of Donation (centered), Donation Consequence (effect coding: -1, 1), and the interaction of these variables on Personal Agency with Spain. The regression revealed a significant two-way interaction between Donation and Donation Consequence, $B = 0.13$, $t(64) = 2.292$, $p = 0.025$, 95% CI: 0.0075, 0.1978.

As seen in Figure 6, among participants assigned to the Group-beneficial condition, Personal Agency scores were higher for participants who donated larger relative to smaller amount of money, $B = 0.27$, $t(64) = 2.78$, $p = 0.007$, 95% CI: 0.0777, 0.4708. In contrast, among participants assigned to the Person-beneficial condition, the amount of money donated was not related to reported levels of Personal Agency. $B = 0.01$, $t(64) = 0.20$, $p = 0.838$ 95% CI: -0.1047, 0.1286. Put differently, among participants who donated more than average (+1SD), those assigned to the Group-beneficial condition reported significantly higher levels of Personal Agency than those assigned to the Person-beneficial condition, $B = .49$, $t(64) = 2.60$, $p = 0.011$, 95% CI: 0.1154, 0.8700. However, among participants who donated

less than average, no difference in reported levels of identity fusion were reported between those assigned to the Group-beneficial vs. Person-beneficial condition, (-1SD), , B = -.17, $t(64)$ = -0.85, p = 0.394, 95% CI: -0.5727, 0.2288.

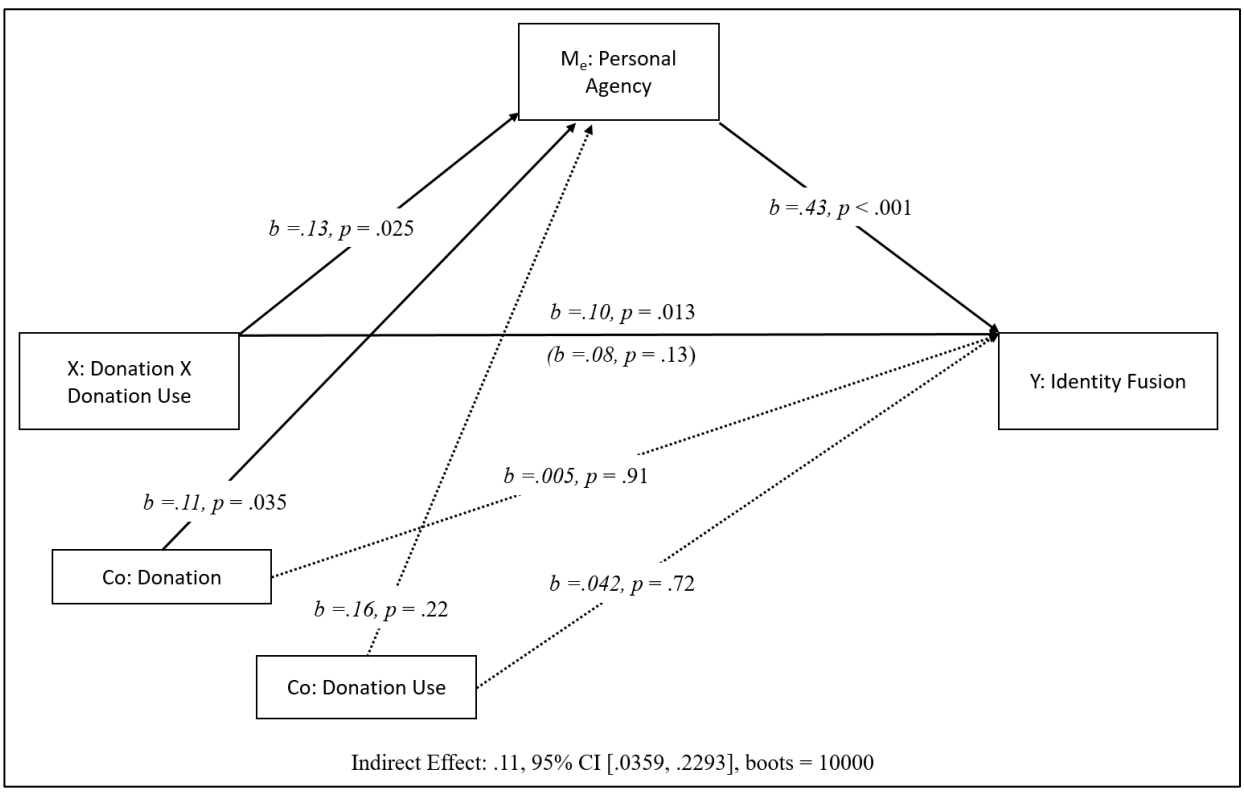
Figure 6. *Feelings of Personal Agency scores as a function of Amount of Donation to an in-group member and Donation Consequence (Study 3).*



Personal Agency mediates the interactive effect of Donation and Consequence of Donation on Identity fusion. Our next step was to examine whether personal agency mediated the interactive effect of Donation and Use of Donation on identity fusion. In particular, following Hayes (2017), we conducted a bootstrapping test (n boots = 10,000) using the PROCESS macro for SPSS to test whether the effects of Donation and Use of Donation on identity fusion were mediated by personal agency both on the Group-beneficial and the Person-beneficial condition (see Hayes, 2017, Model 8). As seen in Figure 7, this analysis confirmed that the indirect effect of the interaction between Donation and Donation Consequence on identity fusion through the feelings of Personal Agency was significant, IE = 0.1130, 95% CI: 0.0384, 0.2292. Additionally, PROCESS provided an index of conditional indirect effects Donation on identity fusion in the two experimental conditions, revealing that the indirect path was significant through personal agency in the Group-beneficial condition, IE = 0.1182, 95% CI:

0.0475, 0.2229, but not in the Person-beneficial condition, IE = 0.0052, 95% CI: -0.0365, 0.0536. That is, feelings of personal agency impacted levels of identity fusion only when participant´s donations were used to the benefit of the group.

Figure 7. *Personal Agency with Spain mediates the interactive effects of Donation x Donation Consequence on identity fusion with Spain. CI = Confidence Interval (Study 3).*



Discussion

Study 3 provides additional replication of the interaction between Donation and Donation consequence on identity fusion with the in-group. Once again a higher amount of pro-group donation was associated with higher levels of identity fusion among those whose pro-group donation had beneficial consequences for the group. In contrast, no difference in identity fusion was found among those whose pro-group donation was beneficial for the person. Additionally, Study 3 controlled for identification, suggesting that this effect might be unique to identity fusion.

Study 3 also provided a manipulation check. Results confirmed that participants perceived Player B´s response in the Group-beneficial condition to be significantly more positive for the group than in the Person-beneficial condition.

More critical to our concerns, Study 3 provides initial support consistent with the notion that making a positive impact in one´s group (e.g., providing a relatively large donation that turns out to be beneficial for the group) increases the sense of personal agency that, in turn, increases levels of fusion with one´s group.

The results of this study show that the attenuation of willingness to fight and die displayed by strongly-fused participants is not related to need for uniqueness, implying that fused identities may already meet the need for optimal distinctiveness (for a more detailed explanation of this notion, see the relational ties principle of identity fusion, Swann et al., 2012). However, our results provide some initial support consistent with the notion that strongly fused participants are confident enough to delay self-sacrifice if necessary, and therefore reduced their willingness to fight and die for the group after learning that other fused individuals would do it for morally relevant reasons. Lastly, the high level of confidence reported by strongly-fused individuals in this study also rules out the possibility that fused participants in the morally relevant condition may have temporarily reduced their levels of fusion with the group as a response to the reasons provided in the message.

General Discussion

The present investigation addressed, for the first time, the question of how the consequences of one´s pro-group behavior can impact feelings of agency and, in turn, levels of identity fusion with one´s in-group. We found that the consequences of one´s pro-group behavior matter. Specifically, a higher amount of pro-group donation was associated with higher levels of identity fusion among those whose pro-group behavior had beneficial consequences for the group, but it did not predict identity fusion when the consequences were non-beneficial for the group. Study 2 replicated findings from Study 1 and

identified conditions under which the effect is more likely to occur (e.g., when interacting with in-group members) as well as corroborated the substantial nature of changes in identity fusion, for they mediated changes in willingness to fight and die for the in-group. Lastly, Study 3 suggests that donation consquences effects on identity fusion are due to changes in the confidence on the impact of one´s actions on the group (e.g., personal agency).

This research has, of course, a few limitations that should be noted. First of all, although this is the first research that provides a full laboratory manipulation of identity fusion, it does so through a quasi-experimental design. Future research should randomly assign participants to a high vs low pro-group behavior manipulation in order to be able to infer causality from our findings with complete confidence. Secondly, although this research offers three instances of the same significant interaction, Study 3 is somewhat underpowered (e.g., 68 participants). Lastly, future research should address alternative explanations to the effects found in this research (e.g., similarity, guilt, etc).

The present work makes several substantial contributions to identity fusion theory. First, although recent research has accomplished succesful manipulations of identity fusion by reminding participants of actual historic threats to the in-group (Vázquez, Gómez & Swann, 2017), this is the first systematic laboratory manipulation of identity fusion that does not draw upon external occurred events to obtain an effect. Study 2 showed that the changes in identity fusion caused by an experimental manipulation were not merely incidental, as identity fusion scores predicted willingness to fight and die for the group. Third, this research offers a new role for personal agency within identity fusion theory, as an antecedent or a cause for changing in feeling one with the group. It may be that the feeling of one´s personal actions having a beneficial impact on one´s group (e.g., personal agency) help bring the social and the personal selves closer together to strengthen the bond individual and collective. Lastly, this research offers a novel, group-related, cognitive method to experimentally induce feelings on personal agency.

The results of the current project make one wonder whether succesful manipualtions of other facets of identity fusion (e.g., identity synergy, irrevocability, relational ties, invulnerability) could also lead to changes in identity fusion through different explanatory mechanisms, with arguably different consequences. For instance, it is plausible to think that heightened levels of fusion manipulated through, say, feelings of invulnerability may be more predictive of group-related consequences conceptually connected to invulnerability (e.g., fight and die) than to any other construct. Conversely, heightened fusion manipulated through relational ties might be better a predictor of group-related consequences conceptually connected to relational ties, such as helping behavior. Future research should look further into these ideas.

CHAPTER 3

Identity fusion Leads To Willingness To Fight and Die for the Group: The Moderating Impact of Being Informed of the Reasons behind Other Members' Sacrifice[1]

[1] This chapter has been accepted as a stand-alone manuscript in *Self and Identity.*

Article ID: PSAI 1419503.

Abstract

Identity fusion is capable of predicting pro-group behaviors. We examine how highly-fused individuals react when they learn the motivations driving other highly-fused individuals' pro-group behaviors. Three studies were conducted in two countries (Spain and USA). Results showed an attenuation effect in their willingness to fight and die for in-group members when highly-fused individuals learned that other participants displaying their same level of visceral commitment to the group would self-sacrifice for morally relevant reasons, compared to pragmatic reasons. The final study suggests that the attenuation in willingness to fight and die observed in those highly fused can be interpreted as a sign of confidence to delay their self-sacrifice rather than a need to be unique.

Introduction

Identity fusion is a visceral feeling of oneness with a group, wherein the personal self (characteristics of individuals that make them unique) joins with the social self (characteristics of individuals that align them with a group), and the borders between the two become porous (Gómez & Vázquez, 2015; Swann, Jetten, Gómez, Whitehouse, & Bastian, 2012). Identity fusion predicts pro-group behaviors such as endorsement of fighting and dying for the group (Gómez, Brooks, Buhrmester, Vázquez, Jetten, & Swann, 2011), willingness to self-sacrifice for in-group members in the trolley dilemma (Swann, Gómez, Buhrmester, López-Rodríguez, Jiménez, & Vázquez, 2014), readiness to deny group wrong-doing (Besta, Gómez, & Vázquez, 2014), and donating personal funds to in-group members (Gómez, Morales Hart, Vázquez, & Swann, 2011).

Research has identified several factors that amplify the consequences of identity fusion and the mechanisms that explain such consequences. In particular, identity fusion predicts the endorsement of pro-group behavior by activating personal or social identity (Gómez, Brooks et al., 2011; Swann et al, 2009), increasing arousal (Swann et al., 2010), and priming shared biological characteristics or core values (Swann, Buhrmester et al., 2014). Moreover, the relationship between strong levels of fusion and an individual's willingness to fight and die for the group has been linked to feelings of agency (Gómez et al., 2011; Swann et al., 2010), perceptions of invulnerability (Gómez, Brooks et al., 2011), familial ties (Swann, Buhrmester et al., 2014), and emotional engagement (Swann, Gómez et al., 2014).

Despite the growing body of research on identity fusion, only one study has examined individuals´ reasoning when deciding to self-sacrifice for in-group members according their degree of fusion with the group. Results indicated that, in a simulated trolley dilemma (Swann et al., 2014), strongly fused participants were more likely to self-sacrifice, and moreover justified their decision using moral reasoning (e.g., "It is the right thing to do"). In contrast, participants scoring low in fusion, who also decided to self-sacrifice for in-group members, justified their decision for pragmatic reasons (e.g.,

"They are more than me, it makes sense that one dies to save many, even if that one is me"), (see also Gawronski, Armstrong, Conway, Friesdorf, & Hütter, 2017).

Given that *self-generated reasons* can vary as a function of identity fusion, the goal of the present research was to examine the extent to which *reasons generated by others* can also play a role in understanding willingness to fight and die.

In thinking about how knowledge of others' reasons to fight and die for the in-group might influence one's own willingness to fight and die, several possibilities emerge: First, being informed about other's morally relevant reasons could either bolster or attenuate the desire of fused individuals to self-sacrifice. A second possibility is that other's reasons may actually have relatively little impact on self-sacrificial behaviors. Indeed, once fused individuals have already decided to self-sacrifice for the group, they may care very little about the reasons behind others' extreme behaviors.

The current experiments were designed to explore these different possibilities. Specifically, our primary goal was to examine whether the perceived reasons behind others' willingness to self-sacrifice for the group matters for one´s own willingness to fight and die. Importantly, others' willingness to engage in self-sacrificial behavior was kept constant and convergent with that of strongly fused individuals (e.g., high willingness to self-sacrifice). Only others' reasons behind their behavior were manipulated across conditions. Assuming that others' reasons play a role, our second goal was to examine *which* reasons (morally relevant or pragmatic) are more likely to affect one's willingness to fight and die, and the extent of their impact. Our final goal was to explore the psychological mechanisms that may be responsible for any potential outcome obtained throughout the studies.

Study 1

Study 1 tested the effects of being informed that other strongly fused individuals are willing to self-sacrifice for in-group members. We examined whether providing information that other people´s self-sacrifice was motivated by morally relevant reasons moderates the relationship between identity

fusion and willingness to fight and die for the group compared to a control condition in which no rationale for the relationship between group bond and self-sacrifice was provided.

Method

Participants and Design. Seventy five North-Americans (72% women, mean age = 24.91, *SD* = 10.72) received a small fee for participating in an online Qualtrics survey through MTurk. This relatively small sample size was all that could be collected, but the data were analyzed nonetheless since the main purpose of the study was to see if a clear direction of the effect emerged. Participants were assigned to one of two experimental conditions (Self-Sacrifice rationale: Morally Relevant vs. Control), with scores on Identity Fusion as a continuous predictor and willingness to fight and die as the dependent variable.

Procedure and Materials. First, participants responded to the 7-item (e.g., "I am one with the United States", "I make the United States strong") *verbal fusion* scale (Gómez, Brooks, et al., 2011), on which responses ranged from 1 (completely disagree) to 7 (completely agree), with higher numbers reflecting more fusion. Reliability was $\alpha = .92$. Next, all participants were provided with a summary of the results of an ostensibly different scientific study explaining that in-group members who hold a strong visceral feeling of oneness with the group are more willing to self-sacrifice to save other in-group members in fictitious situations. The summary reported that American participants were given a version of the trolley dilemma in which it was necessary to decide whether they would (a) let a runaway trolley crush and kill five members of their country or (b) jump off a bridge in front of the trolley to save the five in-group members at the expense of their own lives (adapted from Swann et al., 2010). The text went on to say that the study showed that a representative sample of Americans who have a visceral feeling of oneness with the group showed significantly greater intentions of jumping in front of the trolley and dying to stop it from killing five in-group members.

Participants were then randomly assigned to the *morally relevant reasons to self-sacrifice condition* or to the *control condition.* Participants in the *morally relevant reasons to self-sacrifice condition* learned that the individuals who decided to self-sacrifice for the five in-group members were motivated by morally relevant reasons. Specifically, when confronted with the dilemma, these individuals reported experiencing tension, distress, and anxiety regarding the difficult situation of the in-group members, and instantly believed that the morally correct course of action was to sacrifice oneself. Many of these individuals also noted an immediate impulsive reaction to sacrifice their life to save the lives of their fellow in-group members. The summary also included an example of an individual's line of thinking as follows: "It would be horrible if they should die and I did nothing knowing that sacrificing myself is the right thing to do" (adapted from Swann, Gómez, et al., 2014, p. 719). Participants in the *control condition* received only the summary information but did not receive any feedback about the reasons motivating the self-sacrifice. Finally, all participants completed the willingness to fight and die scale (e.g. I would fight someone physically threatening another member of my country", "I would sacrifice my life if it saved another group member´s life"), taken from Swann et al., (2009). Responses were recorded on a 1 (completely disagree) to 7 (completely agree), scale, with $\alpha = .88$.

We report all measures, manipulations, and exclusions in these three studies. Furthermore, we report all studies conducted in this line of research.

Results

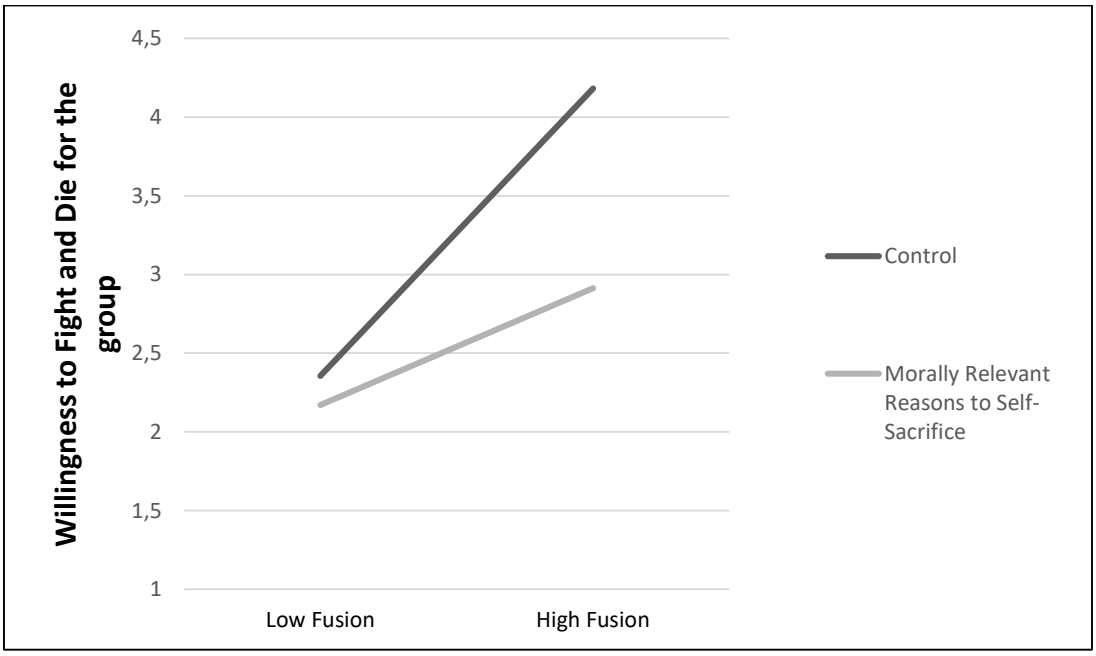
Willingness to fight and die for the group. Linear regression analyses were conducted to examine the effects of identity fusion (centered), self-sacrifice rationale (effect coded: -1 control, 1 morally relevant reasons to self-sacrifice), and the interaction of the two variables on willingness to fight and die for the group. The data revealed a significant main effect of fusion, $B = 0.55$, $t(71) = 6.52$, $p < .001$, 95% CI: 0.3831, 0.7201, indicating a greater willingness to fight and die for strongly fused

relative to weakly fused participants. Also, a significant main effect of self-sacrifice rationale, B = -0.72, $t(71)$ = -3.26, p = .002, 95% CI: -1.1692, -0.2828, indicated that participants in the morally relevant reasons to self-sacrifice condition expressed *less* willingness to fight and die for the group than participants in the control condition.

More critical to our primary concerns, a significant two-way interaction also emerged, B = -0.41, $t(71)$ = -2.21, p = .030, 95% CI: -0.7916, -0.0413. As can be seen in Figure 1, the effect of identity fusion on willingness to fight and die for the group in the morally relevant reasons to self-sacrifice condition was lower than in the control condition, B = 0.28, $t(71)$ = 1.75, p = .083, 95% CI: -0.0381, 0.6083, vs. B = 0.70, $t(71)$ = 7.34, p < .001, 95% CI: 0.5112, 0.8920, respectively.

Put differently, among those participants who scored relatively high in identity fusion (+1SD), those who were assigned to the control condition reported greater willingness to fight and die for their group than those who were assigned to the morally relevant reasons condition, B = -1.26, $t(71)$ = -3.69, p < .001, 95% CI: -1.9529, -0.5830. In contrast, for those who scored relatively low in identity fusion (-1SD) no difference in willingness to fight and die emerged between the control and morally relevant reasons condition, B = -0.18, $t(71)$ = -0.58, p = .563, 95% CI: -0.5638, 0.4487.

Figure 1. *Willingness to Fight and Die for the group as a function of Identity Fusion and Other´s Reasons to Self-sacrifice for in-group members (Study 1).*



Discussion

Study 1 indicated that among strongly fused participants, those receiving morally relevant reasons to sacrifice their lives for members of their in-group were significantly less willing to fight and die for the group than participants who did not receive any justification. In contrast, no difference in willingness to fight and die for the group was found among weakly fused participants as a function of self-sacrifice rationale.

When interpreting these results, it is important to note that among strongly fused individuals, the attenuation in willingness to fight and die for the group in the morally relevant reasons condition could have emerged as a consequence of being informed of the reasoning behind other's behavior. In other words, it is possible that merely the act of providing "reasons" that justify other peoples self-sacrifice, regardless of how morally relevant or irrelevant those reasons might be, could produce an effect. Second, one might argue that this effect could also be driven by cultural characteristics. Therefore, what

has been labelled as "morally relevant" for this sample could have a very different meaning for participants from a different country. To test these possibilities, Study 2 was conducted with a different population and included an experimental condition with a different justification to self-sacrifice for in-group members.

Study 2

Study 2 tested whether knowing the reasons that other strongly fused in-group members are willing to self-sacrifice for their peers would impact one's own willingness to fight and die for the group. Of course, it is possible to provide a variety of different kinds of reasons for one's actions beyond those based on moral principles. For example, previous research has shown that one might engage in self-sacrificial behavior based largely on pramatic reasons (Swann et al., 2014). Interestingly, learning of other in-group members willingness to self-sacrifice for pragmatic reasons may either enhance or attenuate strongly fused individuals' willingness to self-sacrifice for their peers based on factors such as the number of peers their actions might protect from harm. Thus, an important goal of study 2 was to compare the extent to which reasons based on two fundamentally different types of thinking (moral vs. pragmatic reasons) might influence strongly fused in-group members willingness to self-sacrifice for their peers. Similar to study 1, a control group was included as our baseline of comparison in which no reasons to self-sacrifice were provided.

Method

Participants and Design. One hundred and twenty-eight Spanish undergraduate students from UNED (74.2% women, mean age = 38.20, *SD* = 12.14) participated online for course credit. Sample size was determined based on the number of participants collected during the week in which the study was posted. Likewise, our goal was to obtain a minimum of 20 participants per condition, which was achieved. Participants were assigned to one of three experimental conditions (Self-Sacrifice rationale:

Morally Relevant vs. Pragmatic vs. Control), with scores on Identity Fusion as a continuous predictor, and willingness to fight and die as the dependent variable.

Procedure and Materials. First, participants responded to the 7-item (e.g., "I am one with Spain", "I make Spain strong") *verbal fusion* scale (Gómez, Brooks, et al., 2011), on which responses ranged from 1 (completely disagree) to 7 (completely agree), with higher numbers reflecting more fusion. Reliability was $\alpha = .89$. Next, all participants were provided with the same summary as in Study 1; however, in this case the text described a representative sample of Spanish citizens instead of American citizens. Participants were then randomly assigned either to the *control*, to the *morally relevant reasons to self-sacrifice condition,* or to the *pragmatic reasons to self-sacrifice condition.* Participants in the *control* and the *morally relevant reasons to self-sacrifice conditions* received the same experimental manipulation as participants in Study 1, but adapted for a Spanish audience. Participants in the *pragmatic reasons to self-sacrifice* condition learned that the willingness to self-sacrifice shown by in-group members was motivated by pragmatic reasons. Specifically, when confronted with the dilemma, these individuals´ reasoning did not reference any kind of attachment to their fellow group members. Instead, their reasoning focused on minimizing the number of lives lost. The summary also included an example of an individual's line of thinking as follows: "It is better to save five lives than only one, even when the one life is mine." (adapted from Swann et al., 2014, p. 719). Finally, all participants completed the willingness to fight and die scale (Swann et al,. 2009), on which responses ranged from 1 (completely disagree) to 7 (completely agree), with higher numbers reflecting a greater willingness to fight and die. Reliability was $\alpha = .84$.

Results

Willingness to fight and die for the group. We conducted a linear regression analysis using the PROCESS 2.16 macro Model 1 for SPSS with identity fusion (centered), self-sacrifice rationale as a multicategorical predictor (moral reasons vs. pragmatic reasons vs control), and the interaction of the

two variables on participants´ willingness to fight and die. We contrast coded the reasons to self-sacrifice as follows: -1 = Pragmatic condition, 0 = Control condition, 1 = Morally relevant reasons condition. In order to properly probe an interaction that has one multi-categorical predictor, we followed the tutorial by Hayes & Montoya (2017). This required transforming the independent variable into two different dichotomous variables (D_1, D_2). To be able to make all possible comparisons between conditions (Control vs. Morally relevant reasons to self-sacrifice, Control vs. Pragmatic reasons to self-sacrifice, and Morally vs. Pragmatic reasons to self-sacrifice), the analysis was run twice with different coding for D_1 & D_2, namely, dummy (D_1 = 0, 1, 0, D_2 = 0, 0, 1) and sequential (D_1 = 0, 1, 1, D_2 = 0, 0, 1).

The regression revealed a significant main effect of identity fusion on willingness to fight and die for the group, B = 0.65, t(122) = 6.10, p < 0.001, 95% CI: 0.3633, 0.7120, indicating that strongly fused participants were more willing to fight and die for the group. A significant two-way interaction between identity fusion and reasons to self-sacrifice also emerged, $\Delta R2$ = .047, F(2, 122) = 4.846, p = 0.009, meaning that the relationship between identity fusion and willingness to fight and die for the group varied as a function of the different self-sacrifice rationales.

Comparing the control condition and the morally relevant reasons condition, a significant two-way interaction emerged, B = -0.38, t(122) = -2.79, p = .006, 95% CI: -0.6638, -0.1137 , replicating findings from Study 1. As can be seen in Figure 2, the effect of identity fusion on willingness to fight and die for the group in the morally relevant reasons condition was lower than in the control condition, B = 0.21, t(122) = 2.34, p = 0.020, 95% CI: 0.0339, 0.3971, vs. B = 0.60, t(122) = 5.78, p < 0.001, 95% CI: 0.3976, 0.8108, respectively.

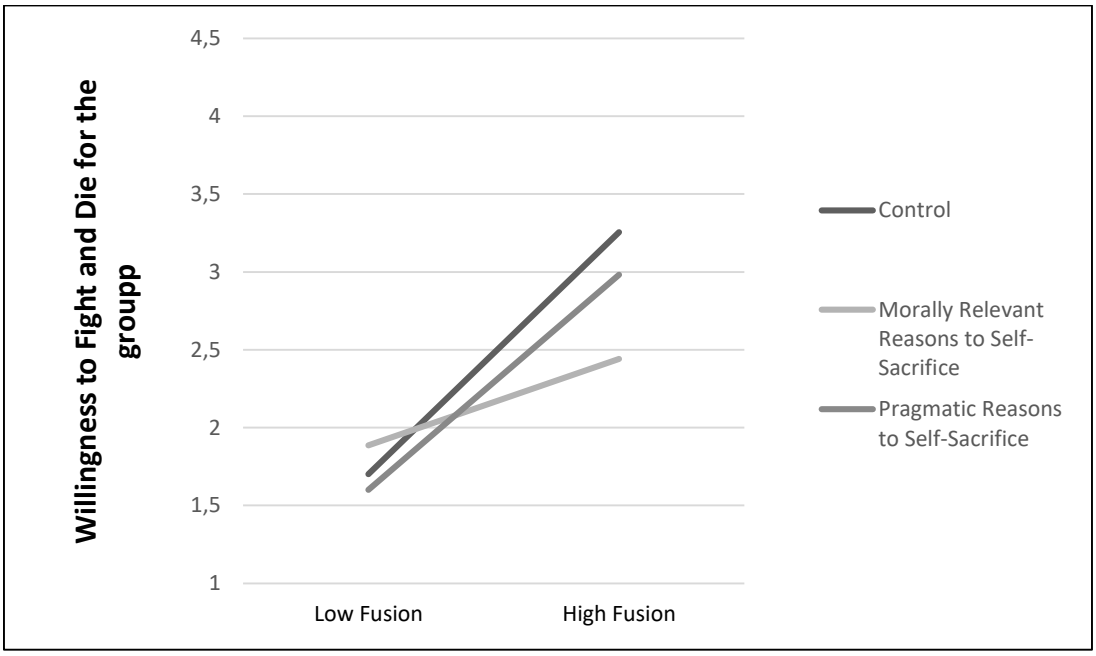
A similar two-way interaction emerged when comparing the pragmatic reasons condition and the morally relevant reasons condition, B = -0.32, t(122) = -2.53, p = 0.012, 95% CI: -0.5739, -0.0705. As can be seen in Figure 2, the effect of identity fusion on willingness to fight and die for the group in

the morally relevant reasons condition was lower than in the pragmatic reasons condition, B = 0.21, t(122) = 2.34, p = 0.020, 95% CI: 0.0339, 0.3971, vs. B = 0.53, t(122) = 6.10, p < 0.001, 95% CI: 0.3633, 0.7120, respectively.

When comparing the control condition with the pragmatic reasons condition, no significant two-way interaction emerged, B = 0.06, t(122) = 0.48, p = 0.626, 95% CI: -0.2037, 0.3369, meaning that the effect of identity fusion on willingness to fight and die for the group in the control condition was not significantly different than the one in the pragmatic reasons condition, B = 0.60, t(122) = 5.78, p < 0.001, 95% CI: 0.3976, 0.8108, vs. B = 0.53, t(122) = 6.10, p < 0.001, 95% CI: 0.3633, 0.7120, respectively.

Put differently, as Figure 2 shows, among participants scoring relatively low in identity fusion (-1SD), no difference in willingness to fight and die for the group emerged between the control condition and the morally relevant reasons condition, B = 0.18, t(122) = 0.75, p = 0.454, 95% CI: -0.3044, 0.6755, the control condition and the pragmatic condition, B = 0.10, t(122) = 0.39, p = 0.693, 95% CI: -0.4049, 0.6072, or the pragmatic condition and the morally relevant reasons condition, B = 0.28, t(122) = 1.25, p = 0.212, 95% CI: -0.1665, 0.7399. However, among participants scoring relatively high in identity fusion (+1SD), the data revealed a significantly greater willingness to fight and die for the group for those participants in the control condition relative to morally relevant reasons condition, B = -0.81, t(122) = -3.21, p = 0.001, 95% CI: -1.3160, -0.3127, yet no difference between the control and the pragmatic condition, B = 0.27, t(122) = 1.12, p = 0.261, 95% CI: -0.2058, 0.7506. Finally, participants in the pragmatic condition were significantly more willing to fight and die than participants in the morally relevant reasons condition, B = -0.54, t(122) = -2.32, p = 0.021, 95% CI: -1.0036, -0.0803.

Figure 2. *Willingness to Fight and Die for the group as a function of Identity Fusion and Other´s Reasons to Self-sacrifice for in-group members (Study 2).*



Discussion

Study 2 replicated the pattern of effects that emerged in Study 1. Specifically, among strongly fused participants, those receiving morally relevant reasons reported significantly less willingness to fight and die for the group compared to those who received no justification. Once again, among weakly fused participants, those assigned to receive morally relevant reasons did not differ from those who received no justification (control condition). Importantly, Study 2 also extended this finding to participants from a different country: Spain.

When we examined the impact of receiving pragmatic reasons on the willingness of participants to fight and die for the group, no difference emerged between those who did not receive any justification (control) and those who received pragmatic reasons and this was true for both strongly and weakly fused participants. However, among strongly fused participants, a significantly greater willingness to fight and die for the group was reported by those who received pragmatic reasons compared with those who

received morally relevant reasons. Among weakly fused participants, those assigned to receive pragmatic reasons did not differ from those who received morally relevant reasons.

Having replicated the basic pattern of effects that emerged in Study 1, we turned our attention to the potential mechanisms that might explain why strongly fused individuals were less inclined to endorse pro-group behaviors after being informed that others were willing to self-sacrifice for morally relevant reasons. We examined two such mechanisms in Study 3.

Study 3

The primary goal of the final study was to explore two reasons why receiving morally relevant reasons might attenuate willingness to fight and die for the group among strongly fused individuals. That is, learning about the moral reasons for the self-sacrifice of others might reduce one's own intention to self-sacrifice for either of two reasons. First, learning about others reasons for self-sacrifice may rob them of the belief that self-sacrifice will make them feel unique. Second, learning this information may bolster their confidence that can sacrifice themselves at a later time.

Study 3 also included an ancillary measure designed to test the extent to which participants perceived to have completely understood the reasons to self-sacrifice manipulation.

Method

Participants and Design. One hundred and fifty-five Spanish undergraduate students from UNED (74.9% women, mean age = 35.21, *SD* = 11.16) participated online for course credit. The procedure was identical to Study 1. Sample size was determined using the same method as in the prior studies. Similar to Studies 1 and 2, our goal was to obtain a minimum of 20 participants per condition, which was achieved. Participants were assigned to one of two experimental conditions (Self-Sacrifice rationale: Morally Relevant vs. Control), with scores on Identity Fusion as a continuous predictor and willingness to fight and die as the dependent variable.

Procedure and Materials. First, participants responded to the 7-item *verbal fusion* scale (Gómez, Brooks, et al., 2011), α = .86. Participants were then randomly assigned to the *morally relevant reasons to self-sacrifice condition* or to the *control condition.* Subsequently, and after being informed that other in-group members would be willing to die for the group, participants were asked to report their perceived *confidence to delay self-sacrifice* using the following 7-point (1 = Strongly Disagree, 7 = Strongly Agree) Likert scale: "I´m confident enough to withhold my sacrifice till next time if there are others who are willing to get involved now". Additionally, participants completed the four-item *Need for Uniqueness scale* from Snyder & Fromkin (1977), (e.g. "I have a high need for uniqueness"), α = .77. Next, participants completed the willingness to fight and die scale, adapted from Swann et al., (2009), α = .81.

Finally, as an ancillary measure, participants responded to the following item: "I thoroughly read and understood the text I was just shown" on a scale ranging from 1 (*totally disagree*) to 7 (*totally agree*). A linear regression analyses including fusion (centered), self-sacrifice rationale(effect coded, -1 control, 1 morally relevant reasons to self-sacrifice), and its interaction on the ancillary measure yielded no significant effects, $B < 0.14$, $t(151) < 0.96$, $p > 0.335$ indicating that neither self-sacrifice rationale nor the level of identity fusion moderated the comprehension of the experimental manipulation. An additional one sample t-test comparing the mid-point of the scale (3.5) with the average score of the comprehension measure indicated that participants reported a significantly higher than mid-point level of understanding of the information, ($M = 5.53$, $SD = 0.72$), $t(144) = 33.64$, $p < 0.001$.

Results

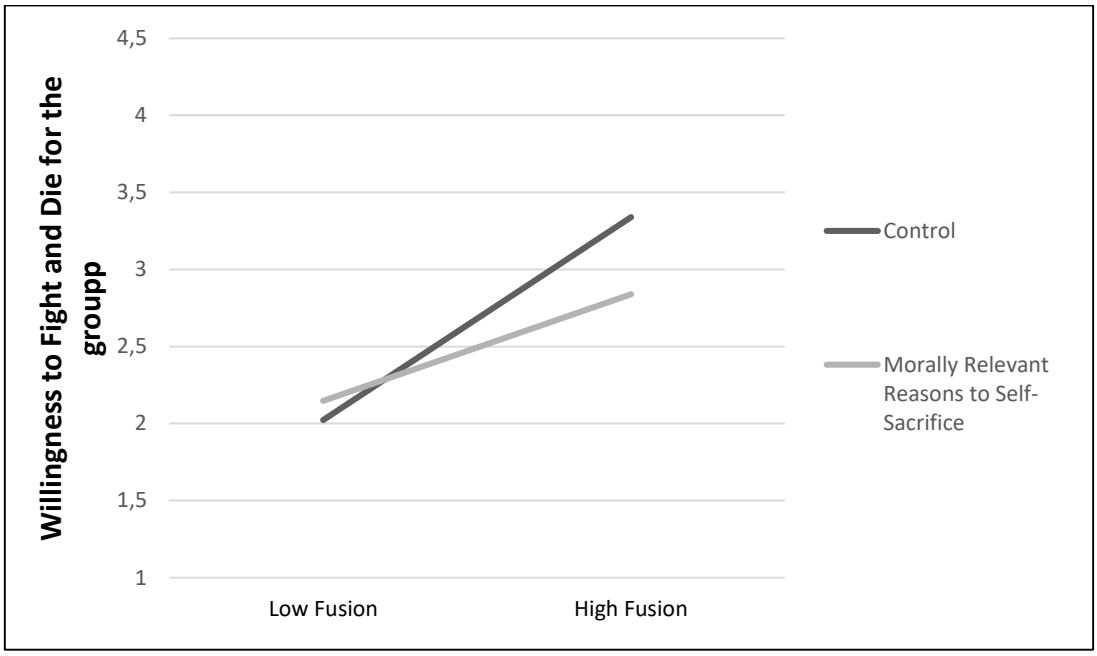
Willingness to fight and die for the group. A linear regression analyses including identity fusion (centered), self-sacrifice rationale (effect coded, -1 control, 1, morally relevant reasons to self-sacrifice), and its interaction, revealed a significant main effect of identity fusion on willingness to fight

and die, B = 0.40, t(151) = 7.53, p < 0.001, 95% CI: 0.2959, 0.5063. No main effect of self-sacrifice rationale emerged, B = -0.18, t(151) = -1.45, p = 0.149.

More critical to our primary concerns, the interaction between identity fusion and self-sacrifice rationale was significant, B = -0.25, t(151) = -2.41, p = 0.016, 95% CI: -0.4663, -0.0471. As can be seen in Figure 3, identity fusion was significantly more predictive of willingness to fight and die for the group in the control condition, B = 0.54, t(151) = 7.19, p < 0.001, 95% CI: 0.3931, 0.6906, than in the morally relevant reasons to self-sacrifice condition, B = 0.28, t(151) = 3.81, p < 0.001, 95% CI: 0.1375, 0.4329.

Once again replicating the pattern that emerged in Studies 1 and 2, among participants who scored relatively high in identity fusion (+1SD), those assigned to the control condition reported greater willingness to fight and die for their group than those assigned to the morally relevant reasons condition, B = -0.49, t(151) = -2.73, p = 0.007, 95% CI: -0.8592, -0.1384. In contrast, among participants who scored relatively low in identity fusion (-1SD) no significant differences emerged in willingness to fight and die between the control condition and the morally relevant reasons condition, B = 0.12, t(151) = 0.68, p = 0.495, 95% CI: -0.22354, 0.4845.

Figure 3. *Willingness to Fight and Die for the group as a function of Identity Fusion and Other´s Reasons to Self-sacrifice for in-group members (Study 3).*



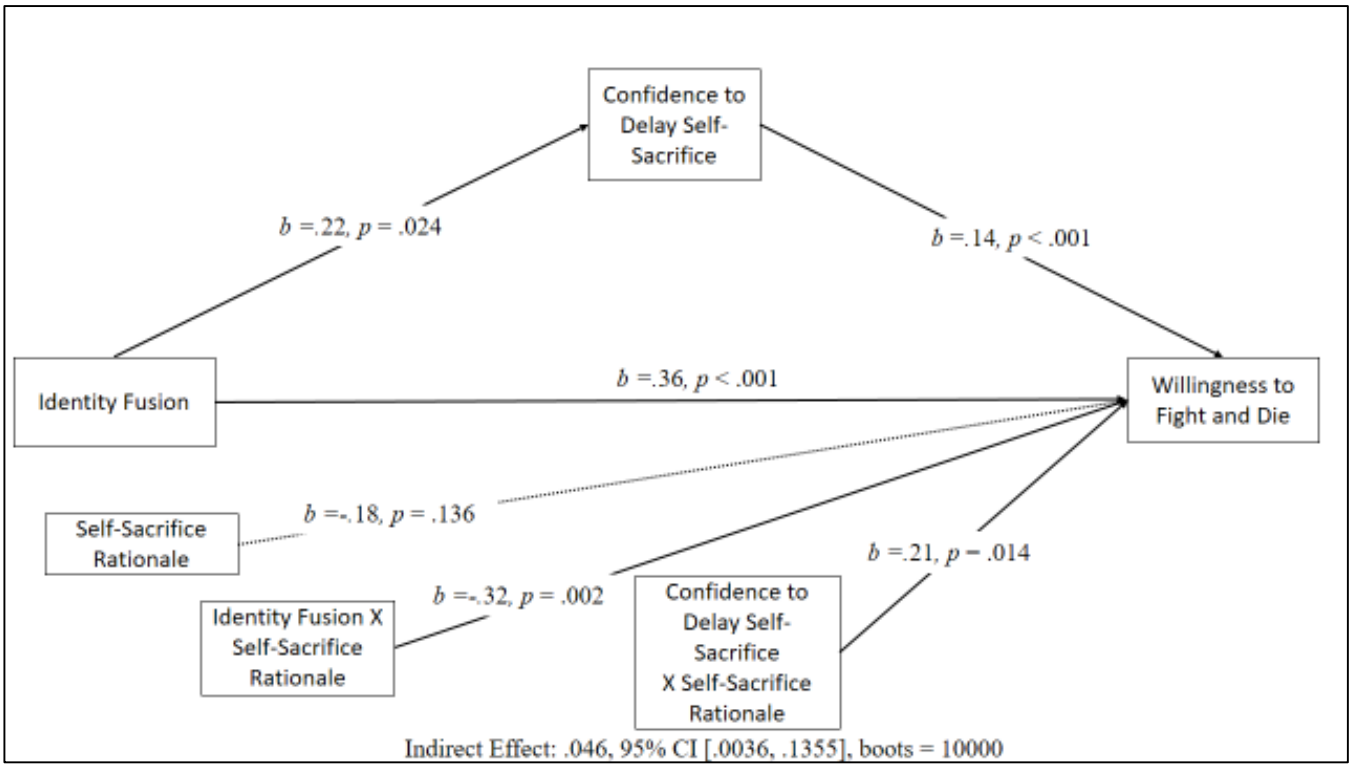
Need for uniqueness. A linear regression analyses including identity fusion (centered), self-sacrifice rationale (effect coded, -1 control, 1, morally relevant reasons to self-sacrifice), and its interaction revealed no significant effects on need for uniqueness, *Bs* < 0.06, *ts*(151) < 0.66, *ps* > 0.505.

Confidence to delay self-sacrifice. A linear regression analyses on the confidence to delay self-sacrifice revealed a significant main effect of identity fusion, *B* = 0.22, *t*(151) = 2.31, *p* = 0.021, 95% CI: 0.0334, 0.4209, indicating that strongly fused participants reported greater confidence that their self-sacrifice could be delayed after being informed that more people were willing to self-sacrifice now. No further effects reached significance, *Bs* < 0.12, *ts*(151) < 0.65, *ps* > 0.513.

Confidence to delay self-sacrifice mediates the interactive effect of identity fusion and others´ reasons for self-sacrifice on willingness to fight and die. Our next step was to examine whether the confidence to delay self-sacrifice mediated the effect of identity fusion on willingness to fight and die for the group, and if this effect was moderated by others´ reasons for self-sacrifice. In particular, following Hayes (2017), we conducted a bootstrapping test (n boots = 10,000) using the

PROCESS macro for SPSS to test whether sacrifice rationale moderated both the direct effect of both identity fusion and confidence to delay self-sacrifice on the dependent variable (i.e., willingness to fight and die), and whether the confidence x reasons two-way interaction on the dependent variable mediated the identity fusion x reasons two-way interaction on the dependent variable (see Hayes, 2017, Model 15). As seen in Figure 4, this analysis confirmed that the indirect effect of identity fusion on willingness to fight and die for the group through the confidence to delay self-sacrifice path was significant, IE = 0.046, 95% CI: 0.0025, 0.1498. Additionally, PROCESS provided an index of conditional indirect effects of identity fusion on willingness to fight and die in the two experimental conditions, revealing that the indirect path was significant through the confidence to delay self-sacrifice in the morally relevant reasons condition, IE = 0.046, 95% CI: 0.0050, 0.1301, but not in the control condition, IE = 0.000, 95% CI: -0.0372, 0.0368. Others´ reasons for self-sacrifice moderated the effect of confidence to delay self-sacrifice on willingness to fight and die as well as the direct effect of identity fusion on willingness to fight and die. That is, strongly fused participants showed greater levels of confidence to delay self-sacrifice. But that confidence only mediates the effects of identity fusion on willingness to fight and die in the morally relevant reasons condition, suggesting that this attenuation is indeed a consequence of strongly fused individuals´ high confidence to delay self-sacrifice if others are willing to step up now.

Figure 4. *Second-stage moderated mediation model with identity fusion and self-sacrifice rationale as predictors, confidence to delay self-sacrifice as the mediator and willingness to fight and die as the dependent variable. (Study 3)* (PROCESS macro, Model 15).



Discussion

Study 3 provides another replication of the interaction between identity fusion and self-sacrifice rationale on willingness to fight and die for the group. Once again, strongly fused participants exhibited lower levels of willingness to fight and die for the group when they were informed that other fused individuals were willing to self-sacrifice for morally relevant reasons compared to a control condition. However, among weakly fused participants, no difference in willingness to fight and die emerged across levels of self-sacrifice rationale. Importantly, Study 3 examined two potential mechanisms for this effect: need for uniqueness and confidence to delay self-sacrifice.

The results of this study show that the attenuation of willingness to fight and die displayed by strongly fused participants is not related to need for uniqueness, implying that fused identities may already meet the need for optimal distinctiveness (for a more detailed explanation of this notion, see the relational ties principle of identity fusion, Swann et al., 2012). However, our results provide some initial support consistent with the notion that strongly fused participants are confident enough to delay self-sacrifice if necessary, and therefore reduced their willingness to fight and die for the group after learning that other fused individuals would do it for morally relevant reasons. Lastly, the high level of confidence reported by strongly fused individuals in this study also rules out the possibility that fused participants in the morally relevant condition may have temporarily reduced their levels of fusion with the group as a response to the reasons provided in the message.

General Discussion

The present investigation addressed for the first time the question of how fused individuals would react to the in-group devotion expressed by others. We tested the role of others´ reasons behind sacrifice because past research suggested that they can play a role in self-justifications of sacrifices (Swann et al., 2014). We found that others reasons matter for self-sacrifice. Being informed that other fused members are also willing to self-sacrifice for the group for moral reasons attenuated their willingness to fight and die for the group. Moreover, Study 3 suggests that this effect may be due to the confidence to delay the sacrifice of those individuals strongly fused with their group rather than to the reduction in feeling unique.

Implications and Future Directions

Our results advance identity fusion theory in at least three ways. These studies represent the first systematic investigation of the effects of identity fusion and the expression of in-group devotion from others and their interactive effects on willingness to fight and die. Second, while previous research has

focused on the moderators that *amplify* the effects of fusion, in this case we have identified a process that *attenuates* its effects. Third, and perhaps most notably, our data suggest that confidence to delay one's self-sacrificial behavior mediates the relationship between identify fusion and pro-group behavior.

Put differently, despite strongly fused individuals´ predisposition towards extreme pro-group sacrifices, previous literature had already explained how thoughtful and elaborated these predispositions can be for strongly fused individuals (Swann et al., 2014; Swann et al., 2010). This research takes this idea one step further by showing that strongly fused individuals are not only willing to give their life to save in-group members, but they also show the necessary patience to delay such extreme action for when it may best serve the group.

Our research is, of course, not exempt of limitations. First of all, although participants expressed having understood our experimental conditions, we have no evidence of what they actually thought. Future research should look into participants actual perceptions (e.g., thought listing, open essay, think aloud). Along these lines, although previous research provides clear evidence in favor of strongly fused persons having moral reasons for their self-sacrifices, no empirical evidence of what reasons strongly fused people hold for self-sacrifice is provided in this research. Given that the two mechanisms we tested rely on the assumption that the moral reasons provided in the summary passage are similar to the ones strongly fused persons hold to engage in self-sacrifice, future research should test to what extent strongly fused people´s reasons for self-sacrifice are indeed similar to the morally relevant reasons provided by others. Lastly, another potential limitation is that our dependent variables do not measure real behavior, but behavioral intentions on hypothetical scenarios. However, past literature has shown that these measures (e.g., willingness to fight and die) do predict actual pro-group behavior (Gómez et al., 2011; Swann et al., 2010).

Importantly, not only do these data contribute to the advancement of identity fusion theory, but their implications can also be extended to other relevant contexts. For example, one area that could

benefit from examining the role of confidence is the domain of delay of gratification (Mischel, Shoda, & Rodriguez, 1989). The research conducted so far in that domain has demonstrated that for most people to be able to delay gratification they need to engage in extensive amounts of thoughtful self-control and deliberative thinking about the consequences associated with their actions in order to overcome their impulses (Metcalfe & Mischel, 1999). However, the higher levels of confidence reported by fused individuals may lead to impulse control through a relatively different route. We speculate that individuals may be willing to wait because they are totally convinced that waiting is the best thing for their group, regardless of the level of self-control resources at the moment. This group-related route to delay of gratification delay is unexplored, thus future research should investigate how levels of fusion could play a role in time-framing effects (Loewenstein & Prelec, 1993). Lastly, the higher levels of confidence reported by fused individuals lead us to speculate that confidence may lead to action in certain scenarios, but to pause (or inaction) in others. Research on attitude-behavior correspondence (Rucker, Tormala, Petty & Briñol, 2014) has shown that the more confident one is in their attitude, the more their attitude is predictive of their behavior. Taken together, the current work offers initial empirical evidence suggesting that, under some circumstances, high confidence might lead to reducing and/or delaying inmediate action in favor of later consistent-behavior (for another recent example of confidence reducing action, see Durso, Briñol, & Petty, 2016).

CHAPTER 4

Matching Person and Situation Increases Reliance on Thoughts:

The Case of Identity fusion

Abstract

In four studies, we examine whether matching Identity Fusion with level of sacrifice influences thought confidence and in turn affects attitude change. These data revealed an interactive rather than additive effect of Identity Fusion and level of sacrifice such that people relied on their thoughts more when they matched their nature than when they did not. The results showed that matching individuals' dispositions (level of identity fusion) to the behaviors they are aligned with (level of self-sacrifice for in-group members) produced greater thought use compared to mismatching these variables. Specifically, people scoring high in identity fusion relied more on thoughts about relatively high sacrifices and people scoring low in identity fusion relied more on thoughts about relatively low sacrifices. In addition to identifying a new role for identify fusion, the present research provides evidence for the underlying process by which this matching effect occurred. Matching affected attitude change by influencing the confidence people had in the thoughts they generated. The present research has implications for Identity fusion, thought validation, matching, attitude change and beyond.

Introduction

Identity fusion is a remarkably strong form of group alignment in which the boundary between personal and social identity becomes porous, producing a visceral feeling of oneness with the group (Gómez & Vázquez, 2015; Swann, Gómez, Seyle, Morales & Huici, 2009; Swann, Jetten, Gómez, Whitehouse, & Bastian, 2012). The result is a powerful feeling of connectedness to the group category without diminishing the integrity of either the personal or social self. In turn, this allows fused individuals to experience a high sense of personal agency (i.e. the capacity to initiate and control intentional behavior) as well as derive reciprocal strength from group membership. Thus, strongly fused individuals are likely to develop strong relational ties with the group (involving attachment to fellow group members) because members are valued by virtue of both their group membership as well as their distinctive personal qualities. Moreover, the connections with others and with the group tend to be maintained over time, and thus contextual factors are less likely to affect relational ties and endorsement of pro-group behaviors (Vázquez, Gómez, & Swann, 2017).

This paper seeks to offer a new role for identity fusion. Namely, that identity fusion may moderate the extent to which thoughts predict attitudes within the domain of in-group self-sacrifice. Furthermore, this research builds on the existing identity fusion literature by examining, for the first time, the relationship between identity fusion and self-persuasion.

This research provides several novelties to the existing literature on identity fusion. First of all, this is the first research in which identity fusion is approached from a metacognitive perspective. Examining how certain variables may affect how people use their thoughts in forming their attitudes has successfully shown reversal of classic effects (Briñol & Petty, 2003; Briñol, Petty, & Tormala, 2006; Briñol, Petty & Barden, 2007; Paredes, Stavraki, Briñol & Petty, 2013; Briñol, Petty & Requero, 2017, among others), meaning that this research could shed light on, for instance, what circumstances may lead individuals scoring low in fusion to show more favorable attitudes towards sacrifice than

individuals scoring high in fusion. Our approach uses a paradigm that matches personal attributes (level of identity fusion) with situational variables (level of sacrifice) to assess the extent to which matching increases confidence on thoughts and the resulting effect on persuasion.

Identity fusion Moderates the Correspondence between mental constructs

A growing body of experimental evidence has demonstrated that identity fusion can predict an individual's willingness to fight and die for their group (Gómez, Brooks, et al., 2011; Gómez, Morales, Hart, Vazquez, & Swann, 2011; Swann et al., 2009, 2014a; Swann, Gómez, Huici, Morales, & Hixon, 2010), willingness to self-sacrifice for in-group members on different intergroup and intragroup versions of the trolley dilemma (Gómez, Brooks, et al., 2011; Swann et al., 2014a; Swann, Gómez, Dovidio, Hart, & Jetten, 2010), serving as front-line combatants during the 2011 Libyan revolution (Whitehouse, McQuinn, Buhrmester & Swann, 2014), and a willingness to die for the beliefs associated with the group in order to implement (or to comply with) strict sharia for Moroccans or for democracy in Westernized countries/nations (Sheik, Gómez & Atran, 2016).

Identity fusion has also been shown to function as a mediating variable between specific predictors and certain pro-group behaviors. For instance, a recent study by Whitehouse et al., (2017) using the classic trolley dilemma paradigm revealed that fans of losing football teams in the English Premier League were more likely to moralize group-related actions and choose to sacrifice themselves for the sake of an in-group member than fans of winning football teams. These data further revealed that the effects of team support on self-sacrificial responses and pro-group moral endorsements were both mediated by identity fusion.

Lastly, identity fusion has been shown to moderate the relationship between certain situational variables and the willingness to engage in extreme pro-group behavior. For example, Swann, Gómez, Huici et al, (2010) conducted four studies in which physiological arousal was experimentally induced

by having participants either operate a stationary bicycle, run wind sprints, or play dodge ball. The results indicated that increases in arousal amplified the tendency for individuals scoring high vs low in fusion to translate their feelings of fusion into pro-group behavior, including overt behaviors such as motor activity (i.e., racing a group-related avatar) or donating personal funds to a needy group member. Moreover, perceptions of agency mediated the interactive effects of arousal and fusion on pro-group behavior (e.g. see Gómez, Morales, et al., (2011) for an example of Fusion moderating the reaction to ostracism).

In conclusion, prior work on identity fusion has focused on its predictive power over many forms of primary cognitions and their links to behavior. That is, thoughts or beliefs that occur at a direct level of cognition and involve a person's initial associations of an object with some attribute. As just reviewed, this research has clearly shown that high scores in identity fusion can increase willingness to engage in extreme pro-group behavior. However, no previous work has examined the impact of identity fusion on secondary, meta-cognitive cognition.

Secondary Cognition

In addition to primary cognitions, secondary cognitions are also important determinants of judgment such as whether people rely on their primary generated thoughts or not. Secondary thoughts occur at a metacognitive level and involve reflections on the first-level thoughts (e.g., "I am not sure how mean this company is," or "I am certain that I am a responsible person"). As a result, two people may have the same thought but one relies on that though more than the other person. The individual who relies more on his thought will use it to form their judgments. Therefore, metacognition has an important role in social judgment because secondary thoughts can magnify, attenuate, or even reverse the impact of first-level cognitions (Petty, Briñol, Tormala, & Wegener, 2007). Moreover, metacognition thoughts can produce changes in thought, feeling, and behavior and thus are critical for

a complete understanding of human behavior (Bless & Forgas, 2000; Briñol & DeMarree, 2012; Metcalfe & Finn, 2008; Wells, Olson, & Charman, 2002).The specific proposition relevant to the current research is that one's thoughts alone are not sufficient for predicting subsequent attitudes. Rather, people must also rely on their thoughts for thoughts to have a marked influence on resulting evaluations. And most importantly, we predict that differences in identity fusion can affect thought use. To our knowledge, no previous research has focused on this metacognitive outcome of identity fusion.

Importantly, one common feature among studies on identity fusion is that they have examined relevant issues and behaviors for individuals scoring high in fusion, namely, extreme sacrifices. These studies almost unanimously show that high levels of fusion are associated with extreme pro-group behavior. Yet interestingly, none of these studies explore whether those situations or actions (e.g., extreme pro-group behavior) may lead strongly-fused people to rely more on what they think, or to feel more confident about their thoughts regardless of their valence.

To address this question that is not responded in the literature, in the current research we examine whether identity fusion can affect secondary cognition. That is, we examine the effects of levels of identity fusion on the extent to which people rely on their ongoing thoughts about various topics. As described in more detail shortly, we use a *self-validation* paradigm in which the effect of a variable (in this case, scores on identity fusion) is examined as a factor that increases or decreases the use of one's own thoughts in forming judgments (Petty, Briñol, & Tormala, 2002). For example, if positive thoughts are generated, as thought confidence is increased, evaluations should become more positive, but if thoughts are negative, as thought confidence is increased, evaluations should become more negative. Stated differently, thought confidence interacts with direction of thoughts to influence attitudes (see Briñol & Petty, 2009, for a review). As an example of research on self-validation, Briñol, Petty, and Wagner (2009) asked participants to think about and write down their best or worse qualities while

sitting with their backs erect, pushing their chests out (i.e., confident posture potentially associated with high power) or while sitting slouched forward with their backs curved (i.e., doubtful posture potentially associated with low power; Carney, Cuddy, & Yap, 2010). Then, participants completed a number of measures, including self-esteem. In line with the self-validation hypothesis, it was predicted and found that the thoughts generated about the self only affected self-attitudes in the confident, more powerful posture. Thus, the effect of the direction of thoughts on self-esteem was greater when participants wrote their thoughts in the confident rather than the doubtful body posture.

Identity fusion and Secondary Cognition

Going back to identity fusion, if being scoring high (vs low) in identity fusion had a positive effect on thought confidence of previously generated thoughts, it would increase the impact of thought direction on attitudes. This might occur, for example, if being strongly-fused made people feel good and they came to like their accessible thoughts (Briñol, Petty, & Barden, 2007). Alternatively, scoring high (vs low) in identity fusion might have a negative effect on thought confidence. That would be the case if, for example, being strongly-fused was a compensatory state that stem from people feeling particularly bad, weak or vulnerable (Briñol, et al., 2007). In that case, scoring high (vs low) in identity fusion would decrease the impact of thought direction, reducing persuasion for positive thoughts and enhancing persuasion for negative thoughts.

A third possibility is that each of the above effects is true – being strongly fused can increase or decrease thought use -- but would do so for different situations. This person-by-situation approach suggests that the effect of being strongly fused on thought confidence depends on situational characteristics. In the current research, we consider the role of extremity of self-sacrifice. Given strongly-fused individuals´ documented readiness and willingness to engage in extreme forms of pro-group behavior, it is reasonable to speculate that strongly-fused individuals may feel *more* confident in thoughts when they are assigned to engage in extreme types of pro-group self-sacrifice, increasing

therefore the reliance in any primary thoughts they might have accesible at the time (and regardless of whether thoughts about in favor or against self-sacrifice).

Before returning to the person × situation approach, how might extreme pro-group self-sacrifice on its own affect thought confidence? One possibility is that those who are willing to harm themselves for their group (i.e., engage in extreme self-sacrifice) would show less thought confidence because of a low general self-evaluation and might therefore have negative reactions to their own thoughts. If so, then people willing to engage in high or extreme (*vs* low or non-extreme) self-sacrifices would show a reduced impact of thought direction on evaluations. However, as just explained for identity fusion, it is also possible that those willingn to engage in high or extreme (*vs* low or non-extreme) self-sacrifices might feel more confident in the validity of their thoughts to the extent that being ready to commit such extreme actions might need a certain degree of confidence, readiness, and approach tendencies (Briñol, Petty, & DeMarree, 2015). If so, then those willing to engage in high or extreme (*vs* low or non-extreme) sacrifices would be more likely to use their thoughts when making evaluative judgments.

As noted, rather than relying on main effects for the situation of being high (vs low) in identity fusion and having willingness to engage in high (vs low) self-sacrifices, we proposed that these variables will interact to affect thought use. Specifically, we hypothesized that a match between an individual's personality (i.e., being either high or low in identity fusion) and the situation (i.e., engaging in high or extreme vs low or non-extreme self-sacrifices) would enhance thought reliance over a mismatch. That is, people high in identity fusion would be more likely to rely on their thoughts when engaging in extreme self-sacrifices, but people low in identity fusion would be less likely to rely on their thoughts engaging in extreme self-sacrifices.

In order to identify these hypothesized differences in thought confidence, the present set of experiments seeks to examine whether identity fusion may moderate responses to concepts such as high or extreme *vs* low or non-extreme sacrifices. More specifically, these experiments will investigate

whether strongly-fused individuals respond similarly to relatively higher or extreme forms of sacrifice than to lower or non-extreme forms of sacrifice. A further goal is to test whether certain forms of pro-group behavior may be better suited for individuals low in identity fusion. This matching approach to identity fusion and sacrifice may result in the identification of boundary conditions under which relatively high levels of identity fusion may be less predictive of pro-group behavior than relatively low levels of identity fusion, as well as the reverse. Specifically, we examine the extent to which thoughts predict attitudes in the domain of group behavior under (mis)matching conditions. Our prediction is that participants' thoughts will have a greater influence on their attitudes when there is a match between the level of fusion with the group and the level of the sacrifice for the group. Specifically, we expect individuals scoring high in fusion to have greater thought confidence when considering high vs. low levels of sacrifice, and the reverse for individuals scoring low in fusion.

Matching Person (Fusion) and Situation (Sacrifice) Affecting Thought Usage

The available literature suggests a strong association between fusion and extreme pro-group behaviors. As reviewed above, individuals scoring high (vs. low) in fusion are more likely to express willingness to fight and die for their groups, to engage in strongly dangerous and painful behaviors, and to show a predisposition to other high self-sacrifices. One way of conceptualizing this relationship is by evaluating the extent to which the person (high level of fusion) and situation (high level of sacrifice) are congruent; in other words, the degree of (mis)match between these variables. In the present research we argue that, compared to individuals with relatively low levels of fusion, individuals with relatively high levels of fusion are more likely to match with thoughts about relatively high levels of pro-group behavior. Similarly, compared to individuals with relatively high levels of fusion, individuals with relatively low levels of fusion are more likely to match with thoughts about lower levels of pro-group behavior.

Attitudes researchers have shown that matching the context (e.g., a particular message) to different aspects of an individual's personal characteristics (e.g., personality) can increase its persuasive effect. Matching procedures with the goal of increasing persuasion can be used to match a large variety of individual differences in personality, needs, interests, and concerns of a recipient or a group to which the recipient belongs (for reviews, see, e.g., Salovey & Wegener, 2003; Teeny, Briñol, & Petty, 2016). Indeed, there are many ways in which a treatment can be matched or tailored to some aspect of the person to make it more persuasive, including the use of a personalized message (tailored at the individual level such as including a person's name) and targeted messages (tailored at the group level such as directing the message to one's race or gender), for a review on different types of matching see, Briñol, & Petty, 2006; Fleming & Petty, 2000; Petty, Wheeler, & Bizer, 1999; Maio & Haddock, 2007).Although there are a number of mechanisms by which matching can influence attitudes (see, Briñol, & Petty, 2006; Petty, Wheeler, & Bizer, 2000), of most relevance here is the notion that matching people and situations can impact attitudes by influencing thought-confidence. This meta-cognitive process of thought validation is a key component of the Elaboration Likelihood Model (ELM; Petty & Cacioppo, 1986; Petty & Briñol, 2012).

Briefly described, when the thoughts in response to a message are primarily favorable, increasing confidence in their validity should increase persuasion, but increasing doubt about their validity should decrease persuasion. When thoughts are primarily unfavorable, however, increasing confidence in their validity should decrease persuasion, but increasing doubt about their validity should increase persuasion. Thus, the metacognitive factor of confidence should interact with thought valence in determining persuasion.

This research is the first to investigate the effects of matching individual differences in identify fusion and situational differences in extremity on the extent to which thoughts are used to form attitudes.

Interestingly, the matching hypothesis applied to level of fusion and level of sacrifice predicts that a treatment against pro-group behavior (i.e. having thoughts against a high level of sacrifice) plus a low level of identity fusion could interact in such a way as to produce more *positive* attitudes than the mismatched conditions of people high in identity fusion generating positive thoughts about a low level of sacrifice.

The present research examines the impact of matching on fusion. Over the past 10 years, matching has been shown to validate attitude-relevant thoughts across many contexts, including stereotypes (Clark, Wegener, Briñol, & Petty, 2009), affective experiences (Huntsinger, 2013), and persuasion (Clark, Wegener, Sawicki, Petty, & Briñol, 2013; Evans & Clark, 2012; Karmarkar, & Tormala, 2010). Research has shown that matching leads to more thought use for variety of reasons, including feelings of ease or fluency (Lee & Aaker, 2004), feelings of liking (Chen, Langner, & Mendoza-Denton, 2009), regulatory fit (Cesario, Grant, & Higgins, 2004), and self-relevance (Petty et al., 2000) that accompany matching.

As an illustration of how a match between the person and the situation can validate thoughts, consider research on self-monitoring. High self-monitors are oriented toward social approval whereas low self-monitors are more motivated to be consistent with their internal beliefs and values (Snyder, 1974; Paredes, et al., 2015). Recent research has shown that individual differences in self-monitoring are relevant to meta-cognitive processes of validation. For example, Evans and Clark (2012) matched individuals who varied in their scores on the self-monitoring scale with a communication source that was high in either expertise or attractiveness. Based on prior research, these researchers argued that people high in self-monitoring are particularly interested in image-related information, thus should be drawn to attractive sources, whereas people low in self-monitoring are particularly influenced by quality or merit information thus are particularly drawn to expert sources (e.g., DeBono & Harnish, 1988). In line with these predictions, the data revealed that high self-monitors relied on their thoughts more when

the source was attractive rather than expert, but low self-monitors relied on their thoughts more when the source was expert rather than attractive. Thus, matching the source to the recipient increased thought use. As a consequence, when thoughts were positive matching increased persuasion but when thoughts were negative matching decreased persuasion. In sum, Evans and Clark (2012) showed that people demonstrated increased reliance on their thoughts when the characteristics of the message source were compatible with (i.e., matched) the characteristics of the message recipient.

Applying a similar matching logic to emotions, Huntsinger (2013b) has shown that a match (vs. mismatch) between emotions and activated evaluative concepts can influence the confidence with which people hold their thoughts. In this research, participants first read a message containing strong or weak arguments in favor of senior comprehensive exams and then listed their thoughts. Next, participants received the emotional congruence (matching) manipulation, which required them to listen to either happy or sad music followed by a lexical decision task in which they were subliminally primed with either happy (e.g., smile) or sad (e.g., glum) words. Affective coherence (matching) occurred when happy or sad music was paired with similarly valenced prime words and incoherence (mismatching) occurred when happy or sad music was paired with oppositely valenced prime words. Consistent with the notion that affective coherence (matching) could validate thoughts, participants in the affective coherence conditions showed a larger effect of argument quality on attitudes than those in the incoherence condition. Furthermore, participants were asked to rate the extent to which they were confident in their thoughts about the message (Briñol et al., 2007). Consistent with the self-validation logic, affective coherence led people to have more confidence in their thoughts and this thought confidence mediated the impact of emotional coherence on attitudes.

As noted, research has identified a number of psychological reasons to help understand why people trust and like their thoughts more when there is a match rather than a mismatch, including because they feel right (Cesario, Grant & Higgins, 2004) or are easier to process (e.g., Lee & Aaker,

2004) and generate (Tormala, Petty & Briñol, 2002). In addition to these reasons, matching can lead to more thought validation through other processes. For example, a match between trait aggressiveness and violence can lead people to feelings of anger, and anger is a negative emotion associated with confidence (Tiedens & Linton, 2001; Blankenship, Nesbit & Murray, 2013). Furthermore, high aggressive people playing violent videogames can feel especially powerful, and power can also lead to greater confidence and use of thoughts (Santos, Briñol, Cárdaba & Petty, 2017).

Overview of the present research

In applying the person-by-situation (matching) approach to the relationship between identity fusion and pro-group behavior, our assumption was that a match would occur when fused individuals generated thoughts about a relatively strong (vs. weak) level of sacrifice, or a weakly fused generated thoughts about a relatively low level of sacrifice. Our prediction was that thought use would be greater in the matching vs. mismatching conditions. Thus, if people had generated positive thoughts, matching would lead to more favorable attitudes than mismatching, but if people had generated negative thoughts, matching would lead to more negative attitudes than mismatching.

We tested our matching hypothesis in four studies, each of which used different levels of sacrifice or manipulated it.

Study 1

Study 1 was designed to test whether a match between levels of identity fusion with a group and self-sacrifice for in-group members can influence reliance on thoughts about attitudes toward self-sacrifice. The match between fusion and thinking about a high self-sacrifice was introduced by asking participants to think about sacrificing their lives for an in-group member. Previous identity fusion research has indicated that high levels of fusion are associated with high levels of extreme pro-group behaviors, such as willingness to die for the group (Gómez et al., 2011; Swann et al., 2009, 2010, 2014), and self-sacrifice in the trolley dilemma (Gómez et al., 2011; Swann et al., 2010, 2014).

We predicted an interaction between thought direction and identity fusion. Participants in the matching condition (e.g., high level of self-sacrifice – high level of identity fusion) were expected to rely more on their thoughts in forming subsequent attitudes toward self-sacrifice for in-group members than those in the mismatching condition (e.g., high level of self-sacrifice – low level of identity fusion).

Method

Participants and design. Eighty-eight undergraduate students at the Universidad Nacional de Educación a Distancia, UNED, participated online typically from their homes in exchange for partial fulfillment of a course requirement (60% women, *Mage* = 38.58 years, *SD* = 11.57). Sample size was determined based on the number of participants collected during the week in which the study was posted. Participants were randomly assigned to a between-participants, 2 (Thought Direction: In favor of self-sacrifice vs. Against self-sacrifice) x identity fusion measure design.

Procedure and Materials.

First, participants were instructed to generate and write down their positive and negative thoughts in favor of vs. against sacrificing their lives for an in-group member. Asking participants to write down their thoughts is a reliable strategy to bias the direction of the participants' thoughts and subsequent attitudes toward a particular topic (e.g., Briñol & Petty, 2003; Briñol, Petty, & Wagner, 2009; Hermann, Leonardelli, & Arkin, 2002; Killeya & Johnson, 1998). That is, people who list thoughts in favor of something typically form more positive evaluations about that topic than those who list thoughts against it.

Next, participants completed a verbal scale of identity fusion with their groupand their reported their attitude towards self-sacrifice for in-group members. Finally, participants were debriefed and thanked.

Independent variables.

Thought Direction. Participants assigned to the *in favor of self-sacrifice* condition were asked to generate three reasons why they would sacrifice their lives for a Spaniard who resides in Spain. Participants in the *against self-sacrifice* condition were asked to generate three reasons against sacrificing their lives for a Spaniard who resides in Spain. Participants were told that this was an important task, thus they should think carefully about their reasons. Some examples of reasons in favor of self-sacrifice were: "I could do it if by sacrifing myself I saved many Spaniards", "I could do it in case our country was being invaded by a foreign nation." Some examples of reasons against self-sacrifice were: "I would not do it because I don´t know the person", "Because I would not want to leave my son alone in this world".

Identity fusion. To measure identity fusion to the in-group (My Country), we used the seven-item verbal fusion scale (Gómez, Brooks, et al., 2011). Responses were provided on scales ranging from 0 (*strongly disagree*) to 6 (*strongly agree*), $\alpha = .89$. Items were averaged into a composite index. Higher scores reflected stronger fusion with the country. Examples of items include: "I am one with my country," and "I am strong because of my country." The identity fusion scale was submitted to a one-way ANOVA using thought direction as the predictor. As expected, identity fusion did not change as a function of the thought direction manipulation, $p = .87$, indicating that listing reasons in favor or against self-sacrifice did not alter levels of identity fusion.

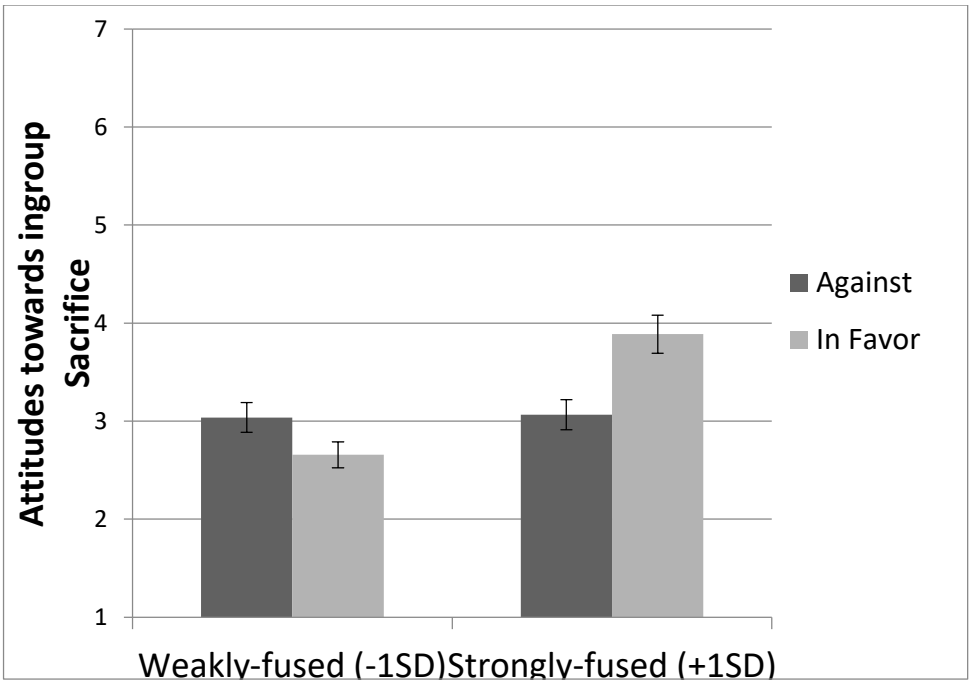
Dependent variables.

Attitudes towards self-sacrifice for in-group members. Participants responded to six items asking them to rate to what extent sacrificing their lifes for other members of their country is "good", "the right thing", "brave", "unintelligent", "absurd", and "reckless". Scales ranged from 1 (strongly disagree) to 7 (strongly agree), $\alpha = .82$. Overall scores were computed by reverse coding the three negative items, then averaging the scores across all scale items. Higher scores represent more favorable attitudes towards self-sacrifice for in-group members.

Results and Discussion

Attitudes toward self-sacrifice for in-group members. A linear regression analysis including thought direction (effect coding: -1 = against self-sacrifice, 1 = in favor of self-sacrifice), identity fusion (centered) and their interaction on attitudes toward self-sacrifice for in-group members revealed a significant two-way interaction, $B = 0.22$, $t(84) = 2.43$, $p = 0.017$, 95% CI: 0.0394, 0.3952. As Figure 1 shows, strongly-fused participants displayed more favorable attitudes toward self-sacrifice for in-group members in the thoughts-in-favor of self-sacrifice, relative to thoughts against self-sacrifice condition, $B = 0.41$, $t(84) = 2.39$, $p = 0.018$, 95% CI: 0.0703, 0.7502. In contrast, no difference in attitudes toward self-sacrifice for in-group members emerged for weakly-fused participants as a function of thought direction, $B = -0.19$, $t(84) = -1.11$, $p = .270$, 95% CI: -0.5317, 0.1508. As anticipated, the data also revealed a significant main effect of identity fusion such that strongly-fused participants had more favorable attitudes towards self-sacrifice for in-group members, $B = 0.21$, $t(84) = 2.40$, $p = 0.018$, 95%, CI: 0.0380, 0.3985. No other significant effects emerged, $ps > 0.356$.

Figure 1. *Attitudes towards self-sacrifice as a function of Identity fusion and thoughts in favor vs against high levels of self-sacrifice (e.g., giving one´s life) (Study 1)*



Study 1 demonstrated that matching an individuals' disposition toward the group (e.g. strong feelings of fusion toward the group) with behaviors (e.g. extreme pro-group behaviors) led to greater thought use compared to a mismatch between disposition and situation. This supports our hypothesis from prior research (Huntsignger, 2013; Evans & Clark, 2012) indicating that a match between an individuals' disposition and situation led participants to rely on their thoughts more than when there was a mismatch. In turn, this increased reliance on thoughts had a powerful effect on attitudes, such that positive (negative) thoughts resulted in more favorable (unfavorable), but only when a person's disposition (Identity fusion) matched the situation (extreme pro-group behaviors). In contrast, thought reliance was lowest when participants' level of identity fusion mismatched the situation. These data are consistent with the idea that a mismatch between disposition and situation should reduce thought reliance relative to when disposition and situation are matched (see Briñol & Petty, 2009). It is important

to note that because the matching was induced after thoughts were generated it is strongly improbable that exposure to the identity fusion measure affected the valence or the number of participants' thoughts.

Critics might argue that our results merely show that strongly fused participants may have dispositionally higher confidence in their thoughts than weakly fused, therefore they are more likely to use their thoughts regardless of the situation. Although we acknowledge this is a possibility, recall that Study 1 used a very extreme type of sacrifice (i.e., giving one's life to save a fellow Spaniard), which we would expect should have a less powerful impact on weakly fused participants. This leads to the possibility that a similar outcome may emerge for weakly fused participants if the type of sacrifice required was more modest in nature. Study 2 sought to address this alternative explanation (e.g, dispositionally higher thought confidence for strongly-fused individuals) while also identifying conditions under which a match would emerge for individuals scoring low in fusion.

Study 2

The goal of Study 2 was to examine whether matching weak levels of fusion with the group to the behaviors they are aligned with (e.g. modest, less extreme pro-group behaviors) leads to greater thought use compared to a mismatch. We chose organ donation as the modest, less extreme self-sacrifice generating and asked participants to generate thoughts in favor or against this behavior. Although in general, donating organs is considered a rather risky behavior, it is important to note that Spain has been the world leader in organ donations for the last twenty five years, (Ministerio de Igualdad, Sanidad y Servicios Sociales, Organización Nacional de Trasplantes [ONT], 2016), and that the number of organ donors has increased every year. In fact, according to Spanish law 30/1979, October 27th, the organs of deceased citizens are extracted for donation by default unless there is explicit oposition expressed by the citizen. Thus, given how common it is within Spanish society to be an organ donor, we assumed this behavior was likely perceived as a significant yet relatively modest level of self-sacrifice.

A pilot study was conducted to provide empirical support for the assumption that donating a non-vital organ was a less extreme self-sacrifice than sacrificing one´s life. In this study, 20 participants were two questions: “How extreme do you think donating a non-vital organ is?” and “How extreme do you think sacrificing your life for someone else is?” (1 = Not extreme at all, to 7 = Very extreme). The presentation order of these items was counter-balanced. A within-subjects one-way ANOVA was run using the two items as the within-subjects factor. Results indicated that participants perceived sacrificing one´s life (M = 6.10, SD = 1.11) as significantly more extreme than donating a non-vital organ, (M = 3.65, SD = 2.27), $F(1, 19) = 22.60$, $p < 0.001$, $\eta_\rho^2 = .54$. Thus, these data confirm that referring to the donation of non-vital organs as a relatively modest level of self-sacrifice is appropriate.

All participants were explicitly instructed to generate thoughts in favor of vs. against a relatively modest level of self-sacrifice (e.g., donating a non-vital organ for an in-group member). Then, participants were asked to complete the verbal scale of identity fusion with their group and the self-reported measure of attitude towards self-sacrifice for in-group members.

Method

Participants and design. Eighty-four undergraduate students at the Universidad Nacional de Educación a Distancia, UNED, participated online in exchange for partial fulfillment of a course requirement (55% women, *Mage* = 37.62 years, *SD* = 11.31). Sample size was determined based on the number of participants collected during the week in which the study was posted. Participants were randomly assigned to a between-participants, 2 (Thought Direction: In favor of donation vs. Against donation) x Identity fusion measure design.

Procedure and Materials.

Study 2 used the same procedures as Study 1.

Independent variables.

Thought Direction. The thought listing task condition in Study 2 used similar instructions as in Study 1, but asked participants either to list three reasons in favor of or against donating a non-vital organ to an in-group member. Some examples of reasons in favor of donation included the following: "I could live without that organ while saving the other person´s life," and "Because I would like someone to donate it to me if I were that person". Some examples of reasons against donation were "The surgical procedure could be life-threatening to me too," and "Because I could need to donate that organ to a family member in the future". As in Study 1, an inspection of the reasons provided by participants showed no differences as a function of identity fusion scores.

Identity fusion. The same scale was used as in Study 1 ($\alpha = .91$). We submitted the Identify Fusion scale to a one-way ANOVA using thought direction as the predictor. As expected, identity fusion did not change as a function of the thought direction manipulation, $p = .88$, indicating that listing reasons in favor or against donating a non-vital organ to an in-group member did not alter reported levels of identity fusion.

Dependent Variables.

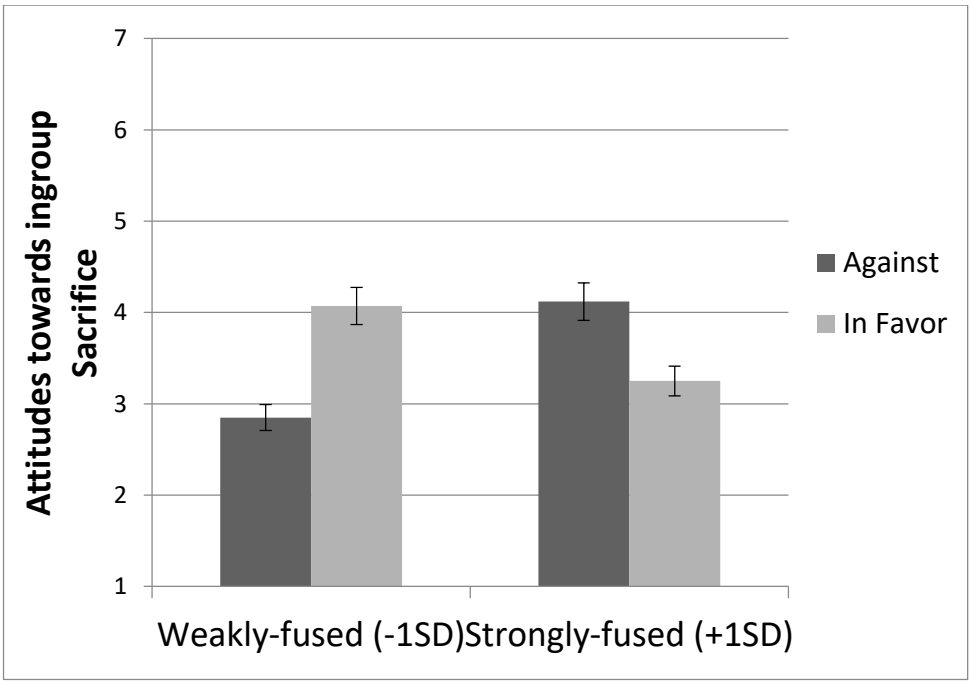
Attitudes toward self-sacrifice for in-group members. The same scale was used as in Study 1 ($\alpha = .74$).

Results and discussion

Attitudes toward self-sacrifice for in-group members. A linear regression analysis including thought direction (effect coding: -1 = against donating, 1 = in favor of donating), identity fusion (centered) and their interaction on attitudes toward self-sacrifice for in-group members revealed a significant two-way interaction, $B = -0.40$, $t(80) = -3.44$, $p < 0.001$, 95% CI: -0.6338, -0.1693. As Figure 2 shows, weakly fused participants displayed more favorable attitudes toward self-sacrifice for an in-group member when their thoughts were in favor of relative to against donating a non-vital organ, $B = 0.61$, $t(80) = 3.01$, $p = .003$, 95% CI: 0.2075, 1.0140. In line with expectations, the opposite pattern

emerged among strongly-fused participants, who displayed *less* favorable attitudes toward self-sacrifice for an in-group member in the thoughts in favor of relative to against donating a non-vital organ. This pattern would be expected if donating a non-vital organ produced so much doubt that strongly-fused individuals wanted to do the opposite of their thoughts, B = -0.43, $t(80)$ = -2.08, p = .040, 95% CI: -0.8489, -0.0195. No other significant effects emerged, ps > 0.51. Notably, this pattern is the exact oposite to the one obtained in Study 1, which tested the effects of matching within the context of a high self-sacrifice.

Figure 2. *Attitudes towards self-sacrifice as a function of Identity fusion and thoughts in favor vs against low levels of self-sacrifice (e.g., donating a non-vital organ) (Study 2)*



Study 2 revealed that matching individuals' dispositions (e.g. weakly feelings of fusion toward the group) with behaviors they are aligned with (e.g. modestely extreme pro-group behaviors) produced greater thought use compared to a mismatch between disposition and situation. This supports and reinforces the results of Study 1, once again demonstrating that a match between an individuals'

disposition and situation led participants to rely on their thoughts more than when there was a mismatch. As a consequence of this increased reliance on thoughts, an effect of thought direction on attitudes (i.e., more positive attitudes after generation of thoughts in favor vs against) was observed only when a person's level of identity fusion matched the situation.

By demonstrating the same matching effect observed in Study 1, but in this case for weakly fused participants, these data help to rule out the alternative explanation that individuals who score high in fusion may have dispositionally higher confidence in their thoughts and therefore are more likely to use their thoughts regardless of the situation. Although these data make the dispositional difference explanation somewhat unlikely, differences in the extremity of our self-sacrifice manipulation between Study 1 and 2 raise a further possibility. That is, perhaps the matching effects observed in Study 1 emerged because sacrifing one's life (vs. donating non-vital organs) highlighted the participants' awareness of their own mortality. Indeed, prior research has shown that mortality salience is able to polarize jugdments in several contexts (Rosenblatt, Greenberg, Solomon, Pyszczynski, & Lyon, 1989), which in turn may have encouraged strongly-fused participants to engage in extreme pro-group behavior. In order to address this alternative explanation, Study 3 sought to test whether matching effects between strongly fused and a relatively high sacrifice could be obtained without making the concept of one´s mortality salient.

Finally, because Studies 1 and 2 were somewhat limited by a relatively small and homogenous sample (i.e., college students), a further goal of Study 3 was to address these shortcomings by obtaining a larger sample drawn from from the general population.

Study 3

The basic goal of Study 3 was to replicate the pattern of effects that emerged in Studies 1 & 2 within a single design. This would then allow us to test the effects of both matching conditions (i.e., stronger fusion and high self-sacrifice, weaker fusion and low self- sacrifice) and mismatching

conditions (stronger fusion and low self-sacrifice, weaker fusion and high self-sacrifice) within the same design. Our expectation was that thoughts in the matching conditions should be more predictive of subsequent attitudes than thoughts in the mismatching conditions. A second goal was to test the generalizability of these patterns by changing participants' focus to thoughts about the self (vs. the in-group), to non-sacrifice-related thoughts (vs. sacrificing one's life or non-vital organs), and replacing the target group from "my country" to "my group of friends". Importantly, by changing participants' focus from thoughts related to self-sacrifice to non-sacrifice related thoughts about the self, this would allow us to empirically support the notion that differences in extremity (not mortality) are responsible for the effect.

Method

Participants and design. Two hundred and twelve Spaniards (77 % women, *Mage* = 41.97, *SD*= 10.84) were obtained from a pool of voluntary participants from Psychology students at Universidad Nacional de Educación a Distancia, UNED, as well as their friends and relatives, and randomly assigned to a between participants, 2 (Thought Direction: positive vs. negative) x Willingness to engage in an extreme self-sacrifice measure x Identity fusion measure design. Sample size was determined based on the number of participants collected during the two-week period in which the study was posted. These last two variables were used as within-subject continuous predictors.

Procedure and materials.

First, participants were asked to list three either positive or negative features about themselves. After completing the thought direction task, participants reported their levels of identity fusion with their group of friends, their willingness to engage in an extreme self-sacrifice for their friends, and their attitudes towards themselves. Finally, participants were debriefed and thanked.

Independent variables.

Thought Direction. In the *positive (negative) thoughts* condition, participants were asked to list three positive (negative) features or strengths (weaknesses) about themselves as friends. We told participants that this was an important task and, therefore, asked them to think carefully as they listed their reasons. In this study, Identity fusion was also measured after the experimental manipulation.

Willingness to engage in an extreme self-sacrifice. Participants reported their perceived *agreement with sacrifice* answering to the following 7-point (1 = Not at all willing, 7 = Absolutely Willing) Likert scale: "Now please answer the following questions about your group of friends: To what extent would you be willing to take the blame for a crime you a member of your group of friends commited and spend the rest of your life in prison?"

Identity fusion. Identity fusion with "my friends" ($\alpha = .76$) was measured using the same scale as in Studies 1 and 2, and completed after the thought direction manipulation. As expected, identity fusion did not change as a function of thought direction, $p > 0.24$.

Dependent Variables.

Attitudes towards the self as friends. Participants responded to six items asking them to rate the extent to which they considered themselves to be "with potential", "self-confident" and to what extent their attitudes towards themselves were "positive", "good", "favorable", and "in favor". Scales ranged from 1 (strongly disagree) to 7 (strongly agree), $\alpha = .90$. Items were coded so higher scores meant more favorable attitudes towards themselves as friends.

Results and Discussion

Linear regressions analyses were used to examine the effects on attitudes toward the self with thought direction, (effect coding: -1 = negative self-thoughts, 1 = positive self-thoughts), willingness to engage in an extreme self-sacrifice, identity fusion (both centered) and their interactions as our independent variables. We ran a Pearson correlation between the identity fusion measure and the willingness to engage in an extreme self-sacrifice measure, for too high a correlation between two

predictors could indeed stifle the validity of subsequent results. Because the correlation between the two was significant (r = .29, p < .001), we ran colinearity tests in a linear regression including thought direction, identity fusion, willingness to engage in an extreme self-sacrifice and their interactions with attitudes towards the self as the dependent variable to determine whether the validity of the results might be reduced because of excessive colinearity among predictors. All statistics were within normal ranges (*VIF* < 1.313, *tolerance* > 0.761), thus indicating that colinearity between predictors did not exceed levels that would compromise the validity of the results.

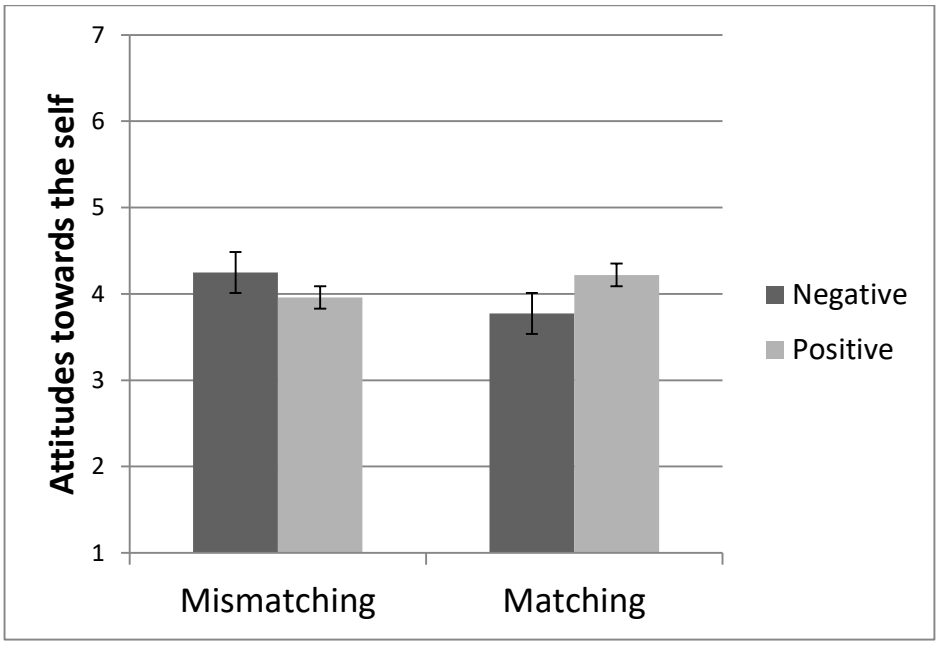
In order to simplify the presentation of the results, we grouped the predicted matching conditions (i.e. high level of self-sacrifice with stronger levels of fusion, and low level of self-sacrifice with weaker levels of fusion) and the predicted mismatching conditions (i.e. high level of self-sacrifice with weaker levels of fusion and low level of self-sacrifice with stronger levels of fusion) into a Matching vs Mismatching independent variable. We also added a factor to account for the type of matching (i.e., match to high levels of self-sacrifice or match to low levels of self-sacrifice). Besides offering a simpler presentation of the results, an additional advantage of conducting the analysis in this way is advantageous because it allows us to compare the extent to which the type of matching moderates our results. For example, if a Matching x Thought Direction interaction emerged, this would support our hypothesis that matching increases thought use. The three way interaction (Matching x Thought Direction x Type of Matching) would test whether the matching effect was comparable across types of matching. This approach to analyzing matching effects has been succesfully used in previous research (Petty & Wegener, 1998).

Attitudes toward the self. Our next step was to conduct a 2 (Thought Direction: Positive vs. Negative) × 2 (Matching: Match vs. Mismatch) × 2 Type of Match (High vs. Low level of self-sacrifice) ANOVA on attitudes toward the self (i.e., thought direction manipulation)[2].

Consistent with our predictions, results revealed a two-way interaction between Thought Direction and Matching, F (1, 204) = 7.039, p = 0.009, η2 = .033. Importantly, this interaction was not qualified by the Type of Match, F (1, 204) = .016, p = 0.687 This suggests that regardless of whether participants were matched on level of self-sacrifice and identity fusion or mismatched on these variables, neither pairing differentially influenced attitudes toward the self. As illustrated in Figures 3A & 3B, for the matching condition, participants who wrote positive thoughts about themselves reported significantly more favorable attitudes toward themselves (M = 4.20, SD = 0.11) than those who wrote negative thoughts about themselves (M = 3.80, SD = 0.12), F(1, 204) = 5.34, p = 0.022, η2 = .026. However, for the mismatching condition, no significant differences in attitudes emerged between those who wrote positive thoughts about themselves (M = 3.92, SD = 0.14), and those who wrote negative thoughts about themselves (M = 4.23, SD = 0.14) F(1, 204) = 2.28, p = 0.133, η2 = .01.

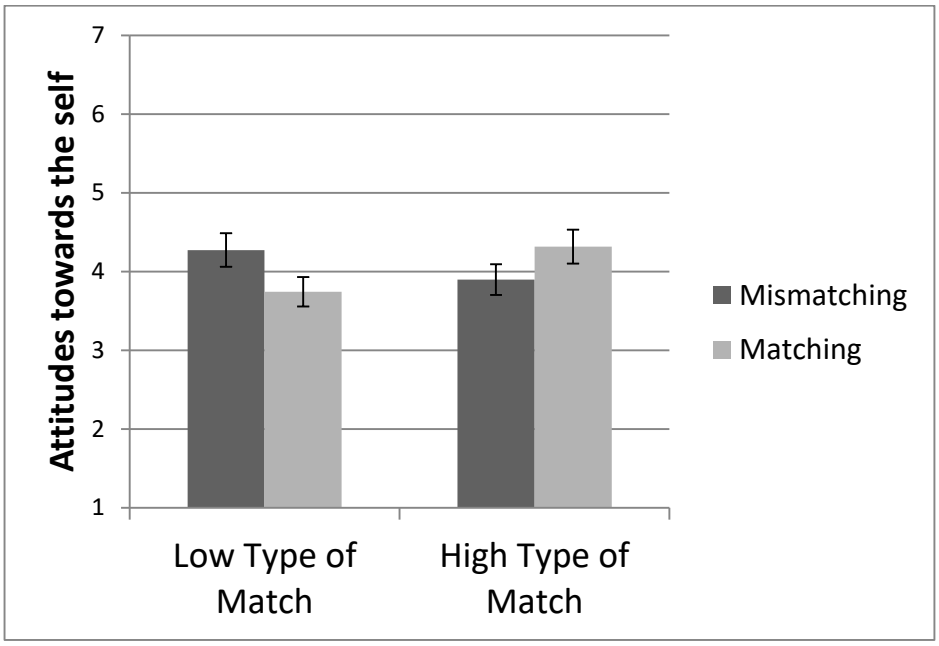
[2] Alternative regression analyses considering Identity fusion, Thought Direction and Level of Sacrifice, and its correspondent interactions, yielded similar significant results in the outcome measures and did not alter the main conclusions in all subsequent analysis in this chapter.

Figure 3A. *Attitudes towards oneself as a function of Matching and Thought Direction (Study 3).*



A significant two-way interaction between Matching and Matching Type also emerged, $F(1, 204) = 12.67$, $p < 0.001$, $\eta 2 = .024$. For the low Type of Match, participants in mismatching conditions reported significantly more favorable attitudes towards themselves ($M = 4.26$, $SD = 0.13$) than did those in matching conditions ($M = 3.89$, $SD = 0.15$), $F(1, 204) = 9.28$, $p = 0 .003$, $\eta 2 = .044$. For the high Type of Matching, participants in matching conditions reported significantly more favorable attitudes towards themselves ($M = 4.29$, $SD = 0.12$) than did those in mismatching conditions ($M = 3.72$, $SD = 0.12$), $F(1, 204) = 4.17$, $p = 0.042$, $\eta 2 = .020$. No further effects reached significance, $F (1, 204) < 0.98$, $p > 0.321$.

Figure 3B. *Attitudes towards oneself as a function of Matching and Type of Matching (Study 3)*



The results of Study 3 showed that matching individuals' dispositions (level of identity fusion) to the behaviors they are aligned with (level of self-sacrifice for in-group members) produced greater thought use compared to mismatching these variables. These data support our hypothesis from prior research, suggesting that matching dispositional and situation varibles leads people to rely on their thoughts more than when there is a mismatch.

As a consequence of this increased reliance on thoughts, an effect of thought direction on attitudes (i.e., more positive attitudes after generation of positive vs. negative thoughts) was observed only when a person's level of identity fusion matched the situation. People relied less on the thoughts they listed when their level of identity fusion mismatched the situation. This is consistent with the idea that mismatching may have undermined the extent to which participants trusted their thoughts relative to when person and situation variables were matched (see Briñol & Petty, 2009). It is important to note that because matching was induced after thoughts were already generated (the identity fusion scale came after the generation of thoughts), it is unlikely that the exposure to the fusion measure affected the valence or the number of participants' thoughts.

Although results of Studies 1, 2 and 3 showed that matching occurred for both participants who scored stronger in fusion with their in-group (Study 1 & 3) and for participants those who scored weaker in fusion with their in-group (Study 2 & 3), no evidence about the underlying mechanism of this matching effect was provided. Previous research has shown that the extent to which participants use their thoughts when forming their judgments in response to matching procedures can be mediated by thought confidence (Huntsinger, 2013; Evans & Clark, 2012). Therefore, our expectation was that thought confidence should mediate the relationship between matching conditions and attitudes towards self-sacrifice for in-group members. As seen in chapters 2 & 3 in this doctoral dissertation, other measures of confidence have shown to mediate the effect that fusion has on judgments. Thus, Study 4 was conducted to test this mechanism and also to replicate the matching effects observed for participants those who scored high and low in fusion found in Studies 1, 2 & 3.

Study 4

The main goal of Study 4 was to identify the underlying mechanism that might explain why matching the level of Identy Fusion with the extremity of self-sacrifice for the in-group influences attitudes toward self-sacrifice for in-group members. Our expection was that the matching conditions would reveal greater thought confidence compared to the mismatching conditions, and that thought confidence would mediate the effect of the thought direction manipulation on attitudes toward self-sacrifice for in-group members. Demonstrating that matching effects does not depend on the level of identity fusion, we do not expect differences according the nature of the match (matching stronger levels of fusion with high self-sacrifice and weaker levels of fusion with low self-sacrifice).

Method

Participants and design.

Five hundred and nine Spaniards (65.1 % women, mean age= 35.83, SD= 12.55) were obtained from a pool of voluntary participants from Psychology students at Universidad Nacional de Educación

a Distancia, UNED as well as their friends and relatives. Sample size was determined based on the number of participants collected during the week in which the study was posted. Participants were randomly assigned to a 2 (Reasons to self-sacrifice for an in-group member: In favor vs. Against) x 2 (Level of self-sacrifice: High vs. Low) x Identity fusion measure between participants design.

Procedure and materials.

Participants were randomly assigned to either list reasons in favor or against low levels of self-sacrifice (e.g., in favor vs. against donating a non-vital organ) or in favor or against high levels of self-sacrifice (e.g., in favor vs against sacrificing one´s life). Next, participants reported their levels of identity fusion with their country. Following this, participants rated the extent to which they were confident in their thoughts as well as attitudes towards self-sacrifice for in-group members. Finally, participants were debriefed, thanked and dismissed.

Independent variables.

Thought Direction. In the *in favor* condition, participants were asked to list three reasons in favor of either donating a non-vital organ or sacrifing their life for a fellow Spaniard. In the *against* condition, participants were asked to list three reasons against either donating a non-vital organ or sacrificing their life for a fellow Spaniard. Participants were informed that this was an important task andto think carefully as they listed their reasons.

Level of Self-Sacrifice. In the *low self-sacrifice* condition, participants were asked to list three reasons about donating a non-vital organ to a fellow Spaniard. In the *high self-sacrifice* condition, participants were asked to list three reasons about sacrificing their life for a fellow Spaniard.

Identity fusion. Identity fusion with "my country" ($\alpha = .87$) was measured using the same scale as in the prior Studies, and measured after the experimental manipulation. As expected, identity fusion did not change as a function of thought direction, level of self-sacrifice or their interaction, *ps* > 0.42.

Dependent Variables.

Attitudes toward self-sacrifice for in-group members. Attitudes were assessed on the same six 7-point scales as in Studies 1 & 2 ($\alpha = .79$).

Thought confidence. Participants were asked to think back to the thought-listing task and report the confidence they had in their thoughts. Self-ratings were provided on four items, including confidence, validity, certainty, and difficulty. Responses were measured on 7-point scales (see Evans & Clark, 2012; Clark et al, 2013), where 1 represented *None at all/Not at all valid/Not at all certain/Not at all difficult*, and 7 represented *Very much/Extremely valid/Very certain/Very difficult*, respectively. A composite of thought confidence was formed by reverse coding the negative item (difficulty), then averaging responses to these four measures ($\alpha = .77$).

Results and Discussion

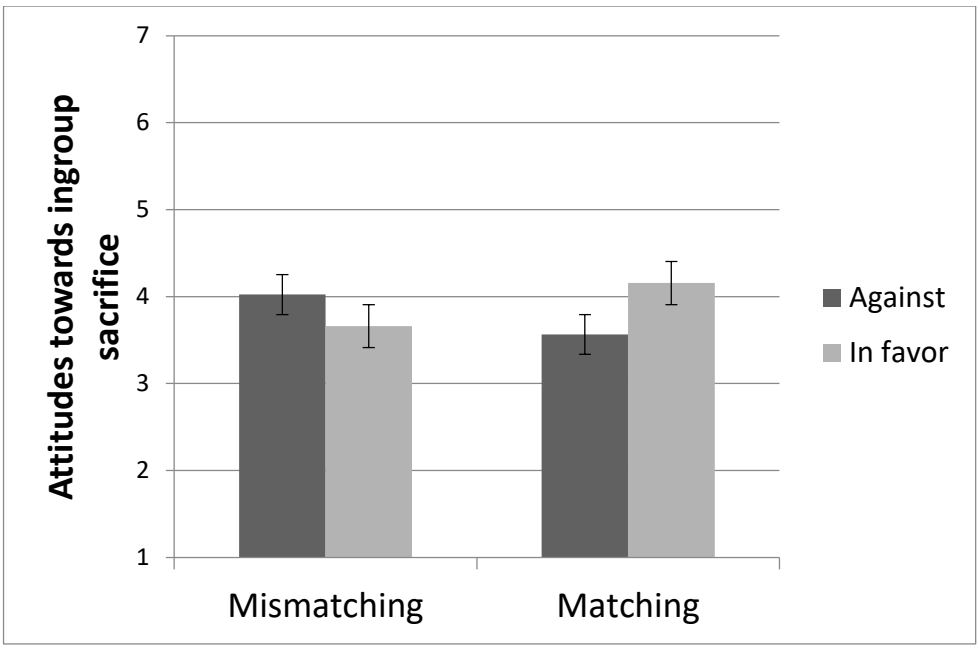
In order to simplify the presentation of the results, we used the same grouping procedures described in Study 3.

Attitudes toward self-sacrifice for an in-group member. We then ran a 2 (Thought Direction: In favor vs. Against sacrifice) × 2 (Matching: Match vs. Mismatch) × 2 Type of Match (High vs. Low level of self-sacrifice) ANOVA on attitudes toward self-sacrifice for the in-group.

Results revealed a two-way interaction between Thought Direction and Matching, F (1, 501) = 24.14, $p < 0.001$, $\eta 2 = .046$. Importantly, this interaction was not moderated by the type of match, F (1, 501) = .01, $p = 0.916$. This suggests that regardless of whether participants were matched on level of self-sacrifice and identity fusion or mismatched on these variables, neither pairing differentially influenced attitudes toward self-sacrifice for the in-group. As illustrated in Figure 4A, for the matching condition, participants who wrote thoughts in favor of self-sacrifice reported significantly more favorable attitudes toward self-sacrifice for in-group members ($M = 4.17$, $SD = 0.10$) than did those who wrote thoughts against self-sacrifice ($M = 3.55$, $SD = 0.10$), $F(1, 501) = 18.56$, $p < 0.001$, $\eta 2 = .036$. As expected, for the mismatching condition, those who wrote thoughts against self-sacrifice reported more

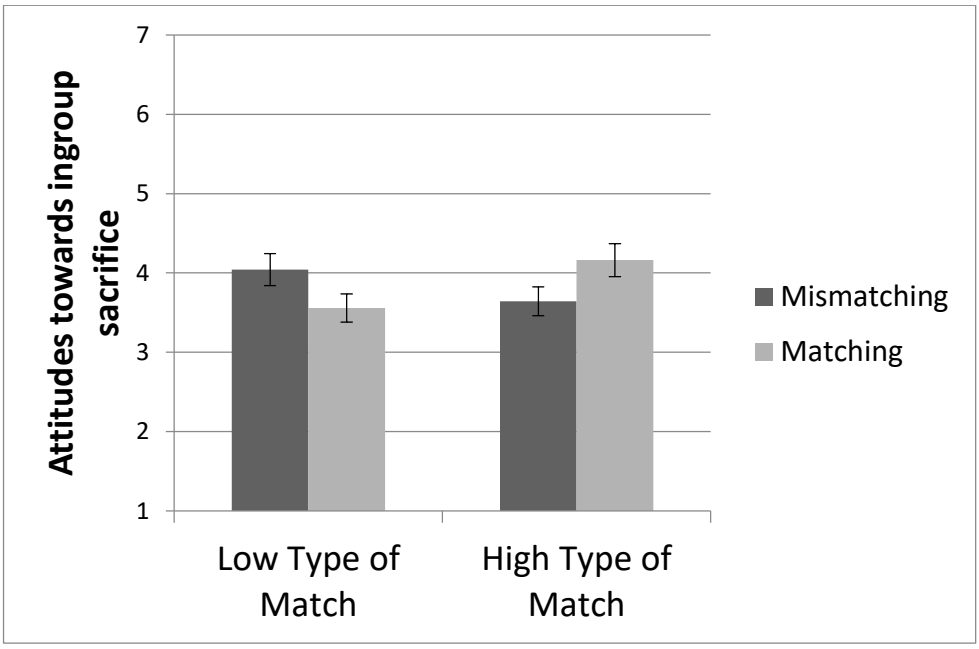
favorable attitudes toward self-sacrifice for the in-group (M = 4.02, SD = 0.09) than those who wrote thoughts in favor of self-sacrifice (M = 3.66, SD = 0.09), $F(1, 501) = 6.70$, $p = 0.010$, $\eta 2 = .013$.

Figure 4A. *Attitudes towards self-sacrifice as a function of Matching and Thought Direction (Study 4).*



A significant two-way interaction between Matching and Type of Match also emerged, $F(1, 501) = 26.72$, $p < 0.001$, $\eta 2 = .05$ (see Figure 4B). For the low Type of Match, participants in mismatching conditions reported significantly more favorable attitudes towards self-sacrifice (M = 4.03, SD = 0.09) than those in the matching conditions (M = 3.54, SD = 0.10), $F(1, 501) = 12.33$, $p < 0.001$, $\eta 2 = .024$. For the high level of sacrifice Type of Match, participants in matching conditions reported significantly more favorable attitudes towards self-sacrifice (M = 4.18, SD = 0.10) than did those in mismatching conditions (M = 3.64, SD = 0.09), $F(1, 501) = 14.42$, $p < 0.001$, $\eta 2 = .028$. No further effects reached significance, $F(1, 501) < 1.90$, $p > 0.170$.

Figure 4B. *Attitudes towards self-sacrifice as a function of Matching and Type of Matching (Study 4).*



Thought Confidence. A 2 (Thought Direction: Positive vs. Negative) × 2 (Matching: Match vs. Mismatch) × Type of Match (High vs Low level of self-sacrifice) ANOVA on thought confidence was run. Results revealed the predicted main effect for Matching, F (1, 501) = 4.49, p = 0.035, η2 = .009. Participants in the Matching conditions reported significantly higher thought confidence (M = 4.95, SD = 1.13) than those in the Mismatching condition (M = 4.73, SD = 1.16). Results also revealed a main effect for Thought Direction, F (1, 501) = 7.99, p = 0.005, η2 = .016, meaning that participants who were assigned to the *thoughts in favor of self-sacrifice* condition (M = 4.39, SD = 0.96) had significantly higher thought confidence than those who were assigned to the *thoughts against self-sacrifice* condition (M = 4.69, SD = 1.30) (See Figures 5A and 5B). No further 3-way or 2-way interaction effects reached significance, F (1, 501) < 3.09, p > 0.080.

Figure 5A. *Thought Confidence as a function of Matching and thoughts in favor vs against low levels of self-sacrifice (Study 4).*

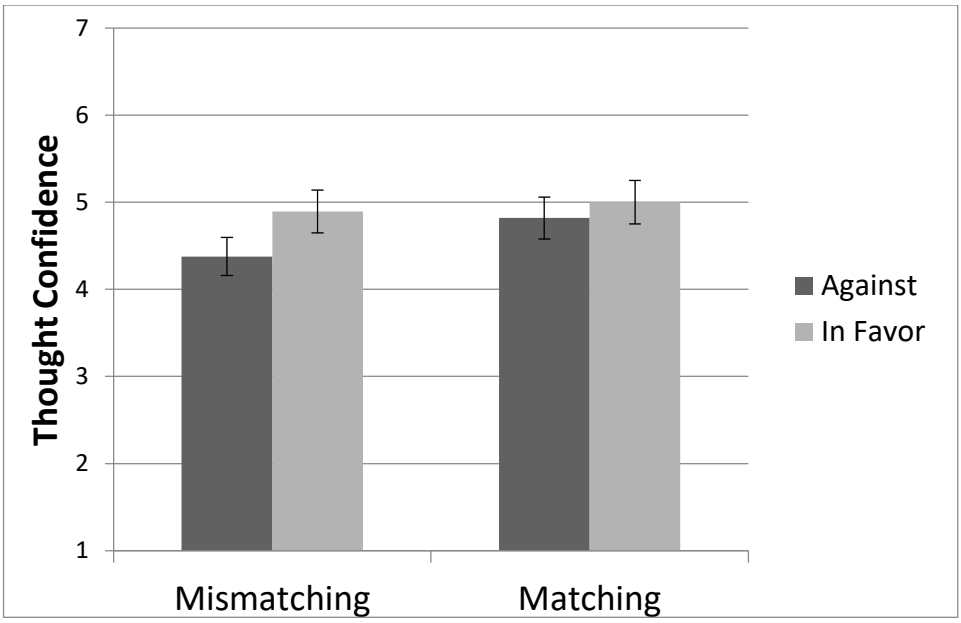
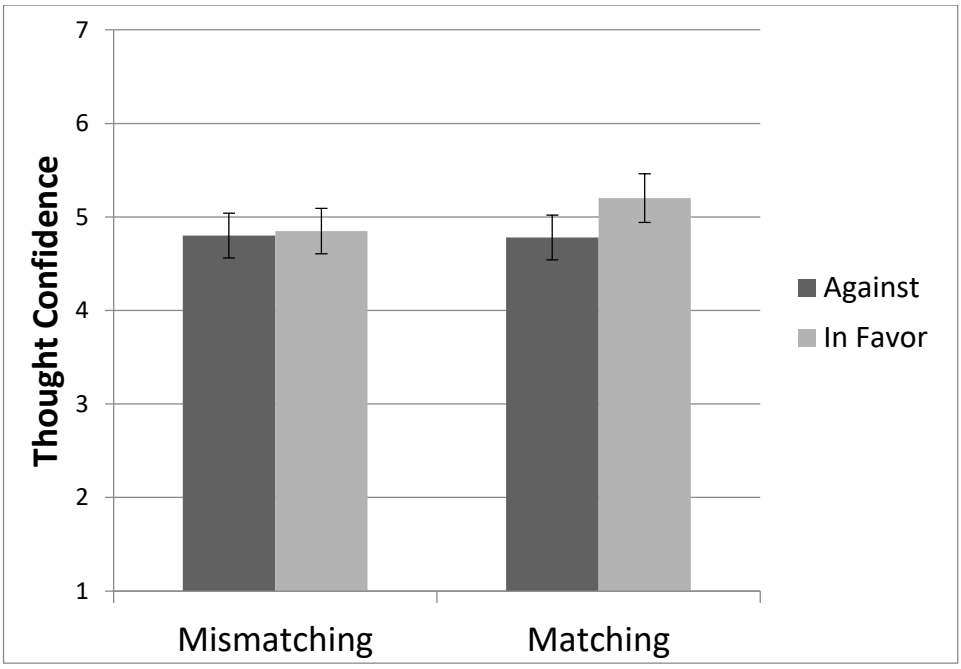


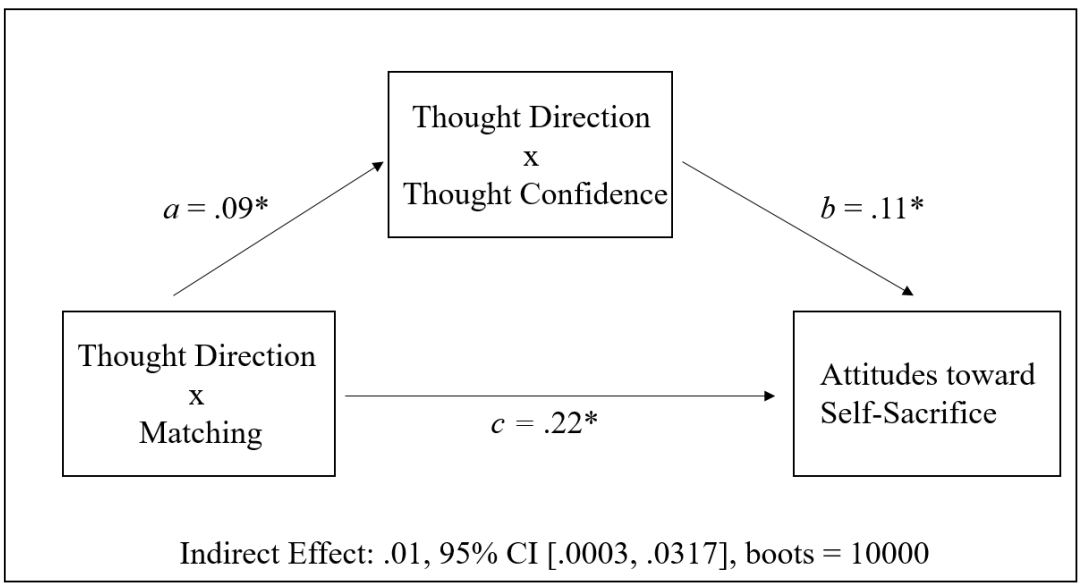
Figure 5B. *Thought Confidence as a function of Matching and thoughts in favor vs against high levels of self-sacrifice (Study 4).*



Thought Confidence mediates the effect of thought direction on attitudes toward self-

sacrifice. In order to examine whether thought confidence mediated the effect of the key theorized interaction on attitudes, we conducted a mediated moderation test using bootstrapping methods (Muller, Judd, & Yzerbyt, 2005). In this procedure, both thought direction (i.e., - 1 = thoughts against sacrifice, 1 = thoughts in favor of sacrifice) and matching (-1 = mismatching, 1 = matching) were contrast coded, and thought confidence was mean-centered. In order to test the hypothesized mediation by thought confidence, we conducted a biased corrected bootstrapping procedure with 10,000 bootstrap re-samples using Hayes process macro (model 4) (Preacher & Hayes, 2004; Shrout & Bolger, 2002). In this analysis, thought direction x matching was the independent variable, attitudes toward self-sacrifice use was the dependent variable, and thought direction x thought confidence was the mediating variable (see figure 6). As predicted, the indirect effect through thought direct x thought confidence significantly was significant, (Indirect Effect a x b = .10, $CI_{95\%}$ = 0.0003, 0.0317). Thus, these data confirmed that the predicted mediation by thought confidence was supported.

Figure 6. *Attitudes towards self-sacrifice mediate interactive effects of Fusion X Thought Direction on Willingness to self-sacrifice for five in-group members (Study 4). CI = Confidence Interval.*



The results of Study 4 showed that matching individuals' dispositions (level of identity fusion) to the behaviors they are aligned with (level of self-sacrifice for in-group members) produced greater thought use compared to mismatching these variables. These data not only support our hypothesis from prior research, suggesting that matching dispositional and situation varibles leads people to rely on their thoughts more than when there is a mismatch, but also provide the first evidence that this process is mediated by thought confidence. As a consequence of this increased reliance on thoughts, an effect of thought direction on attitudes (i.e., more positive attitudes after generation of thoughts in favor vs against) was observed only when a person's level of identity fusion matched the situation. People did not rely on the thoughts they listed when their level of identity fusion mismatched the situation. This is consistent with the idea that mismatching undermined the extent to which participants trusted their thoughts relative to when person and situational variables were matched (see Briñol & Petty, 2009). Once again, it is important to note that because matching was induced after thoughts were already generated (the identity fusion scale came after the generation of thoughts), it is unlikely that the exposure to the fusion measure affected the valence or the number of participants' thoughts.

Study 4 replicates the critical findings of Studies 1, 2 & 3, and importantly, specifices the conditions under which either interaction could take place. These data revealed that when generating thoughts about pro-group behaviors that are not considered extreme, weakly fused participants are more likely to use their thoughts when forming their attitudes. However, when generating thoughts about pro-group behaviors that are considered extreme (e.g., sacrificing one´s life), strongly fused participants are more likely to use their thoughts when forming their attitudes. Importantly, Study 4 provides evidence consistent with previous literature (Huntsinger, 2013; Briñol et al., 2007; Evans & Clark, 2012) that thought confidence serves as the underlying psychological mechanism driving the effects of our matching/mismatching effects on attitudes toward self-sacrifice.

General Discussion

Summary of Findings

Across four Studies, we examined the relationship between levels of Identity fusion, thoughts about different levels of sacrifice, and the subsequent impact of expressed thoughts on evaluative judgments. In Study 1, participants generated either positive or negative thoughts about a high level of sacrifice (giving one´s life) for an in-group member and reported their levels of Identity fusion with their in-group. We found strongly fused participants relied more on the thoughts they had reported when forming subsequent evaluations of in-group self-sacrifice compared to weakly fused participants. In Study 2, participants were asked to generate thoughts in favor or against a relatively lower level of sacrifice (donating a non-vital organ) for an in-group member, and reported their levels of Identity fusion with their in-group. Here we found that weakly fused participants relied more on the thoughts they had reported when forming subsequent evaluations of in-group self-sacrifice compared to strongly fused participants.

Taken together, we speculated that, the results of Studies 1 and 2 were compatible with the idea that the level of identity fusion with one´s group matched the level of sacrifice for the in-group as reflected by one's thoughts toward the level of sacrifice (i.e., stronger fusion with thoughts about high level of sacrifice and weaker fusion with thoughts about low level of sacrifice). In turn, matching dispositional and situational variables led to greater reliance on thoughts when forming attitudes about in-group self-sacrifice. Importantly, these data revealed that one's level of Identity fusion can have opposite effects on attitudes towards self-sacrifice depending on the circumstances. That is, when the topic was more extreme (giving one´s life), stronger (weaker) levels of fusion with the group enhanced (reduced) the subsequent impact of thoughts on attitudes. In contrast, when a topic was relatively less extreme (giving a non-vital organ), stronger (weaker) levels of fusion with the group attenuated (facilitated) the use of thoughts.

In Study 3, we measured willingness to engage in an extreme pro-group sacrifice and thus replicated Study 1(matching high fusion and high levels of sacrifice) as well as Study 2 (matching weaker fusion and low levels of sacrifice) within the same design.

In Study 4, we randomly assigned participants to write thoughts about a relatively high (giving one´s life) or a relatively low (donating a non-vital organ) sacrifice for an in-group member. Participants were then asked to report their levels of identity fusion with their in-group. In addition to replicating the pattern of results found in the previous Studies, participants reported more confidence in their thoughts when their level of Identity fusion matched their level of sacrifice for the in-group than when a mismatch between these variables occurred.

In line with the self-validation hypothesis, we argued that people are likely to rely on their thoughts more when they do something that matches or fits their own nature rather than when the actions they engage in do not fit. Thus, we predicted and found that thought use was increased when a person stronger (vs. weaker) in Identity fusion generated thoughts about a relatively high (vs low) type of sacrifice. That is, we found that individual differences in Identity fusion and generation of thoughts about sacrifice interacted in predicting reliance on thoughts. Ironically, even though both of these variables have been associated with extreme behaviors, they could also lead to milder outcomes (e.g., more unfavorable attitudes towards self-sacrifice) when operating through validation and matching.

Implications

Among other things, the results of the current Studies are important because they reveal that not only can thinking about reasons to self-sacrifice impact the initial thoughts that people have, but it can also influence whether or not people rely on their thoughts. For example, based on differences in thought confidence, strongly-fused individuals may have more unfavorable attitudes towards self-sacrifice after generating thoughts in favor of certain relatively low levels of self-sacrifice. Thus, if a pro-donation message was constructed to elicit primarily positive thoughts, the research just described predicts that

this message would be more effective for weakly fused individuals than for strongly fused because the positive thoughts to the message would be relied upon more.

Applications

According to prior research on self-validation effects, the influence of matching through this meta-cognitive validation process is particularly likely to occur if the situation involves a high degree of thought and the match becomes salient after thought generation (Briñol & Petty, 2009).Also, prior research has shown various kinds of matching can lead to attitude change through different processes under other circumstances (see Briñol, & Petty, 2006; Petty, et al., 2000, for reviews of the processes underlying matching effects in persuasion). Moreover, and in concert with the ELM, the process by which matching works to influence judgments is different depending on the extent of one's thinking within that context. For instance, when elaboration is low (e.g., due to the presence of distractions), a match is more likely to serve as a simple cue ("If it links to my values, it must be good") and promote more positive attitudes regardless of the message (DeBono, 1987). When elaboration is not constrained to be high or low and the match is salient prior to the message rather than after, matching messages to individual differences can increase message processing (Petty & Wegener, 1998). This means that when the arguments are strong, matching should lead to more persuasion, but when the arguments are weak, matching should lead to less persuasion (the opposite effect). Also, when elaboration is high and the match is salient prior to the message, matching can bias processing (e.g., Lavine & Snyder, 1996). Of course, when elaboration is high and the match is salient following the message, matching can affect reliance on one's thoughts as shown in the current research (see also, Evans & Clark, 2012). Specifying the different ways in which matching can affect judgment is important because different mechanisms can lead to different outcomes (e.g., more or less argument quality effects depending on timing, as described above).

In these studies we revealed that when strongly fused participants generated thoughts in favor of

a behavior implying a highy extreme sacrifice for an in-group member, they showed more favorable attitudes towards self-sacrifice. In contrast, when weakly fused participants generated thoughts against a behavior implying a relatively high sacrifice for an in-group member, they showed more unfavorable attitudes and less willingess to die in order to save in-group members in the trolley dilemma. This type of reaction against self-sacrifice by strongly fused participants shows that these individuals´ tendency towards self-sacrifice is not a thoughtless one; rather, it in fact can work through strongly-demanding and thoughtful processes (e.g., Self-Validation; Petty, Briñol, & Tormala, 2002; Briñol & Petty, 2003). This line of research also reveals that strongly fused participants show less thought reliance in the context of relatively low sacrifices than in relatively high sacrifices, suggesting that they hold thoughts about relatively low sacrifices with less confidence, possibly because that is not their preferred action on behalf of the group, or the one they feel they are meant to engage in. This approach helps ensure that individuals who score high in fusion are selective when they behave on behalf of the group.

Limitations and Future Directions

The present research has some limitations that should be mentioned. First, future studies should research whether there may be features of sacrifice for the in-group other than extremity that may match with Identity fusion, such as mortality, target group, self-harm, etc. It would also be interesting to use direct behavioral measures rather than trolley dilemma responses. Although these dilemmas have shown to be useful enough in moral decision-making (Bleske-Rechek, Nelson, Baker, Remiker & Brandt, 2010; Valdesolo & DeSteno, 2006), their usefulness has been also called into question by some authors (Bauman, MacGraw, Bartels & Warren, 2014; Kahane, 2015).

Taken together, these results demonstrate how thought validation can influence attitudes through matching dispositional (Identity fusion) and situational variables (pro-group behavior) of varying levels of extremity. A matching hypothesis is a compelling candidate to explain the current set of results, leading us to conclude that identity fusion can influence pro-group behaviors not only through direct,

and more heuristic, mechanisms, but also by affecting pro-group behavior through a meta-cognitive process, namely self-validation.

CHAPTER 5

General Discussion

Summary of findings

Throughout these chapters, we have tried to shed light on identity fusion processes in novel ways. Literature on this theory has not considered the potential role of confidence before, and how it may explain the relationship between fusion and pro-group behavior under certain circumstances.

Specifically, Chapter 2 examines how differences in a different form of confidence (e.g., personal agency, the certainty that one´s actions will make a difference) may predict changes in levels of identity fusion.

Chapter 3 shows how strongly-fused individuals attenuate their willingness to fight and die for the in-group whe they learn that other in-group members are as willing to self-sacrifice as themselves, plus also holding similar moitivations for it. Additionally, this chapter highlights how strongly-fused individuals are confident enough in their willingness to self-sacrifice to delay it if it benefits the group.

Finally, Chapter 4 highlights differences in thought confidence shown by persons scoring strongly (vs weakly) in fusion according to the behaviors they are asked to reflect upon. Strongly-fused participants display higher levels of confidence in thoughts generated about extreme pro-group behaviors, whereas wekly fused display higher levels of thought confidence in thoughts generated about non-extreme pro-group behaviors.

Taken together, these chapters show how identity fusion and its correlates can, under the right set of circumstances, depart from the automatic, thoughtless, visceral approach it has taken over the last few years. These chapters offer new mediators, boundary conditions, and outcomes that had never been considered before in the literature of identity fusion, and which all suggest that pro-group behavior may happen as a consequence of calculated, paused, involved thinking rather than emotional, rushed, automatic reactions. This doctoral dissertation shows that strongly-fused individuals generally choose to act in favor of the group, but may choose not to if the sacrifice does not match who they are, or if it´s not entirely benefitial to the group.

Limitations and future research

This thesis explored some of the roles that confidence may take as a mediating variable of the effect of experimental manipulations on identity fusion (package 1), moderating the relationship between identity fusion and pro-group behavior by leading to delay of action (package 2) or to action (package 3).

This ten-study doctoral dissertation is, of course, not free from limitations. Taking the three packages together, the first potential limitation refers to the external validity of this research. Although the fact that all studies in this dissertation include random assignment to some experimental conditions providees support for its internal validity, the question of whether these findings apply to a more natural setting still remains. Future research should measure the variables we have accumulated in this research (confidence, identity fusion, attitudes towards sacrifice, willingness to fight and die for the in-group) in both a setting and with samples that may provide a higher degree of external validity (e.g., military, sports fans at sport events, participants in protests, etc). The second limitation is not entirely independent from the first one. All studies in this doctoral dissertation measure relevant constructs that have shown to relate to actual pro-group behavior (attitudes, willingness to engage in extreme sacrifices, responses to extreme versions of trolley dilemmas), but none of them measure actual pro-group behavior. Future research should address this limitation by including real, harmless pro-group behaviors in its studies.

Future research should also explore what variables and/or mechanisms may be responsible for the different roles confidence plays in each of the three distinct packages. For instance, it may be that confidence in certain judgments leads to act on them or to delay action depending on a number of different factors. Specifically, knowing that fused individuals primary motivation is to benefit the group, confidence may predict behavior or delay of such depending on whether fused individuals estimate that getting actively involved will help the group or not. Similarly, it may be that some unexplored individual difference among strongly-fused individuals explain the different roles that confidence may take in the

different packages of this dissertation. Given the fact that identity fusion is not a personality variable, it is plausible to think that there may be some personality differences within fused individuals, and that these differences may account for a significant amount of variance in fused-individuals´ behavior.

Conclusions

This thesis merges two lines of research separated so far: self-validation and identity fusion. Across three packages of studies, we showed several different roles that confidence can play in explaining the relation between identity fusion and pro-group behaviors and in predicting identity fusion itself. These findings maintain consistency with previous literature both on persuasion and fusion while also providing with novel approaches and findings to both traditions in social psychology.

References

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of Other in the Self Scale and the structure of interpersonal closeness. *Journal of personality and social psychology*, *63*(4), 596-612. http://dx.doi.org/10.1037/0022-3514.63.4.596

Avnet, T., Laufer, D., & Higgins, E. T. (2013). Are all experiences of fit created equal? Two paths to persuasion. *Journal of Consumer Research, 23*, 301-316. http://dx.doi.org/10.2139/ssrn.1763396

Balance de Actividad de la Organización Nacional de Transplantes. Nota de prensa (2016. January 12th). *Retrieved from* http://www.ont.es/Documents/Balance_Actividad_2015.pdf

Bandura, A. (2006). Toward a psychology of human agency. *Perspectives on psychological science*, *1*(2), 164-180. https://doi.org/10.1111/j.1745-6916.2006.00011.x

Bauman, C. W., McGraw, A. P., Bartels, D. M., & Warren, C. (2014). Revisiting external validity: Concerns about trolley problems and other sacrificial dilemmas in moral psychology. *Social and Personality Psychology Compass*, *8*(9), 536-554. http://dx.doi.org/10.1111/spc3.12131

Besta, T., Gómez, A., & Vázquez, A. (2014). Readiness to deny group's wrongdoing and willingness to fight for its members: the role of Poles' identity fusion with the country and religious group. *Current Issues in Personality Psychology, 2*, 49-55. https://doi.org/10.5114/cipp.2014.43101

Blankenship, K. L., Nesbit, S. M., & Murray, R. A. (2013). Driving anger and metacognition: The role of thought confidence on anger and aggressive driving intentions. *Aggressive behavior*, *39*(4), 323-334. http://dx.doi.org/10.1002/ab.21484

Bleske-Rechek, A., Nelson, L. A., Baker, J. P., Remiker, M. W., & Brandt, S. J. (2010). Evolution and the trolley problem: People save five over one unless the one is young, genetically related, or a

romantic partner. *Journal of Social, Evolutionary, and Cultural Psychology, 4*(3), 115-127. http://dx.doi.org/10.1037/h0099295

Brehm, J. W. (1966). A theory of psychological reactance. *Oxford, England: Academic Press.*

Brewer, M. B. (1979). In-group bias in the minimal intergroup situation: A cognitive-motivational analysis. *Psychological Bulletin, 86(2),* 307-324. http://dx.doi.org/10.1037/0033-2909.86.2.307

Brewer, M. B., & Kramer, R. M. (1985). The psychology of intergroup attitudes and behavior. *Annual review of psychology*, *36*(1), 219-243. https://doi.org/10.1146/annurev.ps.36.020185.001251

Briñol, P., DeMarree, K. G., & Petty, R. E. (2010). *Processes by which confidence (vs. doubt) influences the self*. Psychology Press.

Briñol, P., & Petty, R. E. (2003). Overt head movements and persuasion: A self-validation analysis. *Journal of Personality and Social Psychology*, *84*(6), 1123-1139. http://dx.doi.org/10.1037/0022-3514.84.6.1123

Briñol, P., & Petty, R. E. (2006). Fundamental processes leading to attitude change: Implications for cancer prevention communications. *Journal of Communication, 56,* 81-104. http://dx.doi.org/10.1111/j.1460-2466.2006.00284.x

Briñol, P., & Petty, R. E. (2009). Persuasion: Insights from the self-validation hypothesis. *Advances in experimental social psychology*, *41*, 69-118. https://doi.org/10.1016/S0065-2601(08)00402-4

Briñol, P., & Petty, R. E. (2012). The history of attitudes and persuasion research. In A. Kruglanski & W. Stroebe (Eds.), *Handbook of the history of social psychology* (pp. 285-320). New York: Psychology Press.

Briñol, P., Petty, R. E., & Barden, J. (2007). Happiness versus sadness as a determinant of thought confidence in persuasion: A self-validation analysis. *Journal of Personality and Social Psychology*, *93*(5), 711-727. http://dx.doi.org/10.1037/0022-3514.93.5.711

Briñol, P., Petty, R. E., & Tormala, Z. L. (2006). The malleable meaning of subjective ease. *Psychological Science*, *17*(3), 200-206. https://doi.org/10.1111/j.1467-9280.2006.01686.x

Briñol, P., Petty, R. E., & Requero, B. (2017). Aggressive primes can increase reliance on positive and negative thoughts affecting self-attitudes. *Self and Identity*, *16*(2), 194-214. http://dx.doi.org/10.1080/15298868.2016.1255251

Briñol, P., Petty, R. E., Valle, C., Rucker, D. D., & Becerra, A. (2007). The effects of message recipients' power before and after persuasion: A self-validation analysis. *Journal of Personality and Social Psychology*, *93*(6), 1040-1053. http://dx.doi.org/10.1037/0022-3514.93.6.1040

Briñol, P., Petty, R. E., & Wagner, B. (2009). Body posture effects on self-evaluation: A self-validation approach. *European Journal of Social Psychology*, *39*(6), 1053-1064. http://dx.doi.org/10.1002/ejsp.607

Buhrmester, M. D., Fraser, W. T., Lanman, J. A., Whitehouse, H., & Swann Jr, W. B. (2015). When terror hits home: Identity fused Americans who saw Boston bombing victims as "family" provided aid. *Self and Identity*, *14*(3), 253-270. https://doi.org/10.1080/15298868.2014.992465

Buhrmester, M. D., Gómez, Á., Brooks, M. L., Morales, J. F., Fernández, S., & Swann Jr, W. B. (2012). My group's fate is my fate: Identity-fused Americans and Spaniards link personal life quality to outcome of 2008 elections. *Basic and Applied Social Psychology*, *34*(6), 527-533. https://doi.org/10.1080/01973533.2012.732825

Cason, T. N., & Mui, V. L. (1998). Social influence in the sequential dictator game. *Journal of mathematical psychology*, *42*(2-3), 248-265. https://doi.org/10.1006/jmps.1998.1213

Cesario, J., Grant, H., & Higgins, E. T. (2004). Regulatory Fit and Persuasion: Transfer From "Feeling Right." *Journal of Personality and Social Psychology*, *86*(3), 388-404. http://dx.doi.org/10.1037/0022-3514.86.3.388

Chaiken, S., Liberman, A., & Eagly, A. H. (1989). Heuristic and systematic processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212-252). New York: Guilford Press.

Chen, S., Langner, C. A., & Mendoza-Denton, R. (2009). When dispositional and role power fit: Implications for self-expression and self–other congruence. *Journal of Personality and Social Psychology, 96*(3), 710-727. http://dx.doi.org/10.1037/a0014526

Clark, J. K., & Thiem, K. C. (2015). Group communicators, perceived entitativity, and persuasion: A self-validation analysis. *Journal of Experimental Social Psychology*, *61*, 5-11. https://doi.org/10.1016/j.jesp.2015.06.005

Clark, J. K., Wegener, D. T., Briñol, P., & Petty, R. E. (2009). Discovering that the shoe fits: The self-validating role of stereotypes. *Psychological Science, 20*, 846-852. https://doi.org/10.1111/j.1467-9280.2009.02375.x

Clark, J. K., Wegener, D. T., Sawicki, V., Petty, R. E., & Briñol, P. (2013). Evaluating the message or the messenger? Implications for self-validation in persuasion. *Personality and Social Psychology Bulletin*, *39*(12), 1571-1584. https://doi.org/10.1177/0146167213499238

Dang, J., Liu, L., Ren, D., & Gu, Z. (2017). "Groupy" Allies Are More Beneficial While "Groupy" Enemies Are More Harmful. *Social Psychological and Personality Science*, 1948550617729409. https://doi.org/10.1177/1948550617729409

Darley, J. M., & Latane, B. (1968). Bystander intervention in emergencies: diffusion of responsibility. *Journal of personality and social psychology*, *8*(4), 377-383. http://dx.doi.org/10.1037/h0025589

DeMarree, K. G., Clark, C. J., Wheeler, S. C., Briñol, P., & Petty, R. E. (2017). On the pursuit of desired attitudes: Wanting a different attitude affects information processing and

behavior. *Journal of Experimental Social Psychology*, *70*, 129-142. https://doi.org/10.1016/j.jesp.2017.01.003

DeBono, K. G. (1987). Investigating the social-adjustive and value-expressive functions of attitudes: Implications for persuasion processes. *Journal of Personality and Social Psychology*, *52*(2), 279-287. http://dx.doi.org/10.1037/0022-3514.52.2.279

DeBono, K. G., & Harnish, R. J. (1988). Source expertise, source attractiveness, and the processing of persuasive information: A functional approach. *Journal of Personality and Social Psychology*, *55*(4), 541-546. http://dx.doi.org/10.1037/0022-3514.55.4.541

DeBono, K. G., & Snyder, A. (1992). Repressors, sensitizers, source expertise, and persuasion. *Social Behavior and Personality: an international journal*, *20*(4), 263-272. https://doi.org/10.2224/sbp.1992.20.4.263

Durso, G. R., Briñol, P., & Petty, R. E. (2016). From Power to Inaction: Ambivalence Gives Pause to the Powerful. *Psychological science*, *27*(12), 1660-1666. https://doi.org/10.1177/0956797616669947

Evans, A. T., & Clark, J. K. (2012). Source characteristics and persuasion: The role of self-monitoring in self-validation. *Journal of Experimental Social Psychology*, *48*(1), 383-386. https://doi.org/10.1016/j.jesp.2011.07.002

Evans, L. M., & Petty, R. E. (2003). Self-guide framing and persuasion: Responsibly increasing message processing to ideal levels. *Personality and Social Psychology Bulletin*, *29*(3), 313-324. https://doi.org/10.1177/0146167202250090

Fazio, R. H., & Zanna, M. P. (1981). Direct experience and attitude-behavior consistency. *Advances in experimental social psychology*, *14*, 161-202. https://doi.org/10.1016/S0065-2601(08)60372-X

Feinberg, M., & Willer, R. (2015). From gulf to bridge: When do moral arguments facilitate political influence? *Personality and Social Psychology Bulletin*, *41(12)*, 1665-1681. https://doi.org/10.1177/0146167215607842

Fleming, M. A., & Petty, R. E. (2000). *Identity and persuasion: An elaboration likelihood approach*. Lawrence Erlbaum Associates Publishers.

Gawronski, B., Armstrong, J., Conway, P., Friesdorf, R., & Hütter, M. (2017). Consequences, norms, and generalized inaction in moral dilemmas: The CNI model of moral decision-making. *Journal of Personality and Social Psychology*, *113*(3), 343-376. http://dx.doi.org/10.1037/pspa0000086.

Gómez, A., Brooks, M. L., Buhrmester, M. D., Vázquez, A., Jetten, J., & Swann Jr, W. B. (2011). On the nature of identity fusion: insights into the construct and a new measure. *Journal of Personality and Social Psychology*, *100(5)*, 918-933. http://dx.doi.org/10.1037/a0022642

Gómez, A., Morales, J.F., Hart, S, Vázquez, A. & Swann, W.B. Jr. (2011). Rejected and excluded forevermore, but even more devoted: Irrevocable ostracism intensifies loyalty to the group among identity fused persons. *Personality and Social Psychology Bulletin*, *37(12)*, 1574-1586. https://doi.org/10.1177/0146167211424580

Gómez, A., & Vázquez, A. (2015). The power of 'feeling one' with a group: identity fusion and extreme pro-group behaviors. *International Journal of Social Psychology*, *30(3)*, 481-511. http://dx.doi.org/10.1080/02134748.2015.1065089

Haddock, G., Maio, G. R., Arnold, K., & Huskinson, T. (2008). Should persuasion be affective or cognitive? The moderating effects of need for affect and need for cognition. *Personality and Social Psychology Bulletin*, *34*(6), 769-778. https://doi.org/10.1177/0146167208314871

Haggard, P., & Tsakiris, M. (2009). The experience of agency: Feelings, judgments, and responsibility. *Current Directions in Psychological Science*, *18*(4), 242-246. https://doi.org/10.1111/j.1467-8721.2009.01644.x

Hayes, A. F. (2016). PROCESS macro for SPSS and SAS (Version 2.16) [Computer software].

Hayes, A. F. (2017). Partial, conditional, and moderated mediation: Quantification, inference, and interpretation. *Communication Monographs*, 1-37. http://dx.doi.org/10.1080/03637751.2017.1352100

Hayes, A. F., & Montoya, A. K. (2017). A tutorial on testing, visualizing, and probing an interaction involving a multicategorical variable in linear regression analysis. *Communication Methods and Measures*, *11*(1), 1-30. http://dx.doi.org/10.1080/19312458.2016.1271116

Hermann, A. D., Leonardelli, G. J., & Arkin, R. M. (2002). Self-doubt and self-esteem: A threat from within. *Personality and Social Psychology Bulletin*, *28*(3), 395-408. https://doi.org/10.1177/0146167202286010

Hirsh, J. B., Kang, S. K., & Bodenhausen, G. V. (2012). Personalized persuasion: Tailoring persuasive appeals to recipients' personality traits. *Psychological science*, *23*(6), 578-581. https://doi.org/10.1177/0956797611436349

Huntsinger, J. R. (2013). Anger enhances correspondence between implicit and explicit attitudes. *Emotion*, *13*(2), 350. http://dx.doi.org/10.1037/a0029974

Huntsinger, J. R. (2013b). Incidental experiences of affective coherence and incoherence influence persuasion. *Personality and Social Psychology Bulletin*, *39*(6), 792-802. https://doi.org/10.1177/0146167213482588

Jiménez, J., Gomez, A., Buhrmester, M. D., Vázquez, A., Whitehouse, H., & Swann, W. B. (2015). The dynamic identity fusion index: A new continuous measure of identity fusion for web-

based questionnaires. *Social Science Computer Review, 34(2),* 215-228. https://doi.org/10.1177/0894439314566178

Johnson, M. K., Rowatt, W. C., & LaBouff, J. P. (2012). Religiosity and prejudice revisited: In-group favoritism, out-group derogation, or both? *Psychology of Religion and Spirituality, 4*(2), 154-168. http://dx.doi.org/10.1037/a0025107

Jong, J., Whitehouse, H., Kavanagh, C., & Lane, J. (2015). Shared negative experiences lead to identity fusion via personal reflection. *PloS one, 10*(12), e0145611. https://doi.org/10.1371/journal.pone.0145611

Joo, M., & Park, S. W. (2017). Effect of identity fusion on decision to make extreme sacrifices in romantic relationships: The moderating role of impulsiveness. *British Journal of Social Psychology*, *56*(4), 819-827. http://dx.doi.org/10.1111/bjso.12218

Karmarkar, U. R., & Tormala, Z. L. (2009). Believe me, I have no idea what I'm talking about: The effects of source certainty on consumer involvement and persuasion. *Journal of Consumer Research*, *36*(6), 1033-1049. https://doi.org/10.1086/648381

Kruglanski, A. W., & Thompson, E. P. (1999). Persuasion by a single route: A view from the unimodel. *Psychological Inquiry*, *10*(2), 83-109. https://doi.org/10.1207/S15327965PL100201

Killeya, L. A., & Johnson, B. T. (1998). Experimental induction of biased systematic processing: The directed-thought technique. *Personality and Social Psychology Bulletin*, *24*(1), 17-33. https://doi.org/10.1177/0146167298241002

Lavine, H., & Snyder, M. (1996). Cognitive processing and the functional matching effect in persuasion: The mediating role of subjective perceptions of message quality. *Journal of Experimental Social Psychology*, *32*(6), 580-604. https://doi.org/10.1006/jesp.1996.0026

Lee, A. Y., & Aaker, J. L. (2004). Bringing the Frame Into Focus: The Influence of Regulatory Fit on Processing Fluency and Persuasion. *Journal of Personality and Social Psychology*, *86*(2), 205-218. http://dx.doi.org/10.1037/0022-3514.86.2.205

Lee, A. Y., Aaker, J. L., & Gardner, W. L. (2000). The pleasures and pains of distinct self-construals: the role of interdependence in regulatory focus. *Journal of personality and social psychology*, *78*(6), 1122-1134. http://dx.doi.org/10.1037/0022-3514.78.6.1122

Lepper, M. R., Greene, D., & Nisbett, R. E. (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the" overjustification" hypothesis. *Journal of Personality and social Psychology*, *28*(1), 129-137. http://dx.doi.org/10.1037/h0035519

Loewenstein, G. F., & Prelec, D. (1993). Preferences for sequences of outcomes. *Psychological review*, *100*(1), 91-108. http://dx.doi.org/10.1037/0033-295X.100.1.91

Ludeke, S. G., & Rasmussen, S. H. R. (2016). Personality correlates of sociopolitical attitudes in the Big Five and Eysenckian models. *Personality and Individual Differences*, *98*, 30-36. https://doi.org/10.1016/j.paid.2016.03.079

Mael, F., & Ashforth, B. E. (1992). Alumni and their alma mater: A partial test of the reformulated model of organizational identification. *Journal of organizational Behavior*, *13*(2), 103-123. https://doi.org/10.1002/job.4030130202

Maio, G. R., & Haddock, G. (2007). Attitude change. *Social psychology: Handbook of basic principles*, *2*, 565-586.

Maio, G. R., & Haddock, G. (2015). *The psychology of attitudes and attitude change* (2nd ed.). London, England: Sage.

McCann, S. J. (2011). Personality and American state differences in obesity prevalence. *The Journal of psychology*, *145*(5), 419-433. https://doi.org/10.1080/00223980.2011.584081

McConnell, A. R., & Leibold, J. M. (2001). Relations among the Implicit Association Test, discriminatory behavior, and explicit measures of racial attitudes. *Journal of experimental Social psychology*, *37*(5), 435-442. https://doi.org/10.1006/jesp.2000.1470

Metcalfe, J., & Mischel, W. (1999). A hot/cool-system analysis of delay of gratification: dynamics of willpower. *Psychological review*, *106*(1), 3-19. http://dx.doi.org/10.1037/0033-295X.106.1.3

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. *Science*, *244*(4907), 933-938.

Muller, D., Judd, C. M., & Yzerbyt, V. Y. (2005). When moderation is mediated and mediation is moderated. *Journal of personality and social psychology*, *89*(6), 852. http://dx.doi.org/10.1037/0022-3514.89.6.852

Nosek, B. A., Banaji, M. R., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice*, *6*(1), 101-115. http://dx.doi.org/10.1037/1089-2699.6.1.101

Pacini, R., & Epstein, S. (1999). The relation of rational and experiential information processing styles to personality, basic beliefs, and the ratio-bias phenomenon. *Journal of personality and social psychology*, *76*(6), 972. http://dx.doi.org/10.1037/0022-3514.76.6.972

Palmgreen, P., Stephenson, M. T., Everett, M. W., Baseheart, J. R., & Francies, R. (2002). Perceived message sensation value (PMSV) and the dimensions and validation of a PMSV scale. *Health Communication*, *14*(4), 403-428. https://doi.org/10.1207/S15327027HC1404_1

Paredes, B., Stavraki, M., Briñol, P., & Petty, R. E. (2013). Smiling after thinking increases reliance on thoughts. *Social Psychology*. 44. 349-353. https://doi.org/10.1027/1864-9335/a000131

Paredes, B., Stavraki, M., Díaz, D., Gandarillas, B., & Briñol, P. (2015). Validity and reliability of the Spanish version of the Revised Self-Monitoring Scale. *Spanish Journal of Psychology*, *18*, 1-11. https://doi.org/10.1017/sjp.2015.64

Petty, R. E., Briñol, P., & Tormala, Z. L. (2002). Thought confidence as a determinant of persuasion: The self-validation hypothesis. *Journal of Personality and Social Psychology, 82*(5), 722-741. http://dx.doi.org/10.1037/0022-3514.82.5.722

Petty, R. E. & Briñol, P. (2012). The Elaboration Likelihood Model. In P. A. M. Van Lange, A. Kruglanski, & E. T. Higgins (Eds.), *Handbook of theories of social psychology* (Vol.1, pp. 224-245). London, England: Sage.

Petty. R. E & Briñol, P. (2014). The elaboration likelihood and meta-cognitive models of attitudes: Implications for prejudice, the self, and beyond. In J. Sherman, B. Gawronski, & Y. Trope (Eds.), *Dual-process theories of the social mind* (pp. 172-187). New York, NY: Guildford Press.

Petty, R. E., & Briñol, P. (2015). Emotion and persuasion: Cognitive and meta-cognitive processes impact attitudes. *Cognition and Emotion*, *29*(1), 1-26.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in experimental social psychology*, *19*, 123-205. https://doi.org/10.1016/S0065-2601(08)60214-2

Petty, R. E., & Wegener, D. T. (1998). Matching versus mismatching attitude functions: Implications for scrutiny of persuasive messages. *Personality and Social Psychology Bulletin*, *24*(3), 227-240. https://doi.org/10.1177/0146167298243001

Petty, R. E., Wheeler, S. C., & Bizer, G. Y. (1999). Is there one persuasion process or more? Lumping versus splitting in attitude change theories. *Psychological Inquiry*, *10*(2), 156-163. http://dx.doi.org/10.1207/S15327965PL100211

Petty, R. E., Wheeler, S.C., & Bizer, G. (2000). Matching effects in persuasion: An elaboration likelihood analysis (pp. 133-162). In G. Maio & J. Olson (Eds.), *Why we evaluate: Functions of attitudes*. Mahwah, NJ: Erlbaum

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior research methods*, *36*(4), 717-731. https://doi.org/10.3758/BF03206553

Prochaska, J. O., DiClemente, C. C., & Norcross, J. C. (1992). In search of how people change: Applications to addictive behaviors. *American psychologist*, *47*(9), 1102-1114. http://dx.doi.org/10.1037/0003-066X.47.9.1102

Rosenblatt, A., Greenberg, J., Solomon, S., Pyszczynski, T., & Lyon, D. (1989). Evidence for terror management theory: I. The effects of mortality salience on reactions to those who violate or uphold cultural values. *Journal of Personality and Social Psychology, 57*(4), 681-690. http://dx.doi.org/10.1037/0022-3514.57.4.681

Rucker, D. D., Tormala, Z. L., Petty, R. E., & Briñol Turnes, P. (2013). Consumer conviction and commitment: An appraisal-based framework for attitude certainty. *Journal of Consumer Psychology 24(1),* 119-136. https://doi.org/10.1016/j.jcps.2013.07.001

Salovey, P., & Wegener, D. T. (2003). Communicating about health: Message framing, persuasion, and health behavior. *Social psychological foundations of health and illness*, 54-81.

Schubert, T. W., & Otten, S. (2002). Overlap of self, ingroup, and outgroup: Pictorial measures of self-categorization. *Self and identity*, *1*(4), 353-376. https://doi.org/10.1080/152988602760328012

See, Y. H. M., Petty, R. E., & Evans, L. M. (2009). The impact of perceived message complexity and need for cognition on information processing and attitudes. *Journal of Research in Personality*, *43*(5), 880-889. https://doi.org/10.1016/j.jrp.2009.04.006

Sheikh, H., Gómez, Á., & Atran, S. (2016). Empirical evidence for the devoted actor model. *Current Anthropology*, *57*(S13), S204-S209. https://doi.org/10.1086/686221

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological methods*, *7*(4), 422-445.

Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, *30*(4), 526-537. http://dx.doi.org/10.1037/h0037039

Snyder, M. (1979). Self-monitoring processes. *Advances in experimental social psychology*, *12*, 85-128. https://doi.org/10.1016/S0065-2601(08)60260-9

Snyder, M., & DeBono, K. G. (1985). Appeals to image and claims about quality: Understanding the psychology of advertising. *Journal of personality and Social Psychology*, *49*(3), 586-597. http://dx.doi.org/10.1037/0022-3514.49.3.586

Snyder, C. R., & Fromkin, H. L. (1977). Abnormality as a positive characteristic: The development and validation of a scale measuring need for uniqueness. *Journal of Abnormal Psychology, 86(5*), 518–527. http://dx.doi.org/10.1037/0021-843X.86.5.518

Strathman, A., Gleicher, F., Boninger, D. S., & Edwards, C. S. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of personality and social psychology*, *66*(4), 742-752. http://dx.doi.org/10.1037/0022-3514.66.4.742

Swann, W. B., Jr., Buhrmester, M. (2015). Identity fusion. *Current Directions in Psychological Science. 24*, 52–57. https://doi.org/10.1177/0963721414551363

Swann Jr, W. B., Buhrmester, M. D., Gómez, A., Jetten, J., Bastian, B., Vázquez, A., ... & Zhang, A. (2014). What makes a group worth dying for? Identity fusion fosters perception of familial ties, promoting self-sacrifice. *Journal of Personality and Social Psychology, 106*, 912-926. http://dx.doi.org/10.1037/a0036089

Swann Jr, W. B., Gómez, Á., Buhrmester, M. D., López-Rodríguez, L., Jiménez, J., & Vázquez, A. (2014). Contemplating the ultimate sacrifice: Identity fusion channels pro-group affect,

cognition, and moral decision making. *Journal of Personality and Social Psychology, 106*, 713-727. http://dx.doi.org/10.1037/a0035809

Swann Jr, W. B., Gómez, Á., Dovidio, J. F., Hart, S., & Jetten, J. (2010). Dying and killing for one's group: Identity fusion moderates responses to intergroup versions of the trolley problem. *Psychological Science, 21*(8), 1176-1183. https://doi.org/10.1177/0956797610376656

Swann Jr, W. B., Gómez, A., Huici, C., Morales, J., & Hixon, J. G. (2010). Identity fusion and self-sacrifice: arousal as a catalyst of pro-group fighting, dying, and helping behavior. *Journal of Personality and Social Psychology*, *99*, 824-841. http://dx.doi.org/10.1037/a0020014

Swann Jr, W. B., Gómez, A., Seyle, D. C., Morales, J., & Huici, C. (2009). Identity fusion: the interplay of personal and social identities in extreme group behavior. *Journal of Personality and Social Psychology*, 96, 995-1011. http://dx.doi.org/10.1037/a0013668

Swann Jr, W. B., Gómez, Á., Vázquez, A., Guillamón, A., Segovia, S., & Carrillo, B. (2015). Fusion with the cross-gender group predicts genital sex reassignment surgery. *Archives of Sexual Behavior, 44*, 1313-1318. http://dx.doi.org/10.1007/s10508-015-0526-0

Swann Jr, W. B., Jetten, J., Gómez, A., Whitehouse, H., & Bastian, B. (2012). When group membership gets personal: A theory of identity fusion. *Psychological Review, 119*, 441-456. http://dx.doi.org/10.1037/a0028589

Tajfel, H. (1982). Social psychology of intergroup relations. *Annual review of psychology*, *33*(1), 1-39. https://doi.org/10.1146/annurev.ps.33.020182.000245

Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worchel (Eds.). *The social psychology of intergroup relations* (pp. 33– 47). Monterey, CA: Brooks-Cole.

Teeny, J., Briñol, P., & Petty, R. E. (2017). The Elaboration Likelihood Model: Understanding consumer attitude change. In C. Jansson-Boyd & M. Zawisza (Eds.), *Routledge international handbook of consumer psychology* (pp. 390-410). Abingdon, UK: Routledge.

Tiedens, L. Z., & Linton, S. (2001). Judgment under emotional certainty and uncertainty: The effects of specific emotions on information processing. *Journal of Personality and Social Psychology, 81*(6), 973-988. http://dx.doi.org/10.1037/0022-3514.81.6.973

Tykocinskl, O., Higgins, E. T., & Chaiken, S. (1994). Message framing, self-discrepancies, and yielding to persuasive messages: The motivational significance of psychological situations. *Personality and Social Psychology Bulletin*, *20*(1), 107-115. https://doi.org/10.1177/0146167294201011

Vázquez, A., Gómez, Á., Ordoñana, J. R., & Paredes, B. (2015). From interpersonal to extended fusion: Relationships between fusion with siblings and fusion with the country/Desde la fusión interpersonal a la fusión extendida: Relaciones entre la fusión con los hermanos y la fusión con el país. *Revista de Psicología Social*, *30*(3), 512-530. http://dx.doi.org/10.1080/02134748.2015.109375

Valdesolo, P., & DeSteno, D. (2006). Manipulations of emotional context shape moral judgment. *Psychological science*, *17*(6), 476-477. https://doi.org/10.1111/j.1467-9280.2006.01731.x

Vázquez, A., Gómez, Á., Ordoñana, J. R., & Paredes, B. (2015). From interpersonal to extended fusion: relationships between fusion with siblings and fusion with the country/Desde la fusión interpersonal a la fusión extendida: Relaciones entre la fusión con los hermanos y la fusión con el país. *Revista de Psicología Social*, *30*(3), 512-530. http://dx.doi.org/10.1080/02134748.2015.1093755

Vázquez, A., Gómez, Á., Ordoñana, J. R., Swann, W. B., & Whitehouse, H. (2017). Sharing genes fosters identity fusion and altruism. *Self and Identity*, 684-702. http://dx.doi.org/10.1080/15298868.2017.1296887

Vázquez, A., Gómez, Á., & Swann, W. B. (2017). Do historic threats to the group diminish identity fusion and its correlates? *Self and Identity, 16(4),* 480-503. http://dx.doi.org/10.1080/15298868.2016.1272485

Waldzus, S., & Mummendey, A. (2004). Inclusion in a superordinate category, in-group prototypicality, and attitudes towards out-groups. *Journal of Experimental social psychology*, *40*(4), 466-477. https://doi.org/10.1016/j.jesp.2003.09.003

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naive theories of bias in bias correction. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol., 29, pp. 141-208). San Diego: Academic Press. https://doi.org/10.1016/S0065-2601(08)60017-9

Wheeler, S. C., Petty, R. E., & Bizer, G. Y. (2005). Self-schema matching and attitude change: Situational and dispositional determinants of message elaboration. *Journal of Consumer Research, 31(4),* 787-797. https://doi.org/10.1086/426613

Whitehouse, H., Jong, J., Buhrmester, M. D., Gómez, Á., Bastian, B., Kavanagh, C. M., ... & Gavrilets, S. (2017). The evolution of extreme cooperation via shared dysphoric experiences. *Scientific Reports*, 7, srep44292. http://dx.doi.org/10.1038/srep44292

Whitehouse, H., McQuinn, B., Buhrmester, M., & Swann, W. B. (2014). Brothers in Arms: Libyan revolutionaries bond like family. *Proceedings of the National Academy of Sciences*, *111*(50), 17783-17785. http://dx.doi.org/10.1073/pnas.1416284111

Williams-Piehota, P., Schneider, T. R., Pizarro, J., Mowad, L., & Salovey, P. (2004). Matching health messages to health locus of control beliefs for promoting mammography

utilization. *Psychology & Health*, *19*(4), 407-423. https://doi.org/10.1080/08870440310001652678

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116(1),* 117-142. http://dx.doi.org/10.1037/0033-2909.116.1.117