

When Trust Goes Wrong: A Social Identity Model of Risk Taking

Tegan Cruwys^{1*+}, Katharine H. Greenaway²⁺, Laura J. Ferris³, Joanne A. Rathbone³,
Alexander K. Saeri⁴, Elyse Williams³, Stacey L. Parker³, Melissa X-L. Chang³, Nicholas
Croft³, William Bingley³, Laura Grace³

1. Research School of Psychology, The Australian National University
2. Melbourne School of Psychological Sciences, University of Melbourne
3. School of Psychology, The University of Queensland
4. BehaviourWorks Australia, Monash Sustainable Development Institute,
Monash University

* The corresponding author is Dr Tegan Cruwys: tegan.cruwys@anu.edu.au, Research School of Psychology, Australian National University, Canberra ACT 2601 AUSTRALIA

+ These authors contributed equally.

This preprint has been accepted for publication 8 March 2020 as:

Cruwys, T., Greenaway, K. H., Ferris, L. J., Rathbone, J. A., Saeri, A. K., Williams, E., Parker, S. L., Chang, M. X-L., Croft, N., Bingley, W., & Grace, L. (2020). When trust is a liability: A social identity model of risk taking. *Journal of Personality and Social Psychology*. 10.1037/pspi0000243

© 2020, American Psychological Association. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/pspi0000243

Author Note

The first two authors are supported by Australian Research Council Discovery Early Career Researcher Awards (DE160100592 & DE160100761). Studies 4 and 13 were also supported by funding from the Queensland Government Department of Child Safety, Youth, and Women. The authors would like to thank Kirsten Axford who collected data for Study 9, Jason Dean and Sophie Gates who assisted with data collection and coding in Study 10, Craig Rice who provided technical expertise for Studies 2 and 8, as well as Eric Vanman and Daniel Skorich for their valuable insights on the design of Study 6. We thank the organisers of the Dark Mofo festival, especially Kate Gould, Miriam McGarry, and Natasha Wardale, for their support for Study 5. Data collection for Studies 4 and 13 was facilitated by Taylor Alati, Sarah Blessing, Joey Brautigam, Thomas Curro, Samuel Fouhy, Petra Harman-Schufft, Sienna Hinton, Yasmin Larasati, Jenna Laroque, Harrison Lee, Sarah Macinnes, Thomas McCarthy, Joseph Martin, Daniel Raab, Younji Ryu, Joshua Santin, Martina Vuckavac, Zoe Weller, Nick Wheeler, Olivia Zeckovic, Lara Ollis, Bhargav Dave, Brittany Chapman, Cassandra Alford, Gabrielle Chatwood, Jessica Gralton, Pratima Kavi, & Yanzhuo Yu.

Abstract

Risk taking is typically viewed through a lens of individual deficits (e.g., impulsivity) or normative influence (e.g., peer pressure). An unexplored possibility is that shared group membership, and the trust that flows from it, may play a role in reducing risk perceptions and promoting risky behavior. We propose and test a Social Identity Model of Risk Taking in eight studies (total $N = 4,708$) that employ multiple methods including minimal group paradigms, correlational, longitudinal, and experimental designs to investigate the effect of shared social identity across diverse risk contexts. Studies 1 and 2 provided evidence for the basic premise of the model, showing that ingroup members were perceived as posing lower risk and inspired greater risk taking behavior than outgroup members. Study 3 found that social identification was a moderator, such that effect of shared group membership was strongest among high identifiers. Studies 4 and 5 among festival attendees showed correlational and longitudinal evidence for the model and further that risk-taking was mediated by trust, not disgust. Study 6 manipulated the mediator and found that untrustworthy faces were trusted more and perceived as less risky when they were ingroup compared to outgroup members. Studies 7 and 8 identified integrity as the subcomponent of trust that consistently promotes greater risk taking in the presence of ingroup members. The findings reveal that a potent source of risk discounting is the group memberships we share with others. Ironically, this means the people we trust the most may sometimes pose the greatest risk.

Keywords: *social identity; risk; trust; disgust; group membership*

When Trust Goes Wrong: A Social Identity Model of Risk Taking

What do unprotected sex, investing in a doomed business venture, or falling for an online dating scam have in common? They have all been characterized as examples of risk taking. As such, these actions have the capacity to compromise safety, personal finances, and even public health (e.g., Tam et al., 2012). The present research introduces a new way of looking at risk, where people do not take risks with just anyone, but often with the very people they trust the most. This research sheds light on the mechanisms through which shared group membership impacts on risk perceptions and risk behavior. We argue that to understand risk taking, we must understand how and why people come to place trust in one another.

Understanding Risk from a Social Perspective

We live in a world of risk. Although global threats such as nuclear war or climate change are ever-present (Stover, 2013), risks in health or social domains hold the potential for harm on a daily basis. Rapid environmental and technological changes are amplifying risk in new domains, including identity fraud, computer hacking, financial theft, drug addiction, and superbug infection (Cole, 2016; Décary-Héту, Paquet-Clouston, & Aldridge, 2016; Reid, 2018; Valenstein et al., 2011). Certain times of life—such as adolescence—and certain people—such as men—are at greater chronic risk of (often self-inflicted) harm to health and safety (Charness & Gneezy, 2012; Gullone & Moore, 2000). Disadvantaged or marginalized social groups such as minorities, immigrants, and refugees face regular risks of physical harm, as well as well-being threats in the form of social disconnection and discrimination (Holt-Lunstad, Smith, Baker, Harris, & Stephenson, 2015; Paradies et al., 2015). Considering the significant (indeed, sometimes life and death) implications of living in a risk society (Beck, 1992), psychological research has sought to document, understand, and change people's perceptions of risk, as well as actual risk behavior.

At the heart of risk is probability—probability that one’s actions will result in harm or not (Leigh, 1999). Indeed, risk taking is often characterized by behaviors that carry potential for benefit as well as cost (Lejuez et al., 2002). Sometimes the potential for cost and reward are material (e.g., financial investment) and sometimes risk behavior involves potential cost that is physical and reward that is psychological (e.g., sky diving). In addition to risk behavior, risk is often assessed subjectively as perceived vulnerability to harm (Janz & Becker, 1984), with the implication that lower perceived risk is inversely related to the likelihood of engaging in risky behavior. Therefore, in this investigation we operationalize risk taking both in terms of relatively lower risk perception and greater risk behavior. We seek to explore the psychological factors that shape these risk outcomes.

Several existing psychological models adopt an *individual* focus and identify personal characteristics as drivers of risk taking (Lauriola, Panno, Levin, & Lejuez, 2014; Nicholson, Soane, Fenton-O’Creevy, & Willman, 2005). These include traits like impulsiveness and sensation seeking (Dawe & Loxton, 2004; Zuckerman & Kuhlman, 2000), and cognitive limits on humans’ ability to calculate risk (Botdorf, Rosenbaum, Patrianakos, Steinberg, & Chein, 2017; Kahneman & Lovallo, 1993), as well as physiological markers such as higher circulating testosterone (Kurath & Mata, 2018; Ronay & von Hippel, 2010).

Another line of research considers the social context in which risk occurs. Early work on the social dimensions of risk taking tended to caricature social influences as purely harmful, leading to ‘groupthink’ that encouraged risky decision making (Janis, 1972; Myers & Lamm, 1976), or ‘peer pressure’ to engage in risky activities (Killen, 1985; Mitchell & West, 1996). A social dimension to risk taking was also given primacy in research on hazing and initiation rituals (e.g., Ramzy & Bryant, 1962). While the ‘group’ in such work might be construed as broadly as a culture, or as narrowly as a fraternity, such research shares an assumption that social factors have a negative influence (Raalte, Cornelius, Linder, &

Brewer, 2007). However, in the intervening decades, a more nuanced understanding of social influences has developed. For instance, research shows that people process and understand information differently depending on the identity of the person communicating the information (e.g., Abrams, Wetherell, Cochrane, Hogg, & Turner, 1990; Greenaway, Wright, Willingham, Reynolds, & Haslam, 2014; Platow et al., 2007). In this paper, we apply this understanding borne of decades of research in social psychology to delve more deeply into understanding risk perception and behavior.

One theoretical perspective for understanding how social forces impact on individual psychology and behavior is the social identity approach, which combines insights drawn from social identity theory (Tajfel, Billig, Bundy, & Flament, 1971; Tajfel & Turner, 1979) and self-categorization theory (Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). Gestalt cognitive theories posited that a collection of objects are more likely to be perceived as a collective entity if they are similar and move together (e.g., Campbell, 1958; see also Lewin, 1948). The social identity approach extended these ideas to *social* objects (i.e., people, including the self; Turner et al., 1987). Self-categorizing as a member of a particular social group fundamentally shapes a person's experience of the world (for a review, see Hornsey, 2008). The psychological process of subjectively perceiving oneself in terms of a particular social group membership (i.e., self-categorization in terms of a social identity) results in qualitative shifts in thoughts, emotions, behavior, and even perception (Turner & Oakes, 1997). For instance, ingroup members are seen to be more similar to the self (van Rijswijk, Haslam, & Ellemers, 2006), to have more positive personality traits (Reynolds & Oakes, 2000), and to hold more valid opinions (McGarty, Haslam, Hutchison, & Turner, 1994) than outgroup members.

While it is well established that social identities influence how individuals appraise and act on information in general, what has not received systematic attention to date is

whether social identities influence risk perception and behavior in particular. It is clear that risk is subjectively experienced (Zinn, 2017). That is, two situations with comparable probabilities of negative outcomes may be highly divergent in the degree to which they are perceived as dangerous. Thus, the question that has occupied researchers is what causes people to *perceive* situations as more or less risky and thus engage in more or less risk behavior. We argue that risk is inherently social—so that *who* we take risks with is vitally important to understanding *why* we take risks. Specifically, we propose that shared group membership may be a psychological proxy for ‘safe’ (i.e., without risk). Indeed, although this specific proposition has not been tested in previous research, we might interpret some extant findings as evidence for shared group membership being used as a kind of ‘safety heuristic’, with ingroup members perceived to be inherently less risky than outgroup members.

The most compelling evidence to date for a link between shared group membership and risk comes from Loersch and Bartholow (2011). In three experiments, these authors found that when participants were presented with beer cans that featured colors of their university, beer consumption was perceived to be less dangerous than when presented with a standard can. Using similar reasoning, Firing and Laberg (2012) investigated the predictors of jumping into the freezing ocean among 128 military and police officers. Highlighting the importance of social context, shared identity with fellow officers was found to be the strongest predictor of taking this risk, explaining more variance than a variety of individual difference factors. There is also indirect evidence of the link between social identity and risk from applied field studies of mass gatherings. Several studies have found that, despite the well-documented public health risks of such gatherings (Khan et al., 2013; Tam et al., 2012), people feel *safer* at these events when they identify with others in the crowd (Alnabulsi & Drury, 2014; Hopkins & Reicher, 2016; Shankar et al., 2013).

While these studies provide promising evidence supporting the relationship between shared identity and risk taking, they are not able to rule out an alternative explanation: specifically, conformity to an unmeasured ingroup norm in favor of risky activities. In particular, both university students and military personnel may perceive their own group to be risk takers, and indeed the experimental paradigms used here may have the confounding effect of making salient an existing risk taking norm within the ingroup. Enhanced normative influence from ingroup members is a useful framework to explain behavior, including risk taking (e.g., Cruwys et al., 2012; Dingle, Stark, Cruwys, & Best, 2015; Louis, Davies, Smith, & Terry, 2007). However, our model proposes that the direct effect of shared identity on risk may occur in addition to, and independent of, normative influence effects.

In the present research, we go beyond previous work by conducting a rigorous test of the relationship between shared social identity and risk. We hypothesize that people who share group membership with others in risky situations will perceive less risk and engage in more risky behavior. Our experimental designs are such that a normative explanation cannot account for the findings (especially in Studies 1, 3, 4, and 6). Moreover, we rule out a number of alternative explanations, such as general positivity toward ingroup (compared to outgroup) members (Study 7). In addition, we identify and test psychological mechanisms through which shared group membership promotes risk (Studies 3 to 8). Specifically, we hypothesize that shared group membership will increase *trust*—the perception that others can be relied upon (e.g., Rotter, 1971)—and reduce *disgust*—an emotion considered to underpin perceptions of risk (e.g., Loewenstein, Weber, Hsee, & Welch, 2001; Sparks, Fessler, Chan, Ashokkumar, & Holbrook, 2018).

The Costs and Benefits of Trust

We know from previous research that ingroup members are more likely to be trusted than outgroup members, even when they are strangers, because of an expectation of fair and

altruistic behavior (e.g., Brewer, 2008; Kramer, Brewer, & Hanna, 1996; Platow, Foddy, Yamagishi, Lim, & Chow, 2012). For instance, in a prisoner's dilemma game, people typically trust ingroup members to act in the group's best interest, even to the detriment of their personal best interest (Güth, Levati, & Ploner, 2008; Tanis & Postmes, 2005). Research on cultural evolution suggests that such a tendency is functional and has benefits for survival. We typically interact more with ingroup members than outgroup members, meaning we have greater experience with successful cooperation in these contexts (Boyd & Richardson, 2009; Henrich et al., 2011). Moreover, sharing cultural norms with fellow ingroup members ensures that people adhere to these norms, thus enhancing coordination within the group (Henrich & Boyd, 2001; McElreath, Boyd & Richardson, 2003). Thus, people have extensive experience of working effectively within ingroups—experiences that are not always afforded by outgroups—making ingroup members particularly desirable interaction partners who people expect to operate fairly, kindly, and competently.

Research to date has overwhelmingly focused on the benefits of ingroup trust, such as cooperation (Brewer, 2008), effective communication (Greenaway et al., 2015), and wellbeing (Cruwys et al., 2014). Being able to trust fellow ingroup members makes interactions fluent and provides the 'grease' that enables effective social functioning (Brewer, 2008; Cooper, 2008; Gilson, 2003; Helliwell & Wang, 2010). Beyond interpersonal exchanges, researchers have argued that such trust is necessary for the effective functioning of society, including the scientific community, the health system, and in emergency situations (Drury, Cocking, & Reicher, 2009; Igarashi et al., 2008; Levine et al., 2005; Ziersch et al., 2005).

However, on the other side of the ledger, we propose the novel extension that trust can have an important downside—leading people to take risks that they would not otherwise take. In economic research, it has been argued that trust always requires *some* risk (or else it

would simply be ‘assurance’; Das & Teng, 2004; Evans & Krueger, 2011). However, these concepts cannot be collapsed, since trust is a cognitive variable (Dietz & Den Hartog, 2006) while risk taking is a behavior. Building on this evidence, we propose that trust is a key mechanism through which shared group membership may attenuate perceptions of risk. For example, we may be more likely to trust that ingroup members will not have a contagious disease, or when they tell us that their investment scheme will succeed, leading to greater likelihood that we will take risks in these domains.

The concept of trust encompasses a willingness to be vulnerable, to allow another person to determine some outcome for oneself, and a belief that the outcome will be favorable (Mayer, Davis & Schoorman, 1995). Indeed, theoretical accounts identify three ‘pillars’ of trust: *ability* (the belief that an individual is capable of meeting their obligations and promises), *benevolence* (the belief that an individual is motivated to act in a target’s best interests), and *integrity* (the belief that an individual is reliable and honest; Mayer, Davis, & Schoorman, 1995). Although previous research identifying the links between shared identity and trust has not unpacked the specific *type* of trust that is engendered by shared group membership, it is possible that perceptions of all constructs may be higher for ingroup members than outgroup members. Ingroup members may be perceived to have greater ability, benevolence, *and* integrity and thus appear more trustworthy on each of these dimensions than outgroup members. Testing the precise mechanisms through which shared group membership predicts greater risk is therefore one goal of our research.

A Role for Disgust

Risk perception is not purely cognitive. Indeed, researchers have argued for a central role of emotion, suggesting that risk perception is better conceptualized as a feeling than as a cost-benefit evaluation (Loewenstein et al., 2001; Slovic & Peters, 2006). In this sense, the emotion of disgust has been particularly linked to risk. For instance, there is robust evidence

from clinical psychology that disgust sensitivity is elevated in several disorders, and plays an important causal role in risk avoidance among people with phobias, eating disorders, and obsessive-compulsive disorder (Olatunji, Ebesutani, Kim, Riemann, & Jacobi, 2017; Olatunji & Sawchuk, 2005; Vicario, Rafal, Martino, & Avenanti, 2017). However, most relevant to our model is a persuasive body of work on the behavioral immune system (Schaller, Murray, & Bangerter, 2015; Schaller & Park, 2011).

The behavioral-immune system refers to a set of behaviors that humans enact to reduce their risk of exposure to pathogens. The emotion of disgust is one of the key tools of the behavioral-immune system, motivating people to avoid stimuli that may be a source of disease. This idea has been supported empirically, with evidence suggesting people do indeed avoid others who might be contagious (Curtis, de Barra, & Aunger, 2011). Furthermore, humans have an elevated disgust response in circumstances when they are particularly vulnerable to contagious disease—for example, during the first trimester of pregnancy (Fessler, Eng, & Navarrete, 2005).

In classic signal detection theory terms (see Figure 1), these are examples of the detection of a *true positive*: a situation in which risk is present and correctly identified (and thus avoided). However, behavioral immune system research has particularly focused on situations in which risk is perceived when it is either absent or minimal (e.g., avoiding people who do not pose a pathogen risk, such as obese people or people from other racial groups; Faulkner, Schaller, Park, & Duncan, 2004; Miller & Maner, 2012; Park, Schaller, & Crandall, 2007). Indeed, a bias towards the overperception of disease cues is said to be a key mechanism of the behavioral immune system (called the “smoke detector principle”; Nesse, 2001; see also Johnston, Blumstein, Fowler & Haselton, 2013; Murray & Schaller, 2016). This represents a *false positive*: a situation in which risk is perceived but is not actually present. What has received comparatively less attention in the literature is the dangerous miss

(or *false negative*): a situation in which risk is present but not perceived as such. It is this under-theorized quadrant in which we locate the present investigation.

	<i>No risk perceived</i>	<i>Risk perceived</i>
<i>Risk present</i>	Face risk (false negative)	Avoid risk (true positive)
<i>Risk absent</i>	No risk (true negative)	No risk (false positive)

Figure 1. Schematic representation of risk detection. *Note.* The false positive quadrant has been the focus of behavioral immune system research. The shaded false negative quadrant is the focus of the new social identity approach to risk.

We propose that such false negatives are particularly likely in the context of salient shared group membership, which may attenuate disgust reactions. Recent evidence for this proposition comes from Reicher and colleagues (Reicher, Templeton, Neville, Ferrari, & Drury 2016; see also Khazaie & Khan, 2019). In two studies, these authors found that participants felt less disgust towards a damp sports-shirt that smelled of perspiration when it bore a logo of their own university, compared to when it bore the logo of a rival university. Furthermore, reduced disgust for ingroup members had behavioral consequences, with participants walking significantly slower to a sink to wash their hands and using less soap after touching the shirt when it had an ingroup logo. Building on this evidence, we propose that disgust is another mechanism through which shared group membership may affect risk.

The Present Research

In this work, we introduce and test a Social Identity Model of Risk Taking. Risk taking is a broad construct that has been conceptualized in many different ways. In this paper,

we aimed to include a mix of risk outcomes, ranging from subjective ratings of disease risk (e.g., Studies 1, 3 and 7) to physical and financial risk behavior (e.g., Studies 2, 5, and 8). In general, we aimed to use methods that could test the impact of (mere) group membership and that would not be influenced by other considerations such as reciprocal behavior and (normative) communication about risk. Therefore, the primary research question in this paper is: independent of contextual cues about actual risk, to what extent do people take risks based simply on sharing group membership with other people?

Our model predicts that shared group membership will have an impact on risk, such that potential threats in the presence of, or arising from, ingroup members will be perceived as less risky and will inspire greater risk taking behavior than potential threats in the presence of, or arising from, outgroup members (H1). Subsequent hypotheses concern the processes through which shared group membership affects risk. Building on research showing that ingroup members are typically trusted more than outgroup members, we hypothesize that shared group membership will affect risk by increasing trust (H2). Building on existing behavioral immune system research that shows disgust is a core driver of risk avoidance, we further hypothesize that shared group membership will affect risk by lowering disgust (H3). These relationships and hypotheses are depicted in Figure 2.

We tested the model in eight studies that span a variety of risk contexts—including disease risk, financial risk, health and physical risk—and a variety of methods, including correlational, longitudinal, and experimental designs. We tested the model in *ad hoc* groups created using minimal group paradigms (Study 1) and in existing groups (Studies 2–8). We also assessed actual risk behavior in addition to risk perceptions.

Study 1 assessed risk perceptions as a function of shared group membership. This study also manipulated risk salience to test whether the model is specific to situations of risk, or is reflective of a general orientation toward ingroup versus outgroup members. Study

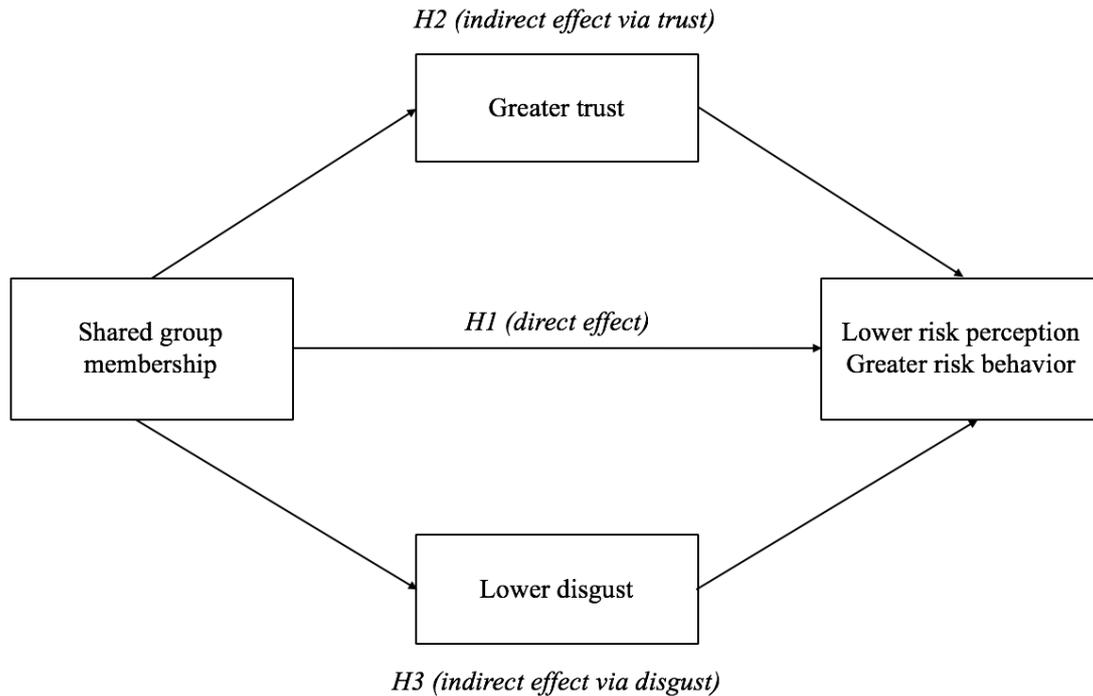


Figure 2. The Social Identity Model of Risk Taking.

2 assessed risk taking behavior as a function of shared group membership. Study 3 tested the moderating influence of social identification, predicting in line with social identity principles that the effect of shared group membership should be particularly pronounced among those who identify strongly with the ingroup. This study also tested the full model, including the mediating paths via trust and disgust. Studies 4 and 5 were conducted in the field, to show ecologically valid evidence for the model among people attending two large-scale festivals with a reputation for risk taking among attendees. Study 6 provided causal evidence for the proposed mediator of trust, by manipulating trust alongside group membership and measuring the impact on risk perceptions. Studies 7 and 8 further probed the nature of the mechanism, testing whether different subdimensions of trust (ability, benevolence, and integrity) explained the relationship between shared group membership and risk.

Open science statement. We aim for transparency in our data and analyses.

Accordingly, we have made our data available on the Open Science Framework (OSF) at this

[link](#). Alongside the data we provide code to reproduce all analyses in the paper. We are transparent about data exclusions made in each study, which were conducted to preserve data quality by excluding automated or inattentive responding, and we applied most of these exclusions prior to conducting main analyses (see Table S2 in supplementary materials for a summary of exclusions). In supplementary materials we report the details of five additional studies that are relevant to the package but had limitations related to design or fit with other studies (Studies 9-12), or which replicated a study in ways that would be redundant to include in the main paper (Study 13). As such, we have no ‘file drawer’ studies. Moreover, we include the results from all 13 studies in the internal meta-analysis conducted at the end of this article.

Study 1: Disease Risk Perception among Undergraduate Students

Study 1 was an initial test of the model. Specifically, this study was designed to test H1: that ingroup members would be perceived as less risky than outgroup members. In order to test this assumption, we employed a minimal group paradigm to assign participants to groups that held no prior meaning. This paradigm allows researchers to establish the basic underpinnings of group behavior (e.g., discrimination; Tajfel et al., 1971) without concern that prior experience with, or norms of, the group might contaminate the effects. In addition, it ensures that group assignment is fully random with no confounding variables associated with group affiliation. Study 1 investigated whether mere shared group membership was enough to lower risk perceptions. In short, we tested whether people would perceive ingroup members as posing less of a disease risk than outgroup members even though the shared group membership was novel and of short duration.

In addition to manipulating shared group membership, we also manipulated disease risk in this study to be low or high. Our model allows for the possibility that ingroup members may be perceived as posing less of a risk than outgroup members regardless of

whether risk is salient or not. However, the effect of shared group membership on risk perception may be stronger when people are particularly aware of risk in the environment. Indeed, such an effect would help to rule out the explanation that the effect is due to a general ingroup favoritism bias, whereby ingroup members are simply liked more and so are perceived as less risky (e.g., in the same way that they might also be perceived as less arrogant or neurotic). Instead, if risk discounting is an outcome of shared group membership, we would expect to see that ingroup members are perceived as less risky than outgroup members *particularly* when risk is salient, and not necessarily when risk is not salient.

Method

Participants and Design

The study employed a 2 (group membership: ingroup vs. outgroup) x 2 (disease salience: high vs. low) between-subjects design. Perceived contagion risk was the dependent variable. Originally, 160 participants completed the study. The majority (89%) of participants were first year psychology students who received partial course credit for participation, and the remainder were volunteers. As detailed below, 37 participants were excluded for incorrectly responding to a comprehension check. These exclusions were performed prior to completing the analyses. This yielded a final sample size of 123 people (71% female) aged from 16 to 55 ($M = 20.39$; $SD = 5.44$). According to G*Power, 128 participants would provide 80% power to detect a medium effect size ($d = 0.50$).

Procedure

Participants were recruited for a study on “color perception and remote visual cooperation”. On arriving, participants were told that remote cooperation involved cooperating on a task without actual interaction, and that their partner had already started the task they would be completing (building a Lego model). The participant was to carry on from where their partner left off. After providing informed consent, participants completed a short

standardized color perception task (five plates from the Ishihara Color Vision Test; Ishihara, 1917). Participants were then randomly allocated to either the red or green color perception group (ostensibly on the basis of their responses).

Participants were then instructed to move to a desk where they would work on the Lego model (LEGO Creator 6911 Mini Fire Truck set, 2011) that had been ostensibly started by their partner. To increase the salience of group membership, a colored sign representing participant's own group membership was placed in front of their seat. Across from the participant (ostensibly where their partner had sat) was a colored sign indicating the group membership of their partner. For all participants, their partner was assigned to the green group. This served as the manipulation of *group membership*: participants assigned to the red color group were aware their partner was an outgroup member, while participants assigned to the green color group were aware their partner was an ingroup member.

Disease salience was manipulated at this point using visual cues. Participants in the high disease salience condition were exposed to what appeared to be used tissues on the table near the Lego model and the partner's group membership sign. The experimenter apologized for the tissues and told the participant that their partner had been sick. Participants in the low disease salience condition saw no tissues and were not told anything about their partner.

At this point, the experiment reinforced the group membership manipulation by gesturing to the partner's colored sign and telling the participants that their partner was in the green group. The experimenter then indicated on the instructions what stage their partner had reached on the Lego model (stage 6). Participants were instructed to try to complete as much of the model as possible in four minutes (up to a total of 16 stages). After four minutes, participants completed the dependent variable, manipulation check, and demographic questions, and were debriefed as to the purpose of the study.

Measures

Risk perception. Following recommendations by Brewer and colleagues (2007), we used a state-based measure of risk perception specific to the experimental context (adapted from a trait measure of Perceived Vulnerability to Disease; Duncan, Schaller, & Park, 2009). A single item assessed risk perception: “When you were completing the Lego task, how worried did you feel about catching a contagious illness?” Participants completed the item on a scale ranging from 1, *I didn’t feel very vulnerable* to 5, *I felt very vulnerable*.

Comprehension check. Participants were asked two questions to assess their understanding. The first question assessed comprehension of the group membership manipulation: “My experiment partner was in the…” with response options “Red group” and “Green group”. The second question assessed comprehension of the disease salience manipulation: “Were you told that the previous participant was sick?” with response options “Yes”, “No” and “I don’t remember”.

Results

Comprehension Check

As noted above, 23% of originally tested participants failed one or both comprehension checks. This was due to participants failing to notice their partner was sick in the high disease salience condition ($n = 24$) or misidentifying their partner’s group membership ($n = 18$), or in some cases committing both of these errors. These participants were excluded to yield the final sample of 123.

Risk Perception

A 2 (group membership: ingroup vs. outgroup) x 2 (disease salience: high vs. low) between-subjects ANOVA was applied to perceived risk of contagion. Suggesting an effective manipulation, there was a significant main effect of disease salience, $F(1,119) = 10.77$, $p = .001$, $\eta_p^2 = .08$, such that participants in the high disease salience condition ($M =$

1.73, $SD = 1.15$) perceived greater risk of contagion than participants in the low salience condition ($M = 1.18$, $SD = 0.65$). Supporting H1, there was a significant main effect of group membership, $F(1,119) = 4.94$, $p = .028$, $\eta_p^2 = .04$, such that participants in the ingroup condition ($M = 1.27$, $SD = 0.73$) perceived lower risk of contagion than participants in the outgroup condition ($M = 1.63$, $SD = 1.13$).

In addition, there was a significant interaction, $F(1,119) = 6.06$, $p = .015$, $\eta_p^2 = .05$. Planned comparisons revealed that when disease salience was high, participants perceived significantly lower risk of contagion with an ingroup partner ($M = 1.33$, $SD = 0.80$) than an outgroup partner ($M = 2.09$; $SD = 1.30$), $F(1,119) = 11.06$, $p = .001$, $\eta_p^2 = .09$. However, when disease salience was low, there was no significant effect of group membership on perceived risk, $F(1,119) = 0.03$, $p = .867$. This interaction is depicted in Figure 3.

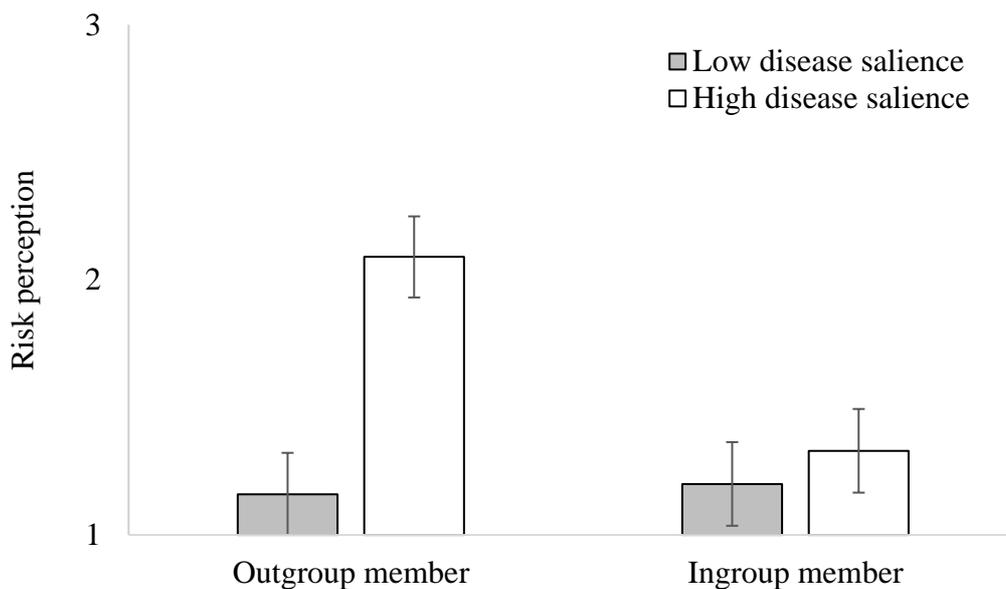


Figure 3. Ingroup members are perceived as posing less disease risk than outgroup members, particularly when disease salience is high. *Note.* Error bars represent standard errors.

Discussion

Study 1 provided preliminary evidence for the model. As predicted by H1, participants viewed ingroup members as posing less of a disease risk than outgroup members.

We observed these effects even though group membership in this case was based on *ad hoc* assignment to groups that had no prior meaning to participants. In addition, we found that the effect of group membership was strongest when the risk of disease was salient. This was an important finding, as it indicates that it is not the case that the results can be explained by an ingroup favoritism effect, where ingroup members are indiscriminately perceived more positively than outgroup members. Rather, perceptions changed according to the context: when there was no disease risk, ingroup members were not perceived as significantly less risky than outgroup members. It was only when the environment cued people to the fact that disease risk may be present that shared group membership provided a method of distinguishing those who pose a risk from those who do not. However, Study 1 had limitations, including insufficient power to test H2 and H3, as well as a high failure rate of the manipulation checks, which we sought to address across the rest of the empirical studies.

Study 2: Financial Risk Behavior among Community Members

Study 2 utilised one of the most commonly used behavioral risk taking tasks in psychological research: the Balloon Analogue Risk Task (BART; Lejuez et al., 2002). In this paradigm, participants are given the opportunity to inflate a number of (virtual) balloons. Balloons have a probabilistic risk of exploding that escalates systematically with each inflation. Participants receive a pay-off for each inflation, however, these earnings are not secure—if a balloon explodes, all the earnings from that balloon are lost. The BART is associated with related psychological constructs, such as impulsivity and sensation seeking (Hunt, Hopko, Bare, Lejuez, & Robinson, 2005; Lauriola, Panno, Levin, & Lejuez, 2014), and has been found to predict risk taking across a variety of domains (Lejuez, Aclin, Zvolensky, & Pedula, 2003).

An advantage of using the BART is that it provides two validated behavioral proxy measures for risk taking: adaptive risk taking, which is the average winnings that participants

received on balloons that did not explode, and maladaptive risk taking, which is the total number of balloons which exploded, yielding no winnings. Therefore, Study 2 was able to provide some insight into the question of whether the link between shared group membership and risk taking is predominantly adaptive or maladaptive in its consequences.

The BART was adapted to explore the role of shared group membership by presenting participants with two types of balloons emblazoned with the name and logo of either the Republican or Democratic parties. Shared group membership was manipulated in this way within-subjects, so that people saw a mixture of ingroup and outgroup balloons (with shared group membership determined by their own political affiliation). Participants were informed that the logo on each balloon represented who had set the explosion point of the balloon for them—an ingroup member or an outgroup member. We expected participants to take more risks for ingroup balloons than outgroup balloons (H1).

To this within-subjects manipulation of shared group membership, we also introduced a between-subjects manipulation of ingroup risk salience. We varied whether ingroup balloons were consistently high risk (with a low average explosion set-point of 32 pumps)—or whether ingroup balloons were consistently low risk (with a high average explosion set-point of 128 pumps). Outgroup balloons were assigned the opposite risk profile. Consistent with prior research on risk taking, we expected participants to both earn a greater proportion of the available money per balloon (i.e., adaptive risks) and explode more balloons (i.e., maladaptive risks) when balloons were high risk compared to low risk. This is because under high risk conditions each pump represents a greater opportunity for gain but also greater opportunity for loss. Critically, we expected this difference to be exacerbated when the balloon explosion set-points were apparently set by ingroup members (rather than outgroup members). In short, this pattern of results would demonstrate that people take more risks with ingroup members than outgroup members, especially under conditions of high risk.

Method

Participants and Design

Study 2 employed a mixed design with two independent variables: group membership (within subjects: ingroup vs. outgroup) and ingroup risk salience (between subjects: outgroup high risk and ingroup low risk, vs. ingroup high risk and outgroup low risk). This yielded a design whereby all participants saw both ingroup and outgroup balloons, but varied across conditions in terms of which type of balloon (ingroup or outgroup) was allocated as high risk. Risk behavior was the focal dependent variable.

A total of 253 people were recruited via Amazon's Mechanical Turk. Following exclusions based on an attention check and suspicious responding, the final sample consisted of 222 American participants (63% male) aged from 18 to 69 ($M=34.43$, $SD=10.58$). Participants were predominantly White (71%), Asian (13%), Hispanic/Latino/a (10%) or Black (7%; note that participants could select multiple ethnic categories). According to G*Power, 244 participants would provide 80% power to detect a small between-groups effect size of $\eta^2 = .04$ (as found in Study 1) at the standard .05 alpha error probability.

Procedure

The study was advertised as investigating Politics and Social Relationships. Participants were offered a minimum \$1.50 incentive for taking part in the study (which took 15 minutes on average), plus their winnings in the balloon game ($M = \$3.69$; $SD = \$1.16$). The study commenced by asking participants "Which political party do you identify with the most?"; response options were *Republican* and *Democrat*. This was a forced choice question, with the clarification instruction provided that people who did not especially identify with either should choose their preference of the two parties.

Participants were then given standard instructions describing the balloon task. Additional instructions described the two types of balloons (Republican and Democrat) and

described how the average set-points had been determined by previous participants. To increase the face validity and comprehension of this information, participants were given the opportunity to provide a setpoint for future Democratic balloons and Republican balloons using a sliding scale with no numerical anchors.

Participants then completed the balloon task, in which they were presented with 30 balloons in a random order, 15 with a Republican party symbol and 15 with a Democratic party symbol (see Figure 4). Participants' winnings (for each balloon and in total) were presented in cents throughout to ensure the behavioral consequences remained salient throughout the task.



Inflate balloon	
Collect \$\$\$	
Earn this round:	15 cents
Pumping times:	15 times
Win last round:	0 cents
Total credit:	8 cents

Figure 4. Study 2 stimuli. *Note.* Participants were presented with 30 balloons in random order (15 Democrat and 15 Republican). Participants could press “Inflate balloon” as many times as they chose and each inflation increased the size of that balloon and its value by one cent. They could also press “Collect \$\$\$” to secure their winnings on this balloon and proceed to the next. If the balloon burst before they pressed “Collect \$\$\$”, all winnings for that balloon were lost.

Measures

Risk behavior. Following Hunt et al. (2005), we calculated two measures of risk taking for each balloon type (ingroup versus outgroup). The first was the adjusted average

number of balloon pumps, excluding those balloons which exploded. This is a measure of *adaptive risk taking*, since a higher score represents a greater total reward for participants.¹

The second was the number of total balloon explosions. This is a measure of *maladaptive risk taking*, since each explosion led to the loss of all rewards accumulated for that balloon. These two measures of risk taking were correlated in this task at $r = .88$ (for same group comparison) or $r = .70$ (totals).

Attention check. Two attention checks were embedded in the study that stated “This item is to check you are paying attention. Please select “strongly agree”.

Results

First, two strategies were used to identify and exclude participants with low quality data. People who failed either attention check were excluded ($n=22$). In addition, suspicious responders were identified via inspection of multivariate outliers, which revealed a small cluster of participants ($n=9$) with the lowest total winnings from the balloon game ($< \$1$), who each also showed at least one of the following anomalies: (1) very short duration in completing the experiment (took less than 7 minutes to complete the whole study), and (2) patterned responding in the balloon task that suggested automated or inattentive responding (e.g., each balloons pumped only once; or each balloon pumped until bursting point such that adaptive risk taking could not be calculated). A total of 31 participants (12%) were excluded from analyses. This left 222 participants in the final sample. Follow-up analyses revealed that entering political affiliation as a covariate did not substantively change the results.

¹ The between-subjects manipulation of risk salience meant that adaptive risk taking needed to be transformed in order to accurately reflect the fact that a single pump involved greater risk taking for high risk balloons than for the low risk balloons. Adaptive risk taking was transformed to a 0-100 scale across the low and high risk conditions. A participant’s score on the adaptive risk taking measure can thus be interpreted as a rough percentage of the available financial bonus that they secured. This transformation was unnecessary for maladaptive risk taking because the potential range was the same for both low and high risk conditions (0-15).

Adaptive Risk Taking

The main effect of ingroup risk salience was not significant, $F(1,220) = 0.01$, $p = .923$, $\eta_p^2 < .01$. Supporting H1, the main effect of group membership was significant, $F(1,220) = 4.17$, $p = .042$, $\eta_p^2 = .02$, such that participants engaged in more adaptive risk taking for ingroup balloons ($M = 38.20$, $SD = 20.25$) than outgroup balloons ($M = 35.50$, $SD = 18.79$) and thus secured a greater proportion of the available winnings for those balloons.

The main effect was qualified by a significant interaction, $F(1,220) = 460.65$, $p < .001$, $\eta_p^2 = .68$. This interaction was such that participants always took more adaptive risks with high compared to low risk balloons, but that this effect was significantly stronger when high risk balloons were ingroup, $F(1,220) = 283.90$, $p < .001$, $\eta_p^2 = .56$, compared to outgroup, $F(1,220) = 183.63$, $p < .001$, $\eta_p^2 = .46$. This meant that when outgroup balloons were high risk, participants tended to secure a greater percentage of available winnings from outgroup balloons ($M = 46.76$, $SD = 16.43$) than ingroup balloons ($M = 26.75$, $SD = 16.11$), but—even more strongly—when ingroup balloons were high risk, participants secured a greater percentage of available winnings from ingroup balloons ($M = 49.05$, $SD = 17.66$) than outgroup balloons ($M = 24.84$, $SD = 14.11$). This interaction is depicted in Figure 5 and demonstrates that adaptive risk-taking was facilitated when ingroup members were high risk.

Maladaptive Risk Taking

The main effect of ingroup risk salience was not significant, $F(1,220) = 1.06$, $p = .304$, $\eta_p^2 = .01$. Supporting H1, the main effect of group membership was significant, $F(1,220) = 4.80$, $p = .029$, $\eta_p^2 = .02$, such that participants engaged in more maladaptive risk taking for ingroup balloons ($M = 3.94$, $SD = 3.01$) compared to outgroup balloons ($M = 3.49$, $SD = 2.74$). That is, people tended to explode more ingroup balloons than outgroup balloons.

The main effect was qualified by a significant interaction, $F(1,220) = 519.74$, $p < .001$, $\eta_p^2 = .70$. This interaction was such that participants always took more maladaptive

risks with high compared to low risk balloons, but that this effect was significantly stronger when high risk balloons were ingroup, $F(1,220) = 320.92, p < .001, \eta_p^2 = .59$, compared to when they were outgroup, $F(1,220) = 206.72, p < .001, \eta_p^2 = .48$. This meant that when outgroup balloons were high risk, participants exploded more outgroup balloons ($M = 5.23, SD = 2.67$) than ingroup balloons ($M = 1.93, SD = 1.48$), but—even more strongly—when ingroup balloons were high risk, participants exploded more ingroup balloons ($M = 5.84, SD = 2.86$) than outgroup balloons ($M = 1.83, SD = 1.49$). This interaction is depicted in Figure 6 and demonstrates that maladaptive risk-taking was enhanced when ingroup members were high risk.

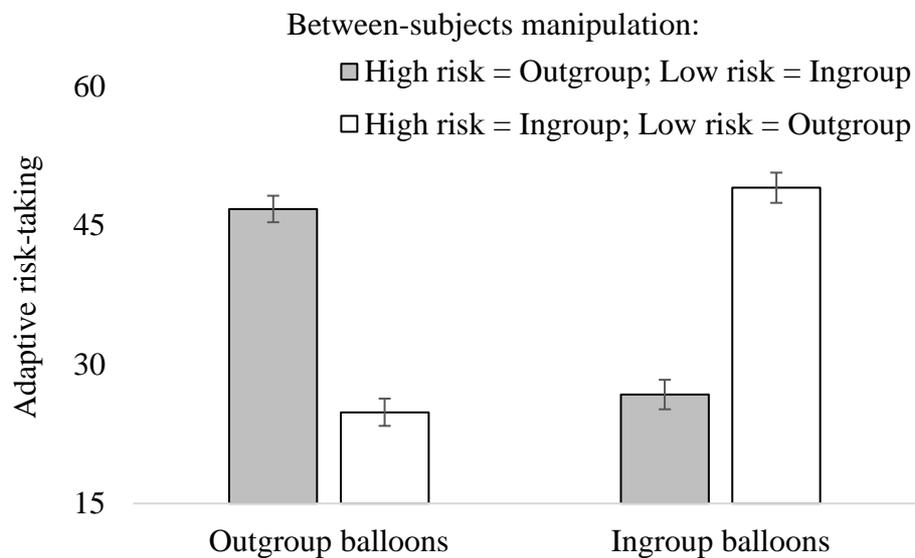


Figure 5. Ingroup members promote greater adaptive risk taking than outgroup members, particularly when they were high risk rather than low risk. *Note.* All participants saw both ingroup and outgroup balloons, with an additional between-subjects manipulation (represented in separate bars) of which balloons were high risk: ingroup or outgroup balloons (with the other balloon type being low risk). Adaptive risk taking was a transformed measure of participants' winnings which roughly corresponded to the percentage of the available financial bonus that participants secured. Error bars represent standard errors.

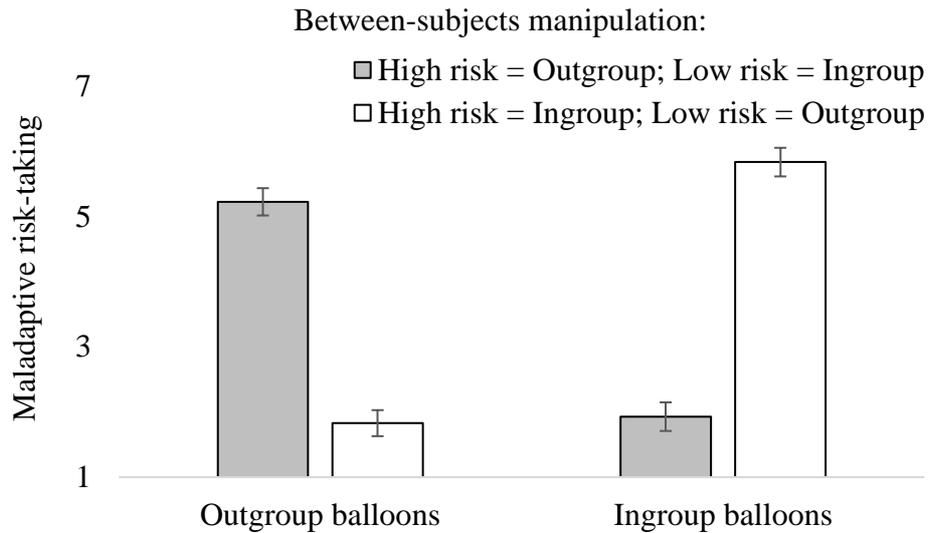


Figure 6. Ingroup members promote greater maladaptive risk taking than outgroup members, particularly when they were high risk rather than low risk. *Note.* All participants saw both ingroup and outgroup balloons, with an additional between-subjects manipulation (represented in separate bars) of which balloons were high risk: ingroup or outgroup balloons (with the other balloon type being low risk). Maladaptive risk taking was the number of balloons which exploded for each participant, leading to a loss of potential winnings. Error bars represent standard errors.

Discussion

Study 2 provided additional support for the model. As predicted by H1, participants engaged in more risk behavior on a task ostensibly set by ingroup members compared to outgroup members. This study builds on Study 1 in several ways. First, it shows that the effect of shared group membership on risk extends to behavior with real-world consequences—in this case, financial earnings. Second, it helps to clarify the nature of the risks taken by participants on the basis of shared group membership. Specifically, the BART distinguishes between adaptive (greater earnings) and maladaptive (more explosions) risk-taking. We found that participants took more adaptive *and* maladaptive risks with ingroup members compared to outgroup members, meaning they made more money on balloons that

did not explode, but also lost more money due to balloons exploding. Thus, the risks people take with ingroup members sometimes pay off, but also have the potential to fail.

Finally, while Study 1 presented ingroup and outgroup members as posing a potentially unwitting and unintended risk (i.e., carrying disease), Study 2 presented ingroup and outgroup members as making a deliberate choice about where to set the level of risk of the BART. Indeed, we found that people were more likely to take risks with ingroup balloons *particularly* when those balloons offered an opportunity for greater risk (as well as reward). This begins to hint at potential mechanisms that may be at play—it suggests that people may have *trusted* ingroup members to choose balloon set-points that were likely to be beneficial for the participant. Of course, we can only speculate on this possibility from these results and requires more targeted investigation. Testing this potential mechanism was a goal of Study 3.

Study 3: Disease Risk Perception in a Global Sample

Study 3 examined the model in the context of another pre-existing group membership: nationality. Like other existing group memberships, nationality is a social category for which we might expect people to vary in their level of commitment. Accordingly, we introduced social identification as a potential moderator in Study 3, predicting that we might be more likely to observe evidence for the model among people for whom nationality is a more relevant psychological group. Such an effect would be consistent with a tradition in social identity research demonstrating that high identifiers are more likely to act in group-based ways and respond to cues that have implications for the group than low identifiers (e.g., Louis et al., 2007).

Study 3 provided the first full test of the model, in particular assessing the hypothesized mediating roles of trust and disgust on risk perception in the context of disease risk. As in previous studies, we hypothesized that ingroup members would be perceived to pose less of a risk than outgroup members (H1) and that the effect of shared group

membership on risk perception would be mediated by trust (H2) and disgust (H3). However, in Study 3, we additionally hypothesized that these effects would be observed most strongly among participants who were highly identified with their national group.

Method

Participants and Design

The study employed a two-level between-subjects design that manipulated group membership (ingroup vs. outgroup) based on nationality. In addition, we measured social identification as a continuous moderator. Risk perception was the focal dependent variable and trust and disgust served as potential mediators.

The study recruited 607 participants (50% female) aged from 17 to 70 ($M=29.77$, $SD=11.06$) using a variety of strategies including a university recruitment pool of community members, online advertising, and several online participant recruitment services. Our recruitment strategy was successful in capturing a diverse global sample, with 40 nationalities represented. Most common nationalities were USA (27%), Indian (25%), Australian (21%), and Chinese (5%). Approximately two-thirds (62%) of participants spoke English at home. Of the original sample, 22 (4%) did not complete all of the key measures, yielding approximately 585 participants for analyses (all available data was used for each analysis, and so the degrees of freedom differ slightly and are reported below). According to G*Power, 506 participants would provide .80 power to detect a small between-groups effect size of $d = .25$ at the standard .05 alpha error probability.

Procedure

Participants completed a national group identification scale and then read a scenario about a meeting with a colleague in a coffee shop, as follows:

You are at a coffee shop having coffee with a colleague who shares your nationality [is of a foreign nationality]. Your colleague asks to try your coffee. You pass it to

your colleague and your colleague drinks from your cup. You then take your coffee cup back so that you can continue drinking your coffee.

Participants were randomly assigned either a condition that described the colleague as having the same nationality (ingroup condition, coded as 1) or having a foreign nationality (outgroup condition, coded as 0). Participants then completed the remaining measures and demographic questions.

Measures

National group identification. The four-item social identification scale (FISI; Doosje, Ellemers, & Spears, 1995; Postmes, Haslam, & Jans, 2013) was included to measure strength of identification with one's nationality. Participants responded to questions such as "I feel strong ties with members of my nationality" on a scale ranging from 1, *strongly disagree* to 7, *strongly agree* ($\alpha = .80$).

Risk perception. Participants were asked to rate their perceptions of risk in the scenario using two items (adapted from Cruwys et al., 2018; Peters, Burraston, & Mertz, 2004): "To what degree do you feel that you have taken a risk?"; "To what degree do you feel that you are at risk of disease?") on a 0-100 sliding scale ranging from *Not at all* to *Extremely* ($r = .85, p < .001$).

Trust. Trust was measured using a four-item scale from the Partner Opinion Questionnaire (POQ; Singh et al., 2009), e.g., "I would find [my colleague] to be dependable"). These items were scored on scale ranging from 1, *strongly disagree* to 7, *strongly agree*, $\alpha = .86$.

Disgust. State feelings of disgust were measured using the 12-item State Disgust Scale (Bates & Chadwick, 2015, e.g., "I felt sick") scored on a scale ranging from 1 *Not at all* to 5, *A great deal*, $\alpha = .95$.

Results

Risk Perception

A hierarchical multiple regression was applied to risk perceptions. Step 1 included the group membership manipulation and centered national identification. Step 2 included the interaction between the group membership manipulation and centered national identification.

Together, the variables in Step 1 accounted for a significant amount of variance, $R^2 = .03$, $F(2,582) = 9.13$, $p < .001$. Consistent with H1, there was a significant effect of shared group membership, $\beta = -.08$, $p = .044$, such that sharing a drink with an ingroup member was perceived as significantly less risky than sharing a drink with an outgroup member.

Unexpectedly, national identification was associated with greater perceived risk, $\beta = .15$, $p < .001$. The interaction was not significant at Step 2, $R^2_{\Delta} < .01$, $F_{\Delta}(1,581) = 1.38$, $\beta = -.07$, $p = .240$.

Trust

A hierarchical multiple regression was conducted with the outcome variable of trust following the procedure described above. Together, the variables in Step 1 accounted for a significant amount of variance, $R^2 = .07$, $F(2,583) = 21.60$, $p < .001$. Unexpectedly, there was no significant effect of shared group membership on trust, $\beta = .04$, $p = .296$, although there was a significant positive effect of national identification on trust, $\beta = .26$, $p < .001$.

However, qualifying these effects there was a significant interaction at Step 2, $R^2_{\Delta} = .01$, $F_{\Delta}(1,582) = 4.51$, $\beta = .12$, $p = .034$. Inspection of the simple slopes revealed that ingroup members were trusted more than outgroup members among people high (+1SD) in national identification, $\beta = .13$, $p = .027$, but were not trusted more among people low (-1SD) in national identification, $\beta = -.05$, $p = .424$.

Disgust

A hierarchical multiple regression was conducted with the outcome variable of disgust following the procedure described above. Together, the variables in Step 1 accounted for a significant amount of variance, $R^2 = .02$, $F(2,583) = 4.77$, $p = .009$. There was no significant effect of shared group membership on disgust, $\beta = -.06$, $p = .123$, although there was a significant positive effect of national identification on disgust, $\beta = .11$, $p = .008$. There was no significant interaction at Step 2, $R^2_{\Delta} < .01$, $F_{\Delta}(1,582) = 1.66$, $\beta = -.08$, $p = .199$.

Indirect Effects

We used Hayes' (2018) PROCESS Model 8 with 10,000 bootstrapping resamples to assess the moderated indirect effect of shared group membership on risk perception via trust and disgust, with national identification as the moderator. The model is depicted in Figure 7.

The index of moderated mediation for the indirect effect via trust was significant, $IE = -1.37$, $SE = 0.73$, 95% $CI = -2.91$ to -0.02 . Supporting H2, this revealed a significant indirect effect of shared group membership on risk perception via trust among high identifiers ($IE = -2.13$, $SE = 1.08$, 95% $CI = -4.53$ to -0.17). However, there was no significant indirect effect of shared group membership on risk perception via trust among low identifiers ($IE = 0.74$, $SE = 0.98$, 95% $CI = -1.14$ to 2.75).

The index of moderated mediation for the indirect effect via disgust was not significant, $IE = -1.99$, $SE = 1.58$, 95% $CI = -5.16$ to 1.09 . Contrary to H3, there were no significant indirect effects of shared group membership on risk perception via disgust among either high identifiers ($IE = -3.88$, $SE = 2.45$, 95% $CI = -8.67$ to 0.94) or low identifiers ($IE = -0.69$, $SE = 2.15$, 95% $CI = -4.94$ to 3.48).

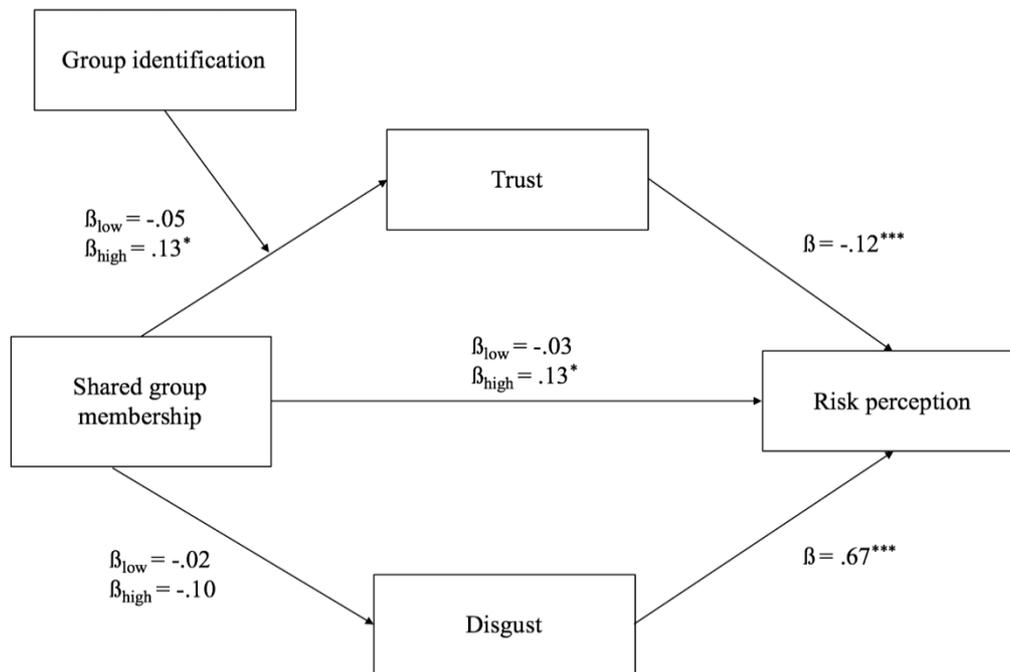


Figure 7. Moderated mediation model showing that shared group membership (ingroup vs. outgroup) lowers risk perceptions via greater trust among people high (but not low) in group identification.

Discussion

Study 3 provided the first full test of the model in a large and culturally diverse sample from around the world. The results revealed, consistent with Studies 1 and 2, that ingroup members were perceived as posing less of a risk than outgroup members. Ingroup members were also trusted more than outgroup members when participants were high identifiers, and it was via this greater trust that shared group membership appeared to lower risk perceptions in this group. This provided support for H1 and H2, but H3 was not supported: there was no effect of shared group membership on disgust; and disgust did not act as a mediator of the effect of shared group membership on risk perceptions.

Importantly, people vary in commitment to existing groups, and consistent with evidence showing that social identification shapes group behavior (Terry & Hogg, 1996), we found that the effects predicted by our model emerged particularly among people who were

high in social identification. Specifically, high identifiers perceived ingroup members to be less risky than outgroup members (supporting H1) because ingroup members are trusted more than outgroup members (supporting H2). However, we found no evidence of a moderated mediation pathway via disgust (contrary to H3).

Study 4: Health Risk Perception in the Field

Studies 1 to 3 provided experimental evidence for the model in *ad hoc* and existing groups. Experimental designs are obviously needed to provide evidence of causality, but the laboratory and online settings of experiments can lack ecological validity. To provide evidence for these processes as they occur naturally in people's lives, we conducted two field studies that investigated risk perception (Study 4) and risk behavior (Study 5) among festival attendees. Study 4 was a large cross-sectional survey among young people attending the largest youth mass gathering in Australia: *Schoolies*, a week-long celebration of the end of formal schooling that occurs annually at coastal locations around the country. Every year tens of thousands of young people attend the festivities (approximately a third of all school leavers; Kang, 2014), which present an opportunity to celebrate graduation but also increase the likelihood of encountering risk. Common risks at *Schoolies* include drug taking, binge drinking, drink spiking, violence, and unprotected sex (Cruwys et al., 2018; Lubman, Droste, Pennay, Hyder, & Miller, 2014; Pettigrew et al., 2015; Salom, Watts, Kinner, & Young, 2005)

In Study 4, we tested whether participants underestimated the health and physical risks of attending *Schoolies* to the extent that they felt a shared identity with other attendees (H1). As we were not experimentally manipulating group membership, and building on the findings of Study 3, we assessed perceptions of shared group membership via measured social identification with fellow attendees. We further tested if the relationship between shared identity and risk perception was mediated by greater feelings of trust (H2). In

addition, we assessed a range of control variables that might plausibly be associated with risk perceptions. Specifically, we tested whether the hypothesized effects remain significant while controlling for trait impulsivity and normative influence. This is relevant because these are influential models that have been invoked to explain risk taking, particularly among adolescents (e.g., Gullone & Moore, 2000; Lauriola et al., 2014) and because the Schoolies context is one in which social norms marketing campaigns are common (e.g., Hutton, 2012), making these strong candidate variables that may influence risk perception.

Method

Participants and Design

The study measured social identification, risk perception, and trust as core variables of interest, and trait impulsivity and normative influence as control variables. Originally, data were collected from 1,596 participants who attended Schoolies in 2016 ($n = 607$) or 2017 ($n = 989$). Participants who did not pass the attention check (202 people; 13% of the sample) or whose self-reported age was implausible (due to reporting or sampling error e.g., 2 years; 7 people; < 1% of the sample) were excluded prior to analysis. The final sample comprised 1387 young people (55% female). As expected given the nature of the event, the majority (78%) of participants were 17, and ranged in age from 16 to 19. According to G*Power, 1387 participants would provide .99 power to detect a small effect size of $\rho^2 = .02$ in a linear multiple regression with two predictors at the standard .05 alpha error probability.

Procedure

Participants were approached during the Schoolies festival in November in 2016 and 2017. Research assistants approached participants at the wristband distribution center (wristbands are used to verify that students are “legitimate” school leavers and Schoolies attendees) and in public areas including a main street mall. Only people wearing wristbands were approached. The samples were independent with no repeat participants and were thus

analyzed as a combined dataset. The sample size for individual analyses varies slightly because participants were not excluded if they had missing data on some but not all measures of interest. Participants received a \$5 voucher for a local sandwich vendor as an incentive for participating.

Measures

Social identification. Identification with fellow attendees at the mass gathering was measured using the Single Item Social Identification scale (SISI; Postmes et al., 2013): “I identify with other Schoolies”, on a scale ranging from 1, *not at all* to 7, *very much so*. This scale has been found to be reliable and valid, with comparable psychometric properties to longer social identification scales (Reysen, Katzarska-Miller, Nesbit, & Pierce, 2013; Postmes et al., 2013).

Trust. A single-item measure of trust was included (“I can trust other Schoolies”, adapted from Helliwell & Wang, 2010) and scored on a scale ranging from 1, *not at all* to 7, *very much so*.

Risk perception. Two measures were employed to assess risk perception. The first was a single-item appraisal measure of global risk perception, similar to the items included in Study 3. This assessed the degree to which respondents felt that “Being at Schoolies is safe” scored on a scale from 1, *not at all* to 7, *extremely*. This item was reverse-scored for inclusion in the analyses, such that higher scores indicated greater global risk perception.

The second measure was made up of ten items on which participants rated how risky they thought it would be to engage in specific behaviors while at Schoolies on a scale from 1, *not at all risky* to 7, *extremely risky*. The items included (1) leaving my belongings on the beach while going for a swim, (2) asking another Schoolie to look after my belongings on the beach, (3) leaving my hotel room unlocked, (4) walking home alone at night, (5) getting separated from my friends at night, (6) having sex with a Schoolie I just met, (7) not using

protection while having sex, (8) taking drinks from a Schoolie I just met, (9) sharing drinks, and (10) leaving my drink unattended.

All items (global and specific risk perceptions) combined to form a reliable scale of Schoolies-related risk perceptions, $\alpha = .83$. Because they were measured on slightly different scales, the global risk perception item and specific risk perception items were standardized prior to being combined into a risk perception scale.

Descriptive norms for risk taking. Participants were asked to consider five different risk behaviors in terms of how common they were in their friendship group. The specific items (e.g., “Most of my friends will be drinking to get drunk”) were informed by norm focus theory (Kallgren, Reno, & Cialdini, 2000) and adapted from Bodimeade and colleagues (2014). The items were scored on a scale ranging from 1, *Strongly disagree* to 7, *Strongly agree*, $\alpha = .81$.

Trait impulsivity. Personality pre-disposition towards impulsivity is a well-established risk factor for risk taking and substance misuse. We included seven items from the Substance Use Risk Profile Scales (Woicik, Stewart, Pihl, & Conrod, 2009; e.g., “I usually act without stopping to think”) measured on scale ranging from 1, *strongly disagree* to 4, *strongly agree*, $\alpha = .84$.

Attention check. The attention check was identical to that used in Study 2.

Results

Risk Perception

A hierarchical multiple regression was used to assess H1, with risk perception as the dependent variable. In Step 1, the control variables—trait impulsivity, descriptive norms, and year—were entered. This allowed us to determine whether the relationship between social identification and risk perception persisted while controlling for the year data were collected

and other risk-related predictors². In Step 2, the main predictor, social identification, was entered. In Step 3, the hypothesized mediator, trust, was entered.

At Step 1, the control variables accounted for a significant amount of variance, $R^2 = .03$, $F(3,1310) = 12.94$, $p < .001$. Trait impulsivity was a significant negative predictor, $\beta = -.15$, $p < .001$, and descriptive norms was a significant positive predictor, $\beta = .12$, $p < .001$, of risk perception. Year of data collection was not a significant predictor of risk perception, $\beta = .01$, $p = .677$. Consistent with H1, social identification accounted for a significant amount of additional variance at Step 2, $R^2_{\Delta} = .02$, $F_{\Delta}(1,1309) = 20.87$, $\beta = -.12$, $p < .001$, such that greater perceived shared identity was associated with lower risk perception. The model at Step 3 was also significant, $R^2_{\Delta} = .04$, $F_{\Delta}(1,1308) = 60.73$, $\beta = -.24$, $p < .001$, such that greater trust in fellow Schoolies attendees was associated with lower risk perception.

Trust

A hierarchical multiple regression was applied to trust in fellow Schoolies attendees. The model was significant at Step 1, $R^2 = .01$, $F(3,1310) = 6.22$, $p < .001$. As with risk perception, trait impulsivity was a significant positive predictor, $\beta = .08$, $p = .008$, and descriptive norms was a negative predictor, $\beta = -.10$, $p = .001$ ($\beta_{\text{year}} = .05$, $p = .066$) of trust. The model at Step 2 was also significant, $R^2_{\Delta} = .26$, $F_{\Delta}(1,1309) = 463.77$, $\beta = .51$, $p < .001$, such that greater social identification was associated with greater trust.

Indirect Effect

We used Hayes' (2018) PROCESS Model 4 with 10,000 bootstrapping resamples to assess the indirect effect of shared group membership on risk perception via trust. Consistent

² Note that all relationships we report here remain significant when not controlling for trait impulsivity, descriptive norms, and year. The relationship between social identification and risk perception remained significant, $R^2 = .02$, $F(1,1333) = 22.44$, $\beta = -.13$, $p < .001$, as did the relationship between trust and risk perception, $R^2_{\Delta} = .05$, $F_{\Delta}(1,1332) = 71.31$, $\beta = -.26$, $p < .001$. Moreover, social identification remained a significant predictor of trust, $R^2 = .26$, $F(1,1336) = 472.62$, $\beta = .51$, $p < .001$, when not including control variables.

with H2, this revealed a significant negative indirect effect of shared identity on perceived risk via trust ($IE = -.07$, $SE = .01$, $95\% CI = -0.08$ to -0.05), where feeling more strongly identified with other Schoolies was associated with greater trust, and greater trust was associated with lower perceived risk. Follow-up tests revealed that the negative indirect effect remained significant when controlling for trait impulsivity, descriptive norms, and year ($IE = -.06$, $SE = .01$, $95\% CI = -0.08$ to -0.04).

Discussion

The results of Study 4 provided evidence for the model in a field context involving a large sample of graduating secondary school students attending a massive celebratory festival. In this study, shared identity predicted reduced perceptions of risk (consistent with H1). We also found evidence of an indirect effect via trust, such that shared identity was associated with greater trust in fellow attendees, which in turn predicted lower perceptions of risk (supporting H2). This study constitutes an important application of this theoretical framework, suggesting new avenues of interventions to prevent youth risk taking that do not rely on traditional models emphasizing peer pressure or impulsivity.

Study 5: Health Risk Behavior in the Field

In Study 4 it was not ethically possible to assess actual risk behavior given that the majority of the sample was under 18 years of age. Study 5 addressed this issue by assessing risk behavior among adult participants at a controversial art festival: Dark Mofo at the Museum of Old and New Art in Hobart, Tasmania. Further, we assessed the key variables longitudinally rather than cross-sectionally as in Study 4.

At this festival we surveyed attendees at two events, both of which involved risk behavior. The first event was the *Solstice Swim*, in which participants strip naked at sunrise on the Winter Solstice and run into the freezing ocean (approximately -2 degrees Celsius / 28 degrees Fahrenheit). The second event was *150.Action*, a performance art event by Nitsch

involving a slaughtered animal carcass. This event received significant negative media attention and attracted more than 22,000 signatures on a petition to ban it, as well as negative commentary from the Premier of the State Government. We posited that participating in either event at the festival involved both health risks (due to the extreme cold for attendees of the Solstice Swim and due to exposure to bovine internal matter and blood for attendees of 150.Action) as well as reputational risks (since both events attracted substantial media coverage, and the Solstice Swim involved public nudity while 150.Action was widely condemned). We tested whether shared identity with fellow attendees at baseline predicted trust and disgust experienced during the event, and risk behavior in the form of more time spent at the risky event.

Method

Participants and Design

Participants were recruited via email invitation from event organizers in the week leading up to the event. This resulted in 226 participants completing the pre-event survey, including our focal measure of social identification with other Dark Mofo attendees. Attendees were also approached in the field in the immediate lead up to each event. Following the events, participants were approached in person and contacted via email to complete a follow up survey; 198 responses were collected. Of these, 149 could be reliably linked to a Time 1 response and had sufficient data available on the key variables at both time points (a 66% retention rate: 93 from the Solstice Swim, and 58 from 150.Action). Participants (50% female) were aged from 19 to 80 ($M=41.50$; $SD = 15.60$).

Measures

Social identification. Identification with fellow attendees at the mass gathering was measured using two items from the FIS (Postmes et al., 2013): “I identify with the other people in the Solstice Swim / at 150.Action” and “I feel a strong connection to the other

people at the Solstice Swim / 150.Action”, on a scale ranging from 1, *not at all* to 7, *very much so* ($r = .74, p < .001$). Identification was measured at Time 1 (before the event).

Trust. A single-item measure of trust was included: “I can trust the other people in the Solstice Swim / 150.Action.” (adapted from Helliwell & Wang, 2010) and scored on a scale ranging from 1, *not at all* to 7, *very much so*. Trust was measured at Time 2 (after the event).

Disgust. The same 12-item state disgust measure was included as used in Studies 3 and 5, $\alpha = .86$. Disgust was measured at Time 2 (after the event).

Risk behavior. Risk behavior was operationalized as the (subjective) duration of participation in the event. Participants were asked “About how long did you stay in the water / at the event?” and specified whether their response was in minutes or hours. Due to the differing nature of the events, responses ranged from 30 seconds to 5 hours, with the average duration of participation in the Solstice Swim being 3.2 minutes ($SD = 1.93$) and the average duration of participation in 150.Action being 2.87 hours ($SD = 1.07$). To enable the data to be analyzed together, a z-score of duration responses was created at the level of the event, and these z-scores were combined to create a measure of risk behavior for analysis.

Results

Risk Behavior

A hierarchical multiple regression was applied to predict Time 2 risk behavior in the form of spending more time at the events. Time 1 social identification was entered as a predictor at Step 1 and Time 2 trust and disgust were entered as predictors at Step 2.

Consistent with H1, the model at Step 1 accounted for a significant amount of variance, $R^2 = .03, F(1,146) = 4.23, \beta = .17, p = .041$, indicating that greater perceived shared identity with fellow festival attendees at Time 1 was associated with greater risk behavior at Time 2. The variables at Step 2 accounted for a marginally significant amount of additional

variance, $R^2_{\Delta} = .04$, $F_{\Delta}(2,144) = 2.71$, $p = .070$, although only Time 2 trust was a significant positive predictor of risk behavior ($\beta_{\text{Trust}} = .19$, $p = .045$; $\beta_{\text{Disgust}} = .13$, $p = .123$).

Trust and Disgust

A hierarchical multiple regression was applied to predict Time 2 trust and disgust from Time 1 social identification. The trust model accounted for a significant amount of variance, $R^2 = .20$, $F(1,147) = 35.79$, $\beta = .44$, $p < .001$, indicating that greater perceived shared identity with fellow festival attendees at Time 1 was associated with higher trust at Time 2. The disgust model accounted for a marginally significant amount of variance, $R^2 = .02$, $F(1,146) = 2.92$, $\beta = -.14$, $p = .089$.

Indirect Effect

We used Hayes' (2018) PROCESS Model 4 with 10,000 bootstrapping resamples to assess the indirect effect of Time 1 social identification on Time 2 risk behavior via Time 2 trust and disgust. Consistent with H2, there was a significant indirect effect of shared identity on risk behavior via trust ($IE = .04$, $SE = .02$, 95% $CI = 0.01$ to 0.10). Contrary to H3, there was no significant indirect effect of shared group membership on risk perception via disgust ($IE = -.01$, $SE = .01$, 95% $CI = -0.03$ to 0.01).

Discussion

Study 5 provided longitudinal evidence for the model with a risk behavior dependent variable in an applied setting. Studies 3 to 5 provide relatively consistent evidence for mediation via trust, whereas no consistent evidence was found for the alternative proposed mediator (disgust). Given this, Study 6 returned to the laboratory to experimentally manipulate this mediator (trust) to provide additional evidence of its causal role in the model.

Study 6: General Risk Perception among Undergraduate Students

Considered together, Studies 1 to 5 provide evidence for the first two hypotheses of the model in disease, financial, health and physical risk contexts and using a mixture of

correlational, longitudinal, and experimental designs. Specifically, we found evidence that ingroup members (others with whom people identify) are perceived to pose less risk and inspire greater risk taking behavior than outgroup members (others with whom people do not identify; H1). We found further that a key process through which this effect operates is increased trust: ingroup members are trusted more than outgroup members, and thus are thought of as “safe” people who are less likely to pose a threat (H2). We found no evidence for disgust as an alternative mediator through which shared group membership impacts on risk taking, despite hypothesizing such a path *a priori* and testing for it in four studies (H3).

Study 6 advanced the empirical chain of evidence by manipulating our consistent mediator—trust. Specifically, we sought to test whether trust causally impacts on risk perception, rather than the other way around (Pirlott & MacKinnon, 2016). Accordingly, in this study we manipulated trust in order to “disable” the mediator experimentally (Jacoby & Sassenberg, 2011; Spencer, Zanna, & Fong, 2005). The model would predict that the penalty for being an outgroup member would be apparent for targets who appear untrustworthy, but negligible for targets who appear trustworthy. More specifically, when cues to trustworthiness in another person are low or medium, shared group membership may allow people to trust more fully, lowering perceptions of risk. However, when cues to trustworthiness in another person are high, the function of shared group membership is moot—here, we would expect people to perceive that other person as low risk, regardless of group membership.

In this study, we manipulated shared group membership within an existing group of college residents and manipulated trust using a well-validated stimulus set of trustworthy and untrustworthy faces (Oosterhof & Todorov, 2008; Todorov, Olivola, Dotsch, & Mende-Siedlecki, 2015). We expected ingroup members would be perceived to be less risky than outgroup members when trustworthiness was low and medium, but predicted no difference in

risk perceptions as a function of shared group membership when trustworthiness was high. Importantly, in Study 6 we distinguish between target *trustworthiness* (an independent variable; manipulated for each target using faces had been pre-validated to have low, medium, or high trustworthy features) and perceived *trust* (the mediator; measured for each target).

Method

Participants and Design

Undergraduate students from 10 different university residential colleges were recruited to participate in the study in exchange for partial course credit. Of the initial sample ($N = 128$), 10 participants were excluded because they did not live on campus, leaving a final sample of 118 participants (69% female) aged from 17 to 24 ($M = 18.68$, $SD = 1.43$). Participants were mostly White (67%) or Asian (26%).

The study used a 2 x 3 within-subjects experimental design that manipulated target group membership (ingroup vs. outgroup) and target trustworthiness (low vs. medium vs. high). Perceived trust in each target served as a mediator and perceived riskiness of each target served as the dependent variable.

Procedure

Participants were first presented with a university college identity prime to make this group membership salient. After completing the prime, participants were told that photos and details of fellow students were collected in a previous study and that those students had given permission for their details to be shown to other college students at their university. Participants were then presented with the battery of target faces, consisting of 12 targets that ranged in (pre-validated) trustworthiness. Target group membership was randomized, but counterbalanced such that all participants viewed six ingroup targets (e.g., living in a “college on campus”) and six outgroup targets (e.g., living in a “share house in [nearby suburb]”).

The 12 target faces were selected from the Karolinska Dataset (Lundqvist, Flykt, & Ohman, 1998; Oosterhof & Todorov, 2008). The battery consisted of an equal number of emotionally neutral male and female faces with low, medium and high pre-validated trustworthiness scores.³ The faces were matched on typicality and attractiveness in order to control for other variables that may impact on risk perceptions.⁴ Each target was randomly paired with fictional demographical information (i.e., living situation, age, year of study, and hobby; see Figure 8 for an example).

UQ Student #1	
Living Situation:	
UQ college on campus	
Age and Year of Study:	
18 years old, First Year	
Hobby:	
Singing	

Figure 8. Example manipulation of group membership (ingroup) and trustworthiness (high).

Materials and Measures

College identity prime. Participants' college identity was primed using a high-identification linguistic framing procedure (adapted from Banas, Cruwys, de Wit, Johnston, & Haslam, 2016; Greenaway, Haslam, Cruwys, Branscombe, Ysseldyk & Heldreth, 2015).

³ Faces were selected whose pre-validated trustworthiness scores corresponded to three levels: low trustworthiness was defined as having a z -score of < -0.6 ; medium trustworthiness as having a z -score between -0.2 and 0.2 ; high trustworthiness as having a z -score of > 0.6 .

⁴ To control for non-focal traits (e.g., face typicality, attractiveness; Oosterhof & Todorov, 2008; Todorov et al., 2015), a standard multiple regression analysis was run to determine Mahalobis distance scores for each face within the Karolinska Dataset that had a pre-validated trustworthiness score within the low, medium and high ranges specified above. The final battery was selected based on the following criteria: 1) the Mahalobis distance score was as close to zero (0) as possible, indicating minimal deviation from the mean across non-focal traits; and 2) the face looked young enough to plausibly be a college student (i.e., 17 – 22 years).

Participants were asked to indicate their agreement with five *very negative* statements (e.g., “I feel no affiliation with [college name]”) and five *moderately positive* statements (e.g., “[College name] has a welcoming atmosphere”) by selecting statements they agreed with and leaving blank statements they disagreed with. To cement the identity priming task, participants were asked to indicate the total number of positive and negative statements they had agreed with. This procedure is designed to make it easy to agree with positive statements about one’s group and disagree with negative statements about one’s group, thus increasing feelings of connection to the group.

Perceived trust. A single-item measure, adapted from Oosterhof and Todorov (2008), was used to assess perceived trust. Participants responded to the item “How trustworthy is this person?”, based on their first impression and “gut feeling”. Responses were recorded for each of the 12 targets on a scale ranging from 1, *not at all* to 9, *extremely*.

Risk perception. Perceptions of risk was assessed using a single-item measure “How risky would it feel to spend time with this person?” (similar to Peters, Burraston, & Mertz, 2004). Responses were recorded for each of the 12 targets using a slider scale, ranging from 0, *not at all risky* to 100, *extremely risky*.

Results

Analytic Strategy

Analyses were conducted using a mixed-effects moderated mediation model with the R *lme4* package (Bates et al., 2015) in combination with the R *mediation* package (Tingley, Yamamoto, Hirose, Keele, & Imai, 2014). This analytic approach was adopted because the study had a multilevel design in which the focal measures of the dependent variable (risk perception) and mediator (perceived trust) were at the level of the target, as were the independent variables (trustworthiness and group membership).

We conducted analyses predicting both the dependent variable (risk perception) and the mediator (perceived trust). The models included random intercepts for participants. This means that the model allowed participants to vary from one another in their average perceived trust and risk ratings, accounting for error variance associated with idiosyncratic differences on these ratings.⁵ First, we modelled the fixed effect of manipulated trustworthiness and group membership—akin to assessing the direct effects of these variables in a traditional regression analysis. Second, we added the interaction between manipulated trustworthiness and group membership. Third, for risk perception only, we added the mediator, perceived trust, to the model. Finally, we conducted moderated mediation analyses to assess the indirect effect of shared group membership on risk perception via perceived trust at high, medium, and low levels of manipulated trustworthiness.

Perceived Trust

Unexpectedly, shared group membership had no impact on perceived trust, $\beta = .07, p = .134$. However, speaking to a successful manipulation, the target trustworthiness manipulation significantly impacted perceived trust, such that low trust targets were perceived as less trustworthy than medium trust targets, $\beta = -.15, p = .009$, and high trust targets were perceived as more trustworthy than medium trust targets, $\beta = .51, p < .001$.

⁵ Normally in this design we would consider including a random intercept for stimuli, which would control for variance due to differences in unique faces across the targets. However, the *R* package for mediation analysis used in this study was limited, in that it only allowed for one random intercept when testing for an indirect effect. This proved unproblematic, however, because individual differences between participants explained more variance in ratings of trust compared to (non-focal) differences between targets. Therefore, we did not include a random intercept for stimuli but retained the random intercept for participants. The effects of the manipulations on perceived risk and perceived trust do not differ substantially when including a random intercept for stimuli as well as a random intercept for participants, with the exception that the trustworthiness manipulation becomes a non-significant predictor—unsurprising, given that this information is already accounted for in a model which includes a random intercept for target.

These effects were qualified by a significant interaction between shared group membership and the trust manipulation, $\chi^2(2) = 6.79, p = .034$. Inspection of the simple effects revealed that group membership had no impact on perceived trust when target trustworthiness was high, $\beta = -.11, p = .193$, but a significant impact on perceived trust when target trustworthiness was medium, $\beta = .16, p = .041$, and a marginal impact on perceived trust when target trustworthiness was low, $\beta = .15, p = .071$, such that ingroup members were trusted more than outgroup members.

Risk Perception

Consistent with H1, shared group membership had a marginally significant impact on risk perceptions $\beta = -.07, p = .082$, such that ingroup members were perceived as less risky than outgroup targets. The target trustworthiness manipulation significantly impacted risk perceptions, such that low trust targets were perceived as more risky than medium trust targets, $\beta = .13, p = .004$, and high trust targets were perceived as less risky than medium trust targets, $\beta = -.35, p < .001$.

There was no significant interaction between shared group membership and the trustworthiness manipulation, $\chi^2(2) = 3.46, p = .177$. However, inspection of the simple effects revealed that, although group membership had no impact on risk perceptions when target trustworthiness was high, $\beta = .03, p = .642$, and no impact on risk perceptions when target trustworthiness was low, $\beta = -.08, p = .217$, group membership did have a significant impact on risk perceptions when target trustworthiness was medium, $\beta = -.15, p = .030$, such that ingroup members were perceived as less risky than outgroup members.

Adding perceived trust as a predictor of risk perception was a significant improvement over the interaction model, $\chi^2(1) = 538.21, p < .001$. This revealed that perceived trust was a significant predictor of risk perceptions, $\beta = -.48, p < .001$, such that greater trust was associated with lower risk perceptions. Both interactions are depicted in

Figure 9.

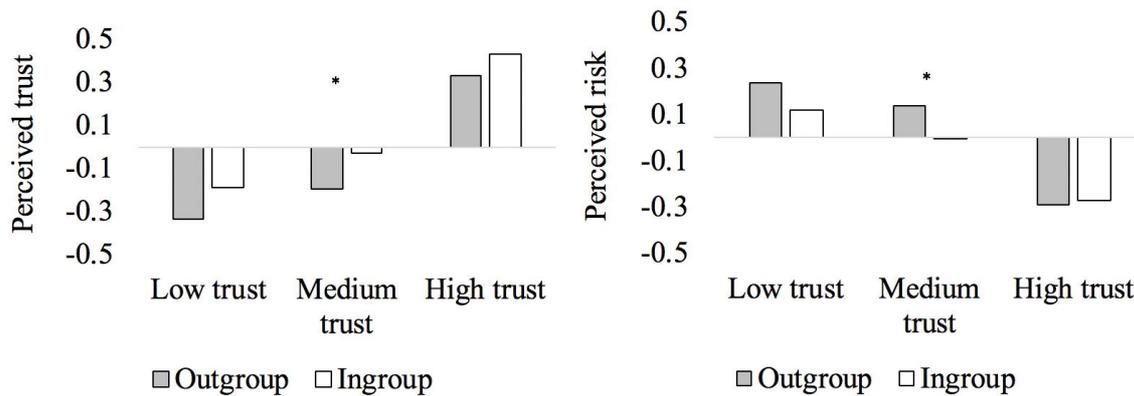


Figure 9. Effects of shared group membership and manipulated target trustworthiness on perceived trust in target (left panel) and perceived risk of target (right panel). *Note.* Asterisks indicate significant comparison between ingroup and outgroup.

Indirect Effects

The indirect effect of group membership on risk perceptions via perceived trust was significant at medium levels of target trustworthiness, $IE = -0.08$, $p = .03$, 95% $CI = -0.15$ to -0.01 , and marginally significant at low levels of target trustworthiness, $IE = -0.07$, $p = .062$, 95% $CI = -0.14$, <0.00 . These indirect effects indicate that, particularly for targets with medium cues to trustworthiness, ingroup members are trusted more, and it is through this process that ingroup members are perceived as less risky. The indirect effect of group membership on risk perceptions via perceived trust was non-significant at high levels of target trustworthiness, $IE = 0.05$, $p = .200$, 95% $CI = -0.02$, 0.12 .

Discussion

Study 6 provided causal evidence for the model by manipulating both group membership and trust (our proposed mediator). As hypothesized, increasing the apparent trustworthiness of targets eliminated the effect of shared group membership on perceived trust—because cues to trustworthiness were already available without the need to infer from the target's group membership. However, when target trustworthiness was medium, shared

group membership bolstered perceptions of trust and lowered perceptions of risk in ways we have seen in previous studies. Moreover, supporting H2, we observed a significant indirect effect whereby shared group membership lowered risk perceptions via increased trust, specifically when target trustworthiness was at medium levels.

It is interesting that we observed the strongest effects among targets who were validated as exhibiting a medium level of trustworthiness. Although we observed some marginal findings for targets who appeared low in trustworthiness, the most consistent findings were observed with targets in the medium range. We believe this is because at extreme ends of the spectrum (high and low), people already have strong cues to who is trustworthy (or not) and who poses the greatest risk (or not). In contrast, it is precisely when trustworthiness is ambiguous that people rely on cues such as shared group membership to guide their decisions about risk.

Study 7: Disease Risk Perceptions among Community Members

Study 7 returned to a disease context to examine the model, with two extensions on previous studies. First, Study 7 used a more extreme disease stimuli than that of Study 3, with participants considering a scenario in which they had been exposed to a contagious disease after shaking hands with a colleague. Second, Study 7 more forensically examined the mediating role of trust. This construct can be understood as comprising three interrelated subdimensions (Mayer, Davis & Schoorman, 1995): *ability* (the degree to which a person is capable of completing their obligations or promises effectively), *benevolence* (the degree to which a person is motivated to act in one's best interests), and *integrity* (the degree to which a person is reliable and honest). Although typically co-varying, these constructs are separable and there may be contexts in which one is of particular importance (Kim, Dirks, Cooper & Ferrin, 2006). In Study 7, we tested whether one or more of these subdimensions of trust was particularly important in explaining the relationship between shared group membership and

risk. However, we made no a priori predictions about which (if any) of the subdimensions would be most important.

In addition, Study 7 measured potential confounds that might plausibly account for the relationship between shared group membership and risk. One possibility is that the relationship between shared group membership and risk is attributable to a generalized halo effect for ingroup members, whereby they are perceived more positively in general, and that it is this (not trust specifically) that accounts for their effects on risk. A wealth of literature supports the ingroup favoritism effect (Voci, 2006). Study 7 therefore sought to establish whether ingroup favoritism, rather trust specifically, could better account for the relationship between shared group membership and risk. We tested for this potential effect by assessing general positive impressions of the target on three dimensions: warmth, competence, and morality (Goodwin, Piazza & Rozin, 2014; Leach, Ellemers & Barreto, 2007).

If ingroup favoritism is responsible for our effects, we should see ingroup members rated more positively than outgroup members on all dimensions, and it should be these dimensions—not trust—that mediate the effect of shared group membership on risk. If ingroup favoritism accounts for the phenomena, we would also expect the indirect effect via trust to disappear when these dimensions are entered into the model. On the other hand, if trust indeed is a primary driver of the risk effect, it should persist even when these dimensions are accounted for.

Method

Participants and Design

The study employed a two-level between-subjects design that manipulated group membership (ingroup vs. outgroup) based on political affiliation. Risk perception was the focal dependent variable and trust and disgust served as potential mediators. Ingroup favoritism was assessed as a potential alternative explanation.

The study recruited 410 Americans aged 19 to 87 years ($M = 36.45$, $SD = 11.38$) via Amazon's Mechanical Turk. Participants were 52% male, 47% female, and 1% other. Education was diverse but relatively high, most commonly a university degree (40%) or some university/community college (26%). Participants were 76% white, 11% black, 7% Asian, 9% Hispanic/Latino/a (this question allowed multiple responses).

After excluding participants who failed comprehension and manipulation checks (more details below), 368 people were included in the analyses. According to G*Power, 352 participants would provide .80 power to detect a small-medium between-groups effect size of $d = .30$ at the standard .05 alpha error probability.

Procedure

Participants were recruited for a study on 'Politics and workplace relationships' and received \$USD1.50 for the study, which took 7.65 minutes on average to complete ($SD = 4.82$ min). Participants indicated their preferred political party and then read a hypothetical scenario about a meeting with a colleague in a coffee shop. In the scenario, they shake hands with their colleague and then notice a rash on the colleague's hand. An image of a leprosy rash was provided, although no information was given about what disease the colleague might have or how contagious it might be. Participants were randomly assigned to either a condition that described the colleague as having the *same political affiliation* (ingroup condition, coded as 1) or having a *different political affiliation* (outgroup condition, coded as 0). Participants then completed the remaining measures and demographic questions.

Measures

Political group membership. Participants were first asked whether they identified most with Republicans or Democrats, identical to Study 2.

Risk perception. The same measure of risk perception from Study 3 was used ($\alpha = .88$).

Trust. Trust was measured using a three-factor scale developed by Mayer and Davis (1999). Minor adaptations were made to ask about the colleague in the scenario, rather than “top management” on a five point scale from “Strongly disagree” to “Strongly agree”. This scale measures the subdimensions of trust: *ability* (4 items; e.g. “I feel very confident about my colleague’s skills”, $\alpha = .88$), *benevolence* (5 items; “My colleague would not knowingly do anything to hurt me”, $\alpha = .89$) and *integrity* (6 items; “My colleague tries hard to be fair in dealings with others”, $\alpha = .85$). In addition to these three subdimensions, the scale also includes 4 additional items measuring general trust (e.g., “I really wish I had a good way to keep an eye on my colleague.” [reverse scored]), that were included when calculating a total score (which we refer to as *global trust*). Two items that could not be readily adapted to the context of this study were dropped, yielding an overall scale of 19 items (global trust scale $\alpha = .92$).⁶

Disgust. The same 12-item measure of disgust from our earlier studies was used.

Target perception. To assess ingroup favoritism, participants were asked to rate the target on 15 items in three subdomains: competence (e.g. “intelligent”, $\alpha = .95$), warmth (e.g., “kind”, $\alpha = .95$), and morality (e.g. “righteous”, $\alpha = .94$) on a five point scale from “Strongly disagree” to “Strongly agree” (Fiske et al., 2002; Leach et al., 2007). The items could also be combined into a single measure of target perception ($\alpha = .98$).

Comprehension and manipulation checks. Following the scenario, participants were asked an open-ended question “So that we know you understand this scenario, please

⁶In addition to this trust measure, we included the same four-item measure of trust used in our other studies ($\alpha = .91$) and a seven-item general trust measure from Schoorman and Ballinger (2006; see also Schoorman, Mayer & Davis, 2007, e.g., “My colleague keeps my interests in mind when making decisions”; $\alpha = .64$). However, given our theoretical interest in the subdimensions of trust, we focus on the three-dimensional measure in the results below. Results for H2 were replicated if these trust measures were used ($ps < .031$).

provide a one sentence summary of the situation that is described above.” Suspicious responses to this question (e.g. “Very nice survey”) were identified and participants were excluded on this basis. A manipulation check was included at the end of the study which asked participants “A colleague in this scenario was described as:” with response options of: *A Republican, A Democrat, or I don’t remember*. Participants who did not answer correctly were excluded from analyses.

Results

A total of 42 people (10% of respondents) were excluded who failed either the manipulation check or comprehension check. The final sample thus comprised 368 people.

A series of one-way ANOVAs were applied to the outcome variables, the results of which are summarized in Table 1. This revealed no significant effect of shared group membership on risk perception, thus not supporting H1. There was a significant effect of shared group membership on the global trust scale such that ingroup members were trusted more than outgroup members. Further inspection of the trust subdimensions revealed that this effect was primarily driven by the integrity subscale. There was no significant effect on disgust. Moreover, there was no significant effect of shared group membership on general target perceptions, although inspection of the subscales revealed ingroup members were rated as significantly more warm than outgroup members.

Table 1. *Effect of shared group membership on outcome variables in Study 7.*

	Omnibus Effect			Ingroup		Outgroup	
	<i>F</i>	<i>p</i>	η_p^2	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Risk Perception	0.55	.460	<.01	56.09	26.15	58.10	26.12
Global Trust Scale	5.18	.023	.01	3.10	0.68	2.94	0.69
Ability subscale	0.28	.600	<.01	3.38	0.74	3.34	0.79
Benevolence subscale	1.45	.229	<.01	2.96	0.82	2.85	0.88

Integrity subscale	12.65	<.001	.03	3.23	0.77	2.95	0.73
Disgust	0.14	.707	<.01	2.90	1.12	2.95	1.04
Target Perceptions	1.75	.187	.01	3.25	0.87	3.13	0.84
Warmth subscale	5.82	.016	.02	3.34	0.92	3.11	0.93
Competence subscale	0.06	.810	<.01	3.27	0.87	3.24	0.83
Morality subscale	1.06	.303	<.01	3.14	0.94	3.04	0.92

Note. Significant effects in bold. Target perception represents the average of warmth, competence, and morality scales.

Indirect Effects

Only two variables met the conditions required to act as a potential mediator (i.e., being associated with both the independent and dependent variables)—the integrity subdimension of trust and warmth perceptions. We therefore entered these variables as simultaneous mediators in Hayes' (2018) PROCESS Model 4 with 10,000 bootstrapping resamples to assess the indirect effect of shared group membership on risk perception. Supporting H2, this revealed a significant indirect effect of shared group membership on risk via integrity-based trust ($IE = -2.07$, $SE = 0.99$, 95% $CI = -4.67$ to -0.56). However, the indirect effect of shared group membership on risk via warmth perceptions was not significant ($IE = -0.81$, $SE = 0.68$, 95% $CI = -2.66$ to 0.11).

For the sake of thoroughness, we also tested for indirect effects of shared group membership via the other trust subdimensions, disgust, and the other target perception subscales. None of these indirect effects was significant. Consistent with H2, there was a significant indirect effect of shared group membership on risk perception via the global trust scale ($IE = -2.32$, $SE = 1.02$, 95% $CI = -4.43$ to -0.41) that persisted even when controlling for warmth perceptions ($IE = -2.49$, $SE = 1.21$, 95% $CI = -5.31$ to -0.49).

Discussion

Unlike the other studies, Study 7 did not find support for H1. However, Study 7 did support H2, and was able to provide insight into the question of whether any of the subdimensions of trust is particularly implicated in the link between shared group membership and risk. This revealed that *integrity*—the degree to which a person is perceived to be reliable, diligent, and honest—was the only trust subdimension that mediated this relationship. Interestingly, integrity is the subdimension of trust that has received the most empirical attention and perhaps most closely corresponds to the way in which trust is typically conceptualized (Dietz & Den Hartog, 2006). As in prior studies, Study 7 found no evidence of mediation via disgust.

Study 7 also assessed several target perceptions to determine whether the effect of shared group membership on risk could be explained by ingroup favoritism more generally. Ingroup members were indeed rated as more warm than outgroup members, but this did not mediate the effect of shared group membership on risk, nor did it eliminate the established indirect effect via integrity-based trust or global trust (H2). This increases our confidence that the effects we have observed to date are attributable to trust specifically.

Study 8: Financial Risk Behavior among Community Members

Our final study aimed to conceptually replicate the findings of Study 7 using a behavioral measure of risk. To do this, we again turned to the BART task to assess adaptive and maladaptive risk-taking with real financial consequences. Study 8 therefore used a similar design to Study 2, but removed the risk manipulation, such that all participants were presented with ingroup and outgroup balloons that had the same average explosion point. Moreover, we no longer told participants that the balloon set points had been selected by ingroup or outgroup members. Study 8 simply displayed ingroup and outgroup logos on the balloons with no additional explanation for why these were present. This represents a

conservative test of our hypothesis, which we were keen to perform to test the potential limits of the effect. Nevertheless, we still expected shared group membership to inspire greater risk-taking (H1). Based on the findings of Study 7, we also expected this effect to operate indirectly via trust (H2), particularly through elevated perceptions of ingroup integrity.

Method

Participants and Design

Participants were adult members of the community in the United States recruited via Amazon's Mechanical Turk. Participants were 53% male and 47% female and ranged in age from 18 to 76 years ($M = 35.70$, $SD = 11.42$). Ethnicity was relatively diverse, including 64.9% white, 16.7% black, 9.1% Hispanic/Latino/a; 10.9% Asian; 1.1% Native American (multiple ethnic categories could be selected).

Procedure

Participants were offered a minimum \$1.00 incentive for taking part in the study on Politics and Social Relationships (which took 14 minutes on average), plus their winnings in the balloon game ($M = \$3.08$; $SD = \$0.93$). The procedure and instructions were largely the same as for Study 2. However, we removed most information about the balloon types, leaving only the logos of the Republican and Democratic political parties. In Study 8, participants were *not* told that the balloons had different average explosion points, nor anything about who had set the explosion points for the balloons. They did not complete the task involving setting the explosion point for future participants as in Study 2. Both ingroup and outgroup balloons had an average explosion point of 64 pumps.

Measures

Risk behavior. The measures of adaptive and maladaptive risk taking were calculated in the same way as in Study 2. No transformation of adaptive risk taking was required because the average explosion points was the same across balloon types.

Trust. Trust was measured as in Study 7. Prior to completing the balloon game, participants completed this scale for both the ingroup and the outgroup.

Disgust. The same 12-item measure of disgust was included as in Studies 3, 5 and 7. Prior to completing the balloon game, participants completed this scale for both the ingroup and the outgroup.

Comprehension check. After reading the instructions for the balloon game, participants were asked to describe the two types of balloon they would see in the study. This was designed to detect suspicious responding. For instance, a correct response was “Democrat and Republican”, while an incorrect response was “INFLATE BALLOON”.

Results

There were 302 complete responses. Of these 18 were excluded who failed the comprehension check. An additional 9 people were excluded who were suspicious responders. As in Study 2, these outliers engaged in patterns of responding such as inflating each balloon exactly once. This left a total of 276 responses available for analysis.

A series of repeated-measures ANOVAs were applied to the outcome variables, the results of which are summarized in Table 2. This revealed significant effects of shared group membership on all variables. Supporting H1, ingroup members inspired greater adaptive and maladaptive risk taking than outgroup members, meaning that people both made more money per non-exploded ingroup balloon and exploded more ingroup balloons.

There were significant effects on all trust scales such that ingroup members were trusted more than outgroup members. Moreover, ingroup members inspired less disgust than outgroup members.

Indirect Effects

The indirect effects of shared group membership on risk taking via trust (H2) and disgust (H3) were examined using the MEMORE macro Model 1 with 10,000 bootstrap

resamples (Montoya & Hayes, 2017). This model examines whether the within subjects independent variable (group membership) predicts differences in the mediators (trust; disgust) for ratings of the ingroup versus the outgroup, and whether this in turn predicts differences on the dependent variable (adaptive and maladaptive risk behavior) for the ingroup versus the outgroup.

Results of the indirect effect analyses are presented in Table 3. These analyses found that all hypothesized mediators—trust subdimensions of integrity, ability, and benevolence, and disgust—significantly mediated the relationship between shared group membership and adaptive risk taking. However, for maladaptive risk taking, the indirect effect of shared group membership was only significant via two subdimensions of trust: integrity and benevolence. Therefore, H2 was supported and H3 partially supported.

Table 2. *Effect of shared group membership on outcome variables in Study 8.*

	Omnibus Effect			Ingroup		Outgroup	
	<i>F</i>	<i>p</i>	η_p^2	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Risk Behavior							
Adaptive risk taking	5.37	.021	.02	15.35	7.22	14.92	6.85
Maladaptive risk taking	4.84	.029	.02	4.11	2.20	3.93	2.06
Global Trust Scale							
Ability subscale	280.97	<.001	.51	3.75	0.79	2.52	0.95
Benevolence subscale	458.46	<.001	.63	3.54	0.88	1.98	0.95
Integrity subscale	400.39	<.001	.59	3.66	0.70	2.21	0.86
Disgust	252.02	<.001	.48	1.58	0.70	2.69	1.12

Table 3. *Indirect effects of shared group membership on risk behavior via trust and disgust.*

	Adaptive Risk Taking	Maladaptive Risk Taking

	<i>IE</i>	<i>SE</i>	95% <i>CI</i>	<i>IE</i>	<i>SE</i>	95% <i>CI</i>
Global Trust Scale	0.15	0.07	0.03, 0.30	0.06	0.03	0.00, 0.12
Ability subscale	0.11	0.06	0.02, 0.24	0.04	0.03	-0.01, 0.09
Benevolence subscale	0.17	0.08	0.04, 0.34	0.07	0.04	0.01, 0.14
Integrity subscale	0.15	0.07	0.03, 0.29	0.06	0.03	0.01, 0.13
Disgust	-0.14	0.06	-0.27, -0.04	-0.03	0.03	-0.09, 0.01

Note. *IE*=indirect effect; *SE*=standard error; *CI*=confidence intervals. Significant effects in bold.

Discussion

With a large sample, a within-subjects design, and a behavioral dependent variable, Study 8 provides strong evidence for the model. We deliberately designed this study to be a conservative test of the hypothesis, given that there was no explanation given to participants that might explain how the group membership displayed on balloons could affect their winnings in the game. Nevertheless, the findings of Study 8 supported H1 and replicated those of Study 2, such that participants took more risks on ingroup balloons, resulting both in more secured winnings (per non-exploded balloon) and in more lost winnings (due to balloon explosions).

Study 8 also provides further evidence that the effect of shared group membership on risk taking is not limited to (either) risk taking that is adaptive or maladaptive, but tends to increase both forms of risk behavior. Thus, risks taken in the context of shared group membership has the potential to pay off and to fail. Indeed, one of the hallmarks of risk taking is that outcomes are uncertain—if they were known, there would be no risk (Das & Teng, 2004; Evans & Krueger, 2011). What our findings show consistently is that risks taken with ingroup members raise these stakes compared to those with outgroup members by virtue of the risk discounting that occurs with someone with whom we share group membership.

As in Study 7, we found evidence that integrity-based trust consistently mediated the relationship between shared group membership and both risk behaviors. In addition, in this study, we found that benevolence-based trust mediated the effect of shared group membership for both risk behaviors. We also found that ability-based trust and disgust mediated the relationship between shared group membership and adaptive risk taking only. This is largely consistent with H2 and H3, although considered together the findings from Studies 7 and 8 suggest that the integrity subdimension of trust may be particularly important for the ingroup risk effect to manifest. That said, findings across Studies 3 to 8 we found that a global measure of trust was a consistent mechanism, and indeed integrity is the most prototypical component of trust (Dietz & Den Hartog, 2006). This body of work provides strong evidence that trust, globally construed, is an important mechanism through which shared group membership impacts on risk.

Meta-Analysis

We conducted eight studies that together provide support for the Social Identity Model of Risk Taking (see Table 4). While we found overall consistent support for the direct effect of shared group membership on risk perceptions and behavior (H1) and for the indirect effect of shared group membership on risk outcomes via trust (H2), we found limited evidence for the indirect effect of shared group membership on risk outcomes via disgust (H3). We therefore conducted a meta-analysis on the effects of shared group membership on all outcome variables (trust, disgust, risk) across all thirteen studies conducted in this line of work (8 studies in the main paper and 5 in supplementary materials; total $N = 4,708$). In addition to determining which effects are robust to changes in context and design, this approach provides an indication of the average effect size for each relationship. The results of this meta-analysis are displayed in Table 5.

We sought to be inclusive in the meta-analysis, adopting the spirit of including all studies that were designed for the purpose of addressing our main research question. However, given that some these studies were ultimately removed from the manuscript due to suboptimal designs, we also report the results of the meta-analysis across only studies included in the paper (Studies 1–8; $N = 3,228$). As seen in Table 5, results are largely similar across both analytic strategies.

We conducted the meta-analysis using the R *metafor* package (Viechtbauer, 2010) with a random effects model, transforming the effect sizes first into raw correlation coefficients and then using the *r*-to-*z* transformation. Heterogeneity statistics were significant for all tests ($Qs > 32.96$, $ps < .001$) except the relationship between shared group membership and risk ($Q = 12.35$, $p = .418$), making it advisable to use random effects models. For all studies we focused on the main effect of shared group membership on the outcome variables, collapsing across any other manipulations that were included in individual studies. For studies with within-subjects designs we calculated within-subjects *d* for the effect size of the relationship between shared group membership and the outcome variable of interest before converting this to effect size *r*. We also reverse-scored effects in studies that assessed risk perceptions so that positive scores on the outcome measures equate to greater risk taking.

The meta-analysis revealed an overall small but significant association between shared group membership and risk ($r_z = .10$, $p < .001$, 95% $CI = 0.07, 0.13$). The forest plot of these effects is displayed in Figure 10 and forest plots for other effects are available on the OSF at this [link](#). There was a moderate and significant association between shared group membership and trust ($r_z = .25$, $p < .001$, 95% $CI = 0.11, 0.39$), such that ingroup members were trusted more and inspired greater risk.

Somewhat surprisingly given the results of individual studies, there was a small but significant association between shared group membership and disgust ($r_z = -.14, p = .020, 95\% CI = -0.26, -0.02$), such that ingroup members were overall rated as less disgusting than outgroup members. Consistent with prior work highlighting trust and disgust as key drivers of risk, we found large and significant associations between trust and risk ($r_z = .30, p < .001, 95\% CI = 0.23, 0.38$) and disgust and risk ($r_z = -.48, p < .001, 95\% CI = -0.74, -0.22$).

Table 4. *Summary of the studies reported in the main paper and supplementary materials.*

<i>Main Paper</i>								
Study	<i>N</i>	Sample	Context	DV	Design	H1	H2	H3
1	123	Undergraduate psychology students	Disease risk	Risk perception	Experimental	✓	-	-
2	222	Community members	Financial risk	Risk behavior	Experimental	✓	-	-
3	585	Global sample (40 nationalities)	Disease risk	Risk perception	Experimental	✓	✓	✗
4	1387	School graduation festival attendees	Health risk	Risk perception	Correlational	✓	✓	-
5	149	Art festival attendees	Health risk	Risk behavior	Longitudinal	✓	✓	✗
6	118	Undergraduate college residents	General risk	Risk perception	Experimental	✓	✓	-
7	368	Community members	Disease risk	Risk perception	Experimental	✗	✓	✗
8	276	Community members	Financial risk	Risk behavior	Experimental	✓	✓	✓
<i>Supplementary Materials</i>								
Study	<i>N</i>	Sample	Context	DV	Design	H1	H2	H3
9	91	Undergraduate psychology students	Disease risk	Risk perception	Experimental	✗	-	-
10	103	Undergraduate psychology students	Corporate risk	Risk behavior	Experimental	✓	-	-
11	157	Community members (pre-registered)	Disease risk	Risk perception	Experimental	✓	✓	✗
12	373	Community members (pre-registered)	Financial risk	Risk behavior	Experimental	✗	✓	-
13	756	School graduation festival attendees	Health risk	Risk perception	Correlational	✓	✓	-

Note. ✓ = Hypothesis supported; ✗ = Hypothesis not supported; - = Hypothesis not tested

H1 = effect of shared group membership on risk perceptions and behavior.

H2 = indirect effect of shared group membership on risk via trust.

H3 = indirect effect of shared group membership on risk via disgust.

Table 5. Results and meta-analysis of all studies.

Study	N	Shared group membership and risk	Shared group membership and trust	Shared group membership and disgust	Trust and risk	Disgust and risk
		<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>	<i>r</i>
1	123	.19*	-.12	-.01	.22*	-.41***
2	222	.05*	-	-	-	-
3	585	.09*	.04	-.07	.25***	-.72***
4	1387	.13***	.51***	-	.26***	-
5	149	.17*	.44***	-.14	.18*	.08
6	118	.03	.02	-	.55***	-
7	368	.04	.12*	-.02	.35***	-.67***
8	276	.14*	.53***	-.43***	-	-
9	91	.06	-	-	-	-
10	103	.08	-	-	-	-
11	157	.19*	.21*	-.16	.38***	-.67***
12	373	-.01	.14**	-	.45***	-
13	756	.13**	.53***	-	.15***	-
<i>Meta-analysis of results across the studies</i>						
		Shared group membership and risk	Shared group membership and trust	Shared group membership and disgust	Trust and risk	Disgust and risk
		<i>r</i> (CI)	<i>r</i> (CI)	<i>r</i> (CI)	<i>r</i> (CI)	<i>r</i> (CI)
S1–13	4,708	.10 (.07, .13)	.25 (.11, .39)	-.14 (-.26, -.22)	.30 (.23, .38)	-.48 (-.74, -.22)
S1–8	3,228	.11 (.07, .14)	.23 (.04, .41)	-.13 (-.27, .00)	.29 (.23, .35)	-.44 (-.75, -.13)

Note. “Risk” refers to greater risk behavior (Studies 2, 5, 8, 10, and 12) and reverse-scored risk perception (Studies 1, 3, 4, 6, 7, 9, 11, and 13). Studies above the dotted line reported in main paper; studies below the dotted line reported in supplementary materials; methods and results for supplementary studies are available on the OSF at this [link](#). Shared group membership scored as 1 = ingroup, 0 = outgroup for experimental studies. Trust results in Studies 7 and 8 reported as global trust scale (not subdimension ability, benevolence, and integrity scales). *r* = effect size. *CI* = 95% confidence intervals. - indicates relationship was not tested in a given study. * $p < .05$. ** $p < .01$. *** $p < .001$

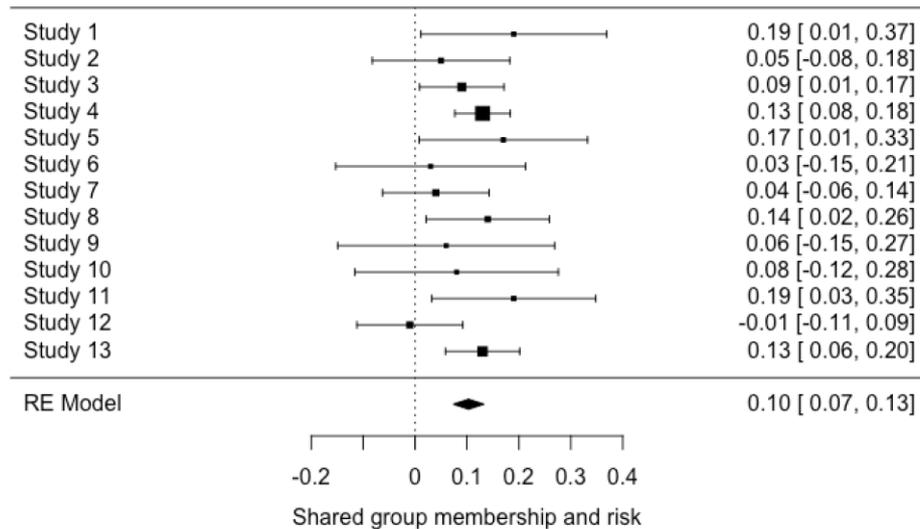


Figure 10. Forest plot for the relationship between shared group membership and risk

General Discussion

The present research reports eight studies that tested a novel Social Identity Model of Risk Taking. These studies were conducted in a range of risk contexts, including disease risk, financial risk, health, and physical risk. The studies provided evidence for the model across correlational, longitudinal, and experimental designs in real groups in the field as well as minimal groups created in the lab. Despite the methodological and contextual diversity across the studies, a meta-analysis revealed consistent effects of shared group membership on greater risk, supporting the first hypothesis (H1). Moreover, we found consistent evidence that shared group membership increased risk via increased trust—ingroup members inspire greater risk because they are trusted more than outgroup members—supporting the second hypothesis (H2). However, we found less consistent evidence for a separate pathway via disgust, contrary to the third hypothesis (H3).

Although we replicated the finding that state disgust is a good predictor of risk avoidance (consistent with e.g., Woody & Tolin, 2002), only one of five studies supported a mediation pathway via disgust. The limited evidence for this mediation pathway was due to a weaker than expected effect of shared group membership on state disgust. However, given that the meta-analysis found evidence for a significant relationship when the samples were

combined across studies, it would be premature to conclude that this pathway does not exist. Instead, we suggest that future research seek to elucidate the conditions under which this pathway might become apparent. For instance, individual differences such as disgust sensitivity or contextual factors such as disease salience may be relevant.

Although robust, the direct effect of shared group membership on risk was modest in size. In part, this is likely because we deliberately sought to design conservative tests of our hypotheses (e.g. by removing explicit information about shared group membership in Study 8), and so we might expect this effect size estimate to be at the lower end of what would occur in non-controlled settings. The fact that the effect size estimates were larger for the field studies supports this idea. There may also be important moderators of the effect, as suggested by different studies. In particular, Study 1 suggests that these theoretical principles apply particularly when risk is contextually salient. Furthermore, Studies 3 to 5 suggest that the *subjective importance* of group membership is also important—we would not expect to see an effect of shared group membership on risk for groups that do not meaningfully inform one’s social identity. Political orientation may also be important in moderating this effect. Prior studies have revealed ideological differences in disgust sensitivity, trust, risk aversion and threat sensitivity (Jost, 2017; Jost et al., 2018); all constructs that are relevant to our findings. Finally, one possibility we have not tested is that causality also flows in the opposite direction, at least in some cases. A bidirectional relationship between risk perception and social identity seems possible, whereby risk taking is also a means of signaling affiliation with others (e.g., Cruwys et al., 2018). Indeed, this is the basis of many predictions made by work on the behavioral immune system (Fincher & Thornhill, 2008).

Theoretical and Practical Implications

This research contributes to a broader body of work that demonstrates the fundamental importance of social groups in shaping perception, emotion, and behavior.

Indeed, our studies demonstrate that the social determinants of risk perception and risk behavior—they are influenced by our affiliative networks and categorization of others as part of the “self”. The minimal group study in particular provided evidence for the causal role of shared group membership in determining risk perception and behavior. Subsequent studies demonstrated that this effect on risk is not limited to a particular domain, and instead has implications for risk in a variety of domains that may apply to a range of fields beyond psychology (e.g., economics, public health, and medicine).

Of course, greater risk taking is not inherently bad, and may sometimes have upsides. This was demonstrated most clearly in Studies 2 and 8 that used the well-known BART to measure adaptive forms of risk taking that yielded benefits as well as maladaptive forms of risk taking that accrued costs. We found that shared group membership encouraged both forms of risk taking, with ingroup members inspiring actions that resulted in the potential for reward as well as loss. Uncertainty is inherent to risk, and ingroup members appear to make people more willing to roll the dice in the face of uncertainty than outgroup members. Our findings consistently show this is because ingroup members are trusted more than outgroup members, and it is this trust that inspires the leap of faith into unknown territory.

This research provides further support for the central role of trust as a social inference that has implications for perception and behavior. Researchers have tended to see trust as the ‘grease’ that enables effective social functioning (Cooper, 2008; Gilson, 2003; Helliwell & Wang, 2010). While we do not dispute the many positive consequences of trust, and indeed found some benefits in this project, we also identified a downside. Specifically, trust can be misplaced, or can lead to unforeseen consequences. Most pertinent to our investigation, trust appears to encourage people to engage in risk taking; thus exposing themselves to potential harm they may otherwise have avoided.

In addition to extending social identity principles into the relatively uncharted risk space, our findings help to further develop the established literature on shared identity and trust. This literature shows robustly that ingroup members inspire greater trust than outgroup members (e.g., Brewer, 2008; Platow et al., 2012; Tanis & Postmes, 2005). What has been relatively unexplored in this work is the precise *nature* of the trust that is inspired by ingroup members. Drawing on work in the organizational psychology literature that identifies three ‘pillars’ of trust, we found some evidence that shared group membership increased all three subdimensions of trust, including the perception of that ingroup members are capable (ability-based trust), caring (benevolence-based trust), and principled (integrity-based trust). This helps to provide insight into why a person’s willingness to take risks with ingroup members versus outgroup members may be dependent on context; a possibility that future research could investigate in targeted empirical work. For example, in situations where another individual could exploit a target (e.g., in the case of Ponzi schemes), risk-taking may be enhanced by the individual’s perceived benevolence (see also Blois & Ryan, 2013). In situations where another party may be a source of risk (e.g., in the case of sexually transmitted disease), risk-taking may be enhanced by that person’s perceived integrity. And in situations where a hazard is communicated as harmless by another individual (e.g., in the case of binge drinking), risk-taking may be enhanced by the perception that the other individual is competent and well informed. These possibilities await further investigation.

Despite some evidence that all three forms of trust were higher with ingroup members than outgroup members in Study 8, we found that only one of these forms of trust was a consistent mediator of the link between shared group membership and risk across two studies. Specifically, ingroup members inspired greater risk because they are considered more reliable and honest (and, at least in one study, because they are considered more benevolent). In addition to further probing the foundations and functions of trust within and between

social groups, this observation accords with prior findings that the integrity component of trust is of particular importance when predicting perceptions and behavior (e.g., Kim, Dirks, Cooper & Ferrin, 2006). Interestingly, this finding also accords with previous evidence that people only trust ingroup members when there is *mutual knowledge* of their shared group membership (Platow et al., 2012). Although such mutual knowledge was not present in all of the studies presented here, it is likely an important moderator that increases the potency of these effects.

The Social Identity Model of Risk Taking also provides a new perspective on the phenomena collectively referred to as the behavioral immune system. Specifically, this body of work has identified a positive relationship between shared group membership (or proxies such as cohesion, collectivism, and sociality) and disease risk (Fincher, Thornhill, Murray & Schaller, 2008; Fincher & Thornhill, 2012). However, the theoretical interpretation of this evidence has typically been in the opposite casual direction—that exposure to disease risk strengthens social bonds (e.g., Moon, Krems, & Cohen, 2018; Schaller & Park, 2011). A social identity perspective, however, conceptualizes social categorization as the more fundamental psychological capability (Turner & Oakes, 1997) and it follows that disgust might have been *co-opted* to promote outgroup avoidance and ingroup cohesion (see also Fessler, Clark & Clint, 2015). The experimental evidence presented here supports this interpretation, although we acknowledge too that the relationship between risk and shared group membership is likely to be bidirectional.

Our findings also have potential to contribute to applied efforts to better manage risk. Policy outcomes based on risk taking research are often less effective than anticipated (Zinn, 2017). For example, policy interventions that attempt to reduce unsafe sex, prevent binge drinking, limit drug use, and increase disaster preparedness have demonstrated only limited success (Miller & Prentice, 2016; Nutbeam, 2000; Pan & Bai, 2009). These disappointing

outcomes may be, in part, because of limited attention to the *social reality* in which these risky actions take place. This research suggests that interventions to reduce risk taking cannot afford to neglect its social determinants. This is an important message, given that interventions typically emphasize individual risk factors like impulsivity, or are purely educational in content. Moreover, when social factors have been acknowledged in interventions, group influences are typically conceptualized in a purely negative fashion—exemplified by the famously ineffective yet stubbornly popular Drug Abuse Resistance Education program, which encourages youth to “just say no” to peer pressure (Pan & Bai, 2009). By contrast, the social identity approach (and the Social Identity Model of Risk Taking in particular) provides the theoretical lens to go beyond these traditional approaches, illustrating how social context and social relationships can impact on individual behavior. One of the benefits of utilizing a comprehensive model of group processes is that it specifies ways in which the power of groups might also be harnessed for positive solutions.

Future research could develop and test interventions based on this model by, for instance, utilizing messaging from ingroup members about strategies to keep fellow ingroup members safe. Educating people about the risks posed by potential ingroup members may also help people adjust against a natural inclination to discount risks with these individuals, for example by highlighting the fact that harassment and physical and sexual violence are statistically more likely to be perpetrated by someone close to us than a stranger. Considered another way, it could sometimes be beneficial to enhance a sense of shared identity before encouraging people to engage in behaviors that to some seem socially or physically risky, such as getting vaccinated or speaking out against injustice.

Strengths and limitations. This research program had many strengths, including the diverse contexts in which the studies were conducted. In particular, we intentionally sought to investigate different kinds of risk, from binge drinking to the spread of disease, and from

financial decision-making to public humiliation. The research program also included controlled laboratory experiments and large field studies, as well as self-report and behavioral outcome variables. These strengths bolster our confidence that (1) these relationships are causal, (2) these phenomena are generalizable across multiple domains, and (3) psychological processes have a meaningful impact on consequential outcomes.

However, like all research these studies are not without limitation. In particular, although the samples were diverse in age and were not predominantly drawn from college students, the participants were overwhelmingly from Western, Educated, Industrialized, Rich Democracies, with only one of eight studies (Study 3) being culturally diverse. For this reason, the model should not be generalized beyond this cultural context (see Henrich et al., 2010) without additional confirmatory evidence.

Conclusions

A young person at a party accepts a ride home from a friend who has had a few drinks. A woman at a festival allows her lipstick to be used by a stranger in the bathroom. An elderly man sends money to a pen pal who has asked for financial support. These are all examples of risky behaviors that may cause harm (injury, disease, financial loss). Yet, they are also behaviors enacted with someone to whom we feel close: a friend; a person with shared interests; a trusted confidante. The present research offers a novel theoretical lens for understanding why the people we trust the most sometimes pose the greatest risk. While sometimes a force for positive social relations and individual benefit, shared group membership can have unforeseen negative consequences in promoting risk taking because, all else being equal, we trust ingroup members more than we trust outgroup members.

References

- Abrams, D., Wetherell, M. S., Cochrane, S., Hogg, M. A., & Turner, J. C. (1990). Knowing what to think by knowing who you are: Self-categorization and the nature of norm formation, conformity and group polarization. *British Journal of Social Psychology, 29*, 97–119.
- Alnabulsi, H., & Drury, J. (2014). Social identification moderates the effect of crowd density on safety at the Hajj. *Proceedings of the National Academy of Sciences of the United States of America, 111*(25), 9091–9096. <http://doi.org/10.1073/pnas.1404953111>
- Banas, K., Cruwys, T., de Wit, J. B. F., Johnston, M., & Haslam, S. A. (2016). When group members go against the grain: An ironic interactive effect of group identification and normative content on healthy eating. *Appetite, 105*, 344–355. <http://doi.org/10.1016/j.appet.2016.06.005>
- Bates, B. R., & Chadwick, A. E. (2015). Measuring state disgust: Initial evidence for the reliability and validity of an English-language version of the Ekel-State-Fragebogen. *Communication Research Reports, 32*(1), 73–82. <http://doi.org/10.1080/08824096.2014.989978>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1-48. <http://doi.org/10.18637/jss.v067.i01>
- Baumeister, R. F., Vohs, K. D., & Funder, D. C. (2007). Psychology as the science of self-reports and finger movements: Whatever happened to actual behavior? *Perspectives on Psychological Science, 2*(4), 396–403. <http://doi.org/10.1111/j.1745-6916.2007.00051.x>
- Beck, U. (1992). *Risk Society: Towards a New Modernity*. London: Sage.
- Blois, K. & Ryan, A. (2013). Affinity fraud and trust within financial markets. *Journal of Financial Crime, 20*(2), 186-202. <https://doi.org/10.1108/13590791311322364>

- Bodimeade, H., Anderson, E., La Macchia, S., Smith, J. R., Terry, D. J., & Louis, W. R. (2014). Testing the direct, indirect, and interactive roles of referent group injunctive and descriptive norms for sun protection in relation to the theory of planned behavior. *Journal of Applied Social Psychology, 44*(11), 739–750.
<http://doi.org/10.1111/jasp.12264>
- Botdorf, M., Rosenbaum, G. M., Patrianakos, J., Steinberg, L., & Chein, J. M. (2017). Adolescent risk-taking is predicted by individual differences in cognitive control over emotional, but not non-emotional, response conflict. *Cognition and Emotion, 31*(5), 972–979. <http://doi.org/10.1080/02699931.2016.1168285>
- Boyd, R., & Richerson, P. J. (2009). Culture and the evolution of human cooperation. *Philosophical Transactions of the Royal Society B: Biological Sciences, 364*(1533), 3281–3288. <https://doi.org/10.1098/rstb.2009.0134>
- Brewer, M. B. (2008). Depersonalized trust and ingroup cooperation. In J. I. Krueger (Ed.), *Rationality and social responsibility: Essays in honor of Robyn Mason Dawes. Modern pioneers in psychological science: An APS-Psychology Press series* (pp. 215–232). New York, NY: Psychology Press.
- Brewer, N. T., Chapman, G. B., Gibbons, F. X., Gerrard, M., McCaul, K. D., & Weinstein, N. D. (2007). Meta-analysis of the relationship between risk perception and health behavior: The example of vaccination. *Health Psychology, 26*(2), 136–145.
<http://doi.org/10.1037/0278-6133.26.2.136>
- Campbell, D. T. (1958). Common fate, similarity, and other indices of the status of aggregates of persons as social entities. *Behavioral Science, 3*, 14–25.
<https://doi.org/10.1002/bs.3830030103>
- Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior and Organization, 83*(1), 50–58.

<http://doi.org/10.1016/j.jebo.2011.06.007>

Cole, J. (2016). Antimicrobial resistance - a “rising tide” of national (and international) risk.

Journal of Hospital Infection, 92(1), 3–4. <http://doi.org/10.1016/j.jhin.2015.10.005>

Cooper, L. A. (2008). Understanding concordance in patient-physician relationships:

Personal and ethnic. *Annals Of Family Medicine*, 198–205.

<http://doi.org/10.1370/afm.821>.Department

Cruwys, T., Haslam, S. A., Dingle, G. A., Jetten, J., Hornsey, M. J., Chong, D. E. M., & Oei,

T. P. S. (2014). Feeling connected again: Interventions that increase social identification reduce depression symptoms in community and clinical settings. *Journal of Affective Disorders*. <http://doi.org/10.1016/j.jad.2014.02.019>

<http://doi.org/10.1016/j.jad.2014.02.019>

Cruwys, T., Platow, M. J., Angullia, S. A., Chang, J. M., Diler, S. E., Kirchner, J. L., ...

Wadley, A. L. (2012). Modeling of food intake is moderated by salient psychological group membership. *Appetite*, 58(2), 754–757. <http://doi.org/10.1016/j.appet.2011.12.002>

Cruwys, T., Saeri, A. K., Radke, H. R. M., Walter, Z. C., Crimston, D., & Ferris, L. J. (2018).

Risk and protective factors for mental health at a youth mass gathering. *European Child and Adolescent Psychiatry*, 28(2), 211–222. <http://doi.org/10.1007/s00787-018-1163-7>

Curtis, V., de Barra, M., & Auger, R. (2011). Disgust as an adaptive system for disease

avoidance behaviour. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 366(1563), 389–401. <http://doi.org/10.1098/rstb.2010.0117>

Das, T. K., & Teng, B. S. (2004). The risk-based view of trust: A conceptual framework.

Journal of Business and Psychology, 19(1), 85–116.

<http://doi.org/10.1023/B:JOBU.0000040274.23551.1b>

Dawe, S., & Loxton, N. J. (2004). The role of impulsivity in the development of substance

use and eating disorders. *Neuroscience & Biobehavioral Reviews*, 28(3), 343–351.

<http://doi.org/10.1016/j.neubiorev.2004.03.007>

- Décary-Héту, D., Paquet-Clouston, M., & Aldridge, J. (2016). Going international? Risk taking by cryptomarket drug vendors. *International Journal of Drug Policy*, *35*, 69–76. <http://doi.org/10.1016/j.drugpo.2016.06.003>
- Dietz, G., & Den Hartog, D. N. (2006). Measuring trust inside organisations. *Personnel Review*, *35*(5), 557–588. <https://doi.org/10.1108/00483480610682299>
- Dingle, G. A., Stark, C. S., Cruwys, T., & Best, D. (2015). Breaking good: Breaking ties with social groups may be good for recovery from substance misuse. *British Journal of Social Psychology*, *54*(2), 236–254. <http://doi.org/10.1111/bjso.12081>
- Doliński, D. (2018). Is psychology still a science of behaviour? *Social Psychological Bulletin*, *13*(2), e25025. <http://doi.org/10.5964/spb.v13i2.25025>
- Doosje, B., Ellemers, N., & Spears, R. (1995). Perceived intragroup variability as a function of group status and identification. *Journal of Experimental Social Psychology*, *31*(5), 410–436.
- Drury, J., Cocking, C., & Reicher, S. (2009). The nature of collective resilience: Survivor reactions to the 2005 London Bombings. *International Journal of Mass Emergencies and Disasters*, *27*(1), 66–95.
- Duncan, L. A., Schaller, M., & Park, J. H. (2009). Perceived vulnerability to disease: Development and validation of a 15-item self-report instrument. *Personality and Individual Differences*, *47*(6), 541–546. <http://doi.org/10.1016/j.paid.2009.05.001>
- Evans, A. M., & Krueger, J. I. (2011). Elements of trust: Risk and perspective-taking. *Journal of Experimental Social Psychology*, *47*(1), 171–177. <http://doi.org/10.1016/j.jesp.2010.08.007>
- Faulkner, J., Schaller, M., Park, J. H., & Duncan, L. A. (2004). Evolved disease-avoidance mechanisms and contemporary xenophobic attitudes. *Group Processes & Intergroup Relations*, *7*(4), 333–353. <http://doi.org/10.1177/1368430204046142>

- Fessler, D. M. T., Clark, J. A., & Clint, E. K. (2015). Evolutionary psychology and evolutionary anthropology. In D. M. Buss (Ed.), *The handbook of evolutionary psychology*. University of California.
- Fessler, D. M. T., Eng, S. J., & Navarrete, C. D. (2005). Elevated disgust sensitivity in the first trimester of pregnancy. *Evolution and Human Behavior*, 26(4), 344–351.
<http://doi.org/10.1016/j.evolhumbehav.2004.12.001>
- Fincher, C. L., Thornhill, R., Murray, D. R., & Schaller, M. (2008). Pathogen prevalence predicts human cross-cultural variability in individualism/collectivism. *Proceedings. Biological Sciences / The Royal Society*, 275(1640), 1279–1285.
<https://doi.org/10.1098/rspb.2008.0094>
- Fincher, C. L., & Thornhill, R. (2008). Assortative sociality, limited dispersal, infectious disease and the genesis of the global pattern of religion diversity. *Proceedings of the Royal Society B: Biological Sciences*, 275(1651), 2587–2594.
<https://doi.org/10.1098/rspb.2008.0688>
- Fincher, C. L., & Thornhill, R. (2012). Parasite-stress promotes in-group assortative sociality: The cases of strong family ties and heightened religiosity. *Behavioral and Brain Sciences*, 35(2), 61–79. <https://doi.org/10.1017/S0140525X11000021>
- Firing, K., & Laberg, K. C. (2012). Personal characteristics and social identity as predictors of risk taking among military officers: An empirical study. *International Journal of Management*, 29(3), 86.
- Gaertner, S. L., Mann, J., Murrell, A., & Dovidio, J. F. (1989). Reducing intergroup bias: The benefits of recategorization. *Journal of Personality and Social Psychology*, 57(2), 239.
- Gilson, L. (2003). Trust and the development of health care as a social institution. *Social Science & Medicine*, 56(7), 1453–1468. [http://doi.org/10.1016/S0277-9536\(02\)00142-9](http://doi.org/10.1016/S0277-9536(02)00142-9)
- Goodwin, G. P., Piazza, J., & Rozin, P. (2014). Moral character predominates in person

- perception and evaluation. *Journal of Personality and Social Psychology*, 106(1), 148–168. <https://doi.org/10.1037/a0034726>
- Greenaway, K. H., Haslam, S. A., Cruwys, T., Branscombe, N. R., Ysseldyk, R., & Heldreth, C. (2015). From “we” to “me”: Group identification enhances perceived personal control with consequences for health and well-being. *Journal of Personality and Social Psychology*, 109(1), 53–74.
- Greenaway, K. H., & Louis, W. R. (2010). Only human: Hostile human norms can reduce legitimization of intergroup discrimination by perpetrators of historical atrocities. *British Journal of Social Psychology*, 49(4), 765–783.
<http://doi.org/10.1348/014466609X479202>
- Greenaway, K. H., Wright, R. G., Willingham, J., Reynolds, K. J., & Haslam, S. A. (2014). Shared identity is key to effective communication. *Personality and Social Psychology Bulletin*. <http://doi.org/10.1177/0146167214559709>
- Gullone, E., & Moore, S. (2000). Adolescent risk-taking and the five-factor model of personality. *Journal of Adolescence*, 23(4), 393–407.
<http://doi.org/10.1006/jado.2000.0327>
- Güth, W., Levati, M. V., & Ploner, M. (2008). Social identity and trust—An experimental investigation. *The Journal of Socio-Economics*, 37, 1293–1308.
- Hayes, A. F. (2018). *Introduction to Mediation, Moderation, and Conditional Process Analysis, Second Edition: A Regression-Based Approach* (2nd Ed). New York: The Guilford Press.
- Helliwell, J. F., & Wang, S. (2010). *Trust and wellbeing* (NBER WORKING PAPER SERIES TRUST No. 15911). Cambridge, MA. Retrieved from <http://www.nber.org/papers/w1591>
- Henrich, J., & Boyd, R. (2001). Why people punish defectors: Weak conformist transmission

- can stabilize costly enforcement of norms in cooperative dilemmas. *Theoretical Biology*, 208, 79–89. <https://doi.org/10.1006/jtbi.2000.2202>
- Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., ... Ziker, J. (2011). Markets, religion, community size and the evolution of fairness and punishment. *Science*, 327 <https://doi.org/10.1126/science.1182238>
- Holt-Lunstad, J., Smith, T. B., Baker, M., Harris, T., & Stephenson, D. (2015). Loneliness and social isolation as risk factors for mortality: A meta-analytic review. *Perspectives on Psychological Science*, 10(2), 227–237. <http://doi.org/10.1177/1745691614568352>
- Hornsey, M. J. (2008). Social identity theory and self-categorization theory: A historical review. *Social and Personality Psychology Compass*, 2(1), 204-222.
- Hopkins, N., & Reicher, S. (2016). Adding a psychological dimension to mass gatherings medicine. *International Journal of Infectious Diseases*, 47, 112–116. <http://doi.org/10.1016/j.ijid.2015.12.017>
- Hunt, M. K., Hopko, D. R., Bare, R., Lejuez, C. W., & Robinson, E. V. (2005). Construct validity of the Balloon Analog Risk Task (BART): Associations with psychopathy and impulsivity. *Assessment*, 12(4), 416–428. <https://doi.org/10.1177/1073191105278740>
- Hutton, A. (2012). Health messages not enough to protect harm at Schoolies Festivals. *Australian Nursing Journal*, 19(10), 41
- Igarashi, T., Kashima, Y., Kashima, E. S., Farsides, T., Kim, U., Strack, F., ... Yuki, M. (2008). Culture, trust, and social networks. *Asian Journal Of Social Psychology*, 11(1), 88–101. <http://doi.org/10.1111/j.1467-839X.2007.00246.x>
- Ishihara, S. (1917). *Tests for color-blindness*. Handaya, Tokyo, Hongo Harukicho.
- Jacoby, J., & Sassenberg, K. (2011). Interactions do not only tell us when, but can also tell us how: Testing process hypotheses by interaction. *European Journal of Social Psychology*, 41(2), 180–190. <http://doi.org/10.1002/ejsp.762>

- Janis, I. L. (1972). *Victims of Groupthink: A Psychological Study of Foreign-Policy Decisions and Fiascoes*. Oxford, England: Houghton Mifflin.
- Janz, N. K., & Becker, M. H. (1984). The health belief model: A decade later. *Health Education Quarterly*, *11*(1), 1-47.
- Johnson, D. D. P., Blumstein, D. T., Fowler, J. H., & Haselton, M. G. (2013). The evolution of error: Error management, cognitive constraints, and adaptive decision-making biases. *Trends in Ecology and Evolution*, *28*(8), 474–481.
<https://doi.org/10.1016/j.tree.2013.05.014>
- Jost, J. T., van der Linden, S., Panagopoulos, C., & Hardin, C. D. (2018). Ideological asymmetries in conformity, desire for shared reality, and the spread of misinformation. *Current Opinion in Psychology*, *23*, 77–83. <https://doi.org/10.1016/j.copsyc.2018.01.003>
- Jost, J. T. (2017). The marketplace of ideology: “Elective affinities” in political psychology and their implications for consumer behavior. *Journal of Consumer Psychology*, *27*(4), 502–520. <https://doi.org/10.1016/j.jcps.2017.07.003>
- Kahneman, D., & Lovallo, D. (1993). Timid choices and bold forecasts: A cognitive perspective on risk taking. *Management Science*, *39*(1), 17–31.
<http://doi.org/10.1287/mnsc.39.1.17>
- Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A focus theory of normative conduct: When norms do and do not affect behavior. *Personality and Social Psychology Bulletin*, *26*(8), 1002–1012. Retrieved from <http://psp.sagepub.com/content/26/8/1002.abstract>
- Kang, M. S. (2014). The health of “emerging adults” in Australia: Freedom, risk and rites of passage. *Medical Journal of Australia*, *201*(10), 562–563.
<http://doi.org/10.5694/mja14.01409>
- Khazaie, D. H., & Khan, S. S. (2019). Shared social identification in mass gatherings lowers health risk perceptions via lowered disgust. *British Journal of Social Psychology*.

<https://doi.org/10.1111/bjso.12362>

- Khan, K., Sears, J., Hu, V. W., Brownstein, J. S., Hay, S., Kossowsky, D., ... Cetron, M. (2013). Potential for the international spread of middle East respiratory syndrome in association with mass gatherings in Saudi Arabia. *PLoS Currents*, 5. <http://doi.org/10.1371/currents.outbreaks.a7b70897ac2fa4f79b59f90d24c860b8>
- Killen, J. (1985). Prevention of adolescent tobacco smoking: The social pressure resistance training approach. *Journal of Child Psychology and Psychiatry*, 26(1), 7–15.
- Kim, P. H., Dirks, K. T., Cooper, C. D., & Ferrin, D. L. (2006). When more blame is better than less: The implications of internal vs. external attributions for the repair of trust after a competence- vs. integrity-based trust violation. *Organizational Behavior and Human Decision Processes*, 99(1), 49–65. <https://doi.org/10.1016/j.obhdp.2005.07.002>
- Kramer, R. M., Brewer, M. B., & Hanna, B. (1996). Collective trust and collective action: The decision to trust as a social decision. In R. M. Kramer & T. R. Tyler (Eds.), *Trust in Organizations: Frontiers of Theory and Research* (pp. 357–389). Thousand Oaks, CA, CA: Sage.
- Kurath, J., & Mata, R. (2018). Individual differences in risk taking and endogenous levels of testosterone, estradiol, and cortisol: A systematic literature search and three independent meta-analyses. *Neuroscience & Biobehavioral Reviews*, 90, 428–446. <http://doi.org/https://doi.org/10.1016/j.neubiorev.2018.05.003>
- Lauriola, M., Panno, A., Levin, I. P., & Lejuez, C. W. (2014). Individual differences in risky decision making: A meta-analysis of sensation seeking and impulsivity with the balloon analogue risk task. *Journal of Behavioral Decision Making*, 27(1), 20–36. <http://doi.org/10.1002/bdm.1784>
- Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of*

Personality and Social Psychology, 93(2), 234–249. <https://doi.org/10.1037/0022-3514.93.2.234>

Leigh, B. C. (1999). Peril, chance, and adventure: Concepts of risk, alcohol use and risky behavior in young adults. *Addiction*, 94, 371–383.

Lejuez, C. W., Aklin, W. M., Zvolensky, M. J., & Pedulla, C. M. (2003). Evaluation of the Balloon Analogue Risk Task (BART) as a predictor of adolescent real-world risk-taking behaviours. *Journal of Adolescence*, 26, 475–479. [https://doi.org/10.1016/S0140-1971\(03\)00036-8](https://doi.org/10.1016/S0140-1971(03)00036-8)

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., ... & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, 8(2), 75.

Levine, M., Prosser, A., Evans, D., & Reicher, S. (2005). Identity and emergency intervention: How social group membership and inclusiveness of group boundaries shape helping behavior. *Personality and Social Psychology Bulletin*, 31(4), 443–453. <http://doi.org/10.1177/0146167204271651>

Lewin, K. (1948). *Resolving social conflicts; selected papers on group dynamics*. Oxford, England: Harper.

Loersch, C., & Bartholow, B. D. (2011). The color of safety: Ingroup associated colors make beer safer. *Journal of Experimental Social Psychology*, 47(1), 190–194. <http://doi.org/10.1088/1367-2630/15/1/015008>.Fluid

Loewenstein, G. F., Weber, E. U., Hsee, C. K., & Welch, N. (2001). Risk as feelings. *Psychological Bulletin*, 127(2), 267–286. <http://doi.org/10.1037//0033-2909.127.2.267>

Louis, W. R., Davies, S., Smith, J. R., & Terry, D. (2007). Pizza and pop and the student identity: The role of referent group norms in healthy and unhealthy eating. *The Journal of Social Psychology*, 147(1), 57–74.

- Lubman, D. I., Droste, N., Pennay, A., Hyder, S., & Miller, P. (2014). High rates of alcohol consumption and related harm at schoolies week: A portal study. *Australian and New Zealand Journal of Public Health, 38*, 536–541. <http://doi.org/10.1111/1753-6405.12266>
- Lundqvist, D., Flykt, A., & Ohman, A. (1998). *Karolinska directed emotional faces* [Database of standardized facial images]. Psychology Section, Department of Clinical Neuroscience, Karolinska Hospital, S-171 76 Stockholm, Sweden.
- Mayer, R. C. & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology, 84*, 123-136.
- Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management Review, 20*(3), 709-734.
- McElreath, R., Boyd, R., & Richerson, P. J. (2003). Shared norms and the evolution of ethnic markers. *Current Anthropology, 33*(1), 122–129.
- McGarty, C., Haslam, S. A., Hutchinson, K. J., & Turner, J. C. (1994). The effects of salient group memberships on persuasion. *Small Group Research, 25*(2), 267–293. <http://doi.org/10.1177/1046496494252007>
- Miller, D. T., & Prentice, D. A. (2016). Changing norms to change behavior. *Annual Review of Psychology, 67*, 339–361. <http://doi.org/10.1146/annurev-psych-010814-015013>
- Miller, S. L., & Maner, J. K. (2012). Overperceiving disease cues: The basic cognition of the behavioral immune system. *Journal of Personality and Social Psychology, 102*(6), 1198–1213. <http://doi.org/10.1037/a0027198>
- Mitchell, L., & West, P. (1996). Peer pressure to smoke—The meaning depends on the method. *Health Education Research, 11*(1), 39–46.
- Moon, J. W., Krems, J. A., & Cohen, A. B. (2018). Religious people are trusted because they are viewed as slow life-history strategists. *Psychological Science, 29*(6), 947-960.

- Montoya, A. K., & Hayes, A. F. (2017). Two condition within-participant statistical mediation analysis: A path-analytic framework. *Psychological Methods, 22*(1), 6-27
- Murray, D. R., & Schaller, M. (2016). The Behavioral Immune System: Implications for social cognition, social interaction, and social influence. *Advances in Experimental Social Psychology, 53*, 75–129. <http://doi.org/10.1016/bs.aesp.2015.09.002>
- Myers, D. G., & Lamm, H. (1976). The group polarization phenomenon. *Psychological Bulletin, 83*(4), 602–627. <http://doi.org/10.1037/0033-2909.83.4.602>
- Nesse, R. M. (2001). The smoke detector principle: Natural selection and defensive responses. *Annals of the New York Academy of Sciences, 935*(1), 75–85.
- Nicholson, N., Soane, E., Fenton-O’Creevy, M., & Willman, P. (2005). Personality and domain-specific risk taking. *Journal of Risk Research, 8*(2), 157–176. <http://doi.org/10.1080/1366987032000123856>
- Nutbeam, D. (2000). Health literacy as a public health goal: A challenge for contemporary health education and communication strategies into the 21st century. *Health Promotion International, 15*(3), 259–267. <http://doi.org/10.1093/heapro/15.3.259>
- Olatunji, B. O., Ebesutani, C., Kim, J., Riemann, B. C., & Jacobi, D. M. (2017). Disgust proneness predicts obsessive-compulsive disorder symptom severity in a clinical sample of youth: Distinctions from negative affect. *Journal of Affective Disorders, 213*, 118–125. <http://doi.org/10.1016/j.jad.2017.02.017>
- Olatunji, B. O., & Sawchuk, C. N. (2005). Disgust: Characteristic features, social manifestations, and clinical implications. *Journal of Social and Clinical Psychology, 24*(7), 932–962. <http://doi.org/10.1521/jscp.2005.24.7.932>
- Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences, 105*(32), 11087-11092. doi: 10.1073/pnas.0805664105

- Pan, W., & Bai, H. (2009). A multivariate approach to a meta-analytic review of the effectiveness of the D.A.R.E. program. *International Journal of Environmental Research and Public Health*, 6(1), 267–277. <http://doi.org/10.3390/ijerph6010267>
- Paradies, Y., Ben, J., Denson, N., Elias, A., Priest, N., Pieterse, A., ... Gee, G. (2015). Racism as a determinant of health: A systematic review and meta-analysis. *PLoS ONE*, 10(9), 1–48. <http://doi.org/10.1371/journal.pone.0138511>
- Park, J. H., Schaller, M., & Crandall, C. S. (2007). Pathogen-avoidance mechanisms and the stigmatization of obese people. *Evolution and Human Behavior*, 28(6), 410–414. <http://doi.org/10.1016/j.evolhumbehav.2007.05.008>
- Parker, S. L., Jimmieson, N. L., & Amiot, C. E. (2013). Self-determination, control, and reactions to changes in workload: A work simulation. *Journal of Occupational Health Psychology*, 18(2), 173–190. <http://doi.org/10.1037/a0031803>
- Parker, S. L., Laurie, K. R., Newton, C. J., & Jimmieson, N. L. (2014). Regulatory focus moderates the relationship between task control and physiological and psychological markers of stress: A work simulation study. *International Journal of Psychophysiology*, 94(3), 390–398. <http://doi.org/10.1016/j.ijpsycho.2014.10.009>
- Peters, E. M., Burraston, B., & Mertz, C. K. (2004). An emotion-based model of risk perception and stigma susceptibility: Cognitive appraisals of emotion, affective reactivity, worldviews, and risk perceptions in the generation of technological stigma. *Risk Analysis*, 24(5), 1349–1367. <http://doi.org/10.1111/j.0272-4332.2004.00531.x>
- Pettigrew, S., Biagioni, N., Daube, M., Jones, S. C., Chikritzhs, T., & Kirby, G. (2015). Understanding and addressing the Schoolies phenomenon. *Australian and New Zealand Journal of Public Health*, 39(2), 135–140. <http://doi.org/10.1111/1753-6405.12324>
- Pirlott, A. G., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of Experimental Social Psychology*, 66, 29–38.

<http://doi.org/10.1016/j.jesp.2015.09.012>

- Platow, M. J., Foddy, M., Yamagishi, T., Lim, L. I., & Chow, A. (2012). Two experimental tests of trust in in-group strangers: The moderating role of common knowledge of group membership. *European Journal of Social Psychology, 42*(1), 30–35.
- Platow, M. J., Voudouris, N. J., Coulson, M., Gilford, N., Jamieson, R., Najdovski, L., ... Terry, L. (2007). In-group reassurance in a pain setting produces lower levels of physiological arousal: Direct support for a self-categorization analysis of social influence. *European Journal of Social Psychology, 37*(4), 649–660.
- Postmes, T., Haslam, S. A., & Jans, L. (2013). A single-item measure of social identification: Reliability, validity, and utility. *British Journal of Social Psychology, 52*(4), 597–617.
- Reicher, S. D., Templeton, A., Neville, F., Ferrari, L., & Drury, J. (2016). Core disgust is attenuated by ingroup relations. *Proceedings of the National Academy of Sciences, 113*(10), 201517027. <http://doi.org/10.1073/pnas.1517027113>
- Reid, A. S. (2018). Financial crime in the twenty-first century: The rise of the virtual collar criminal. In N. Ryder (Ed.), *White Collar Crime and Risk: Financial Crime, Corruption, and the Financial Crisis* (pp. 231–251). London: Palgrave Macmillan.
- Reynolds, K. J., & Oakes, P. J. (2000). Variability in impression formation: Investigating the role of motivation, capacity, and the categorization process. *Personality and Social Psychology Bulletin, 26*(3), 355–373. <http://doi.org/10.1177/0146167200265008>
- Ronay, R., & Hippel, W. V. (2010). The presence of an attractive woman elevates testosterone and physical risk taking in young men. *Social Psychological and Personality Science, 1*(1), 57–64. <http://doi.org/10.1177/1948550609352807>
- Rotter, J. B. (1971). Generalized expectancies for interpersonal trust. *American Psychologist, 26*(1), 24–30. US: American Psychological Association. <http://doi.org/10.1037/h0031464>
- Salom, C., Watts, M., Kinner, S., & Young, D. (2005). Schoolies week in perspective:

Studies of alcohol, drug and risk-taking behaviour. *Of Substance*, 3(1), 26–28.

Schaller, M., Murray, D. R., & Bangerter, A. (2015). Implications of the behavioural immune system for social behaviour and human health in the modern world. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1669).
<http://doi.org/10.1098/rstb.2014.0105>

Schaller, M., & Park, J. H. (2011). The behavioral immune system (and why it matters). *Current Directions in Psychological Science*, 20(2), 99–103.
<http://doi.org/10.1177/0963721411402596>

Schoorman, F. D., Mayer, R. C., & Davis, J. H. (2007). An integrative model of organisational trust: Past, present and future. *Academy of Management Review*, 32(2), 344–354.

Shankar, S., Stevenson, C., Pandey, K., Tewari, S., Hopkins, N. P. P., & Reicher, S. D. D. (2013). A calming cacophony: Social identity can shape the experience of loud noise. *Journal of Environmental Psychology*, 36, 87–95.
<http://doi.org/10.1016/j.jenvp.2013.07.004>

Singh, R., Simons, J. J. P., Young, D. P. C. Y., Sim, B. S. X., Chai, X. T., Singh, S., & Chiou, S. Y. (2009). Trust and respect as mediators of the other- and self-profitable trait effects on interpersonal attraction. *European Journal of Social Psychology*, 39(6), 1021–1038. <http://doi.org/10.1002/ejsp>

Slovic, P., & Peters, E. (2006). Risk perception and affect. *Current Directions in Psychological Science*, 15(6), 322–325. <http://doi.org/10.1111/j.1467-8721.2006.00461.x>

Spencer, S. J., Zanna, M. P., & Fong, G. T. (2005). Establishing a causal chain: Why experiments are often more effective than mediational analyses in examining psychological processes. *Journal of Personality and Social Psychology*, 89(6), 845–51.

<http://doi.org/10.1037/0022-3514.89.6.845>

- Stover, D. (2013, September). How many Hiroshimas does it take to describe climate change? *Bulletin of the Atomic Scientists*. Retrieved from <https://thebulletin.org/2013/09/how-many-hiroshimas-does-it-take-to-describe-climate-change/>
- Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social categorization and intergroup behaviour. *European Journal of Social Psychology, 1*(2), 149–178. <http://doi.org/10.1002/ejsp.2420010202>
- Tajfel, H., & Turner, J. C. (1979). An integrative theory of intergroup conflict. In W. G. Austin & S. Worehel (Eds.), *The social psychology of intergroup relations* (pp. 33–47). Monterey: Brooks/Cole.
- Tam, J. S., Barbeschi, M., Shapovalova, N., Briand, S., Memish, Z. A., & Kieny, M.-P. P. (2012). Research agenda for mass gatherings: A call to action. *The Lancet: Infectious Diseases, 12*(3), 231–9. [http://doi.org/10.1016/S1473-3099\(11\)70353-X](http://doi.org/10.1016/S1473-3099(11)70353-X)
- Tanis, M., & Postmes, T. (2005). A social identity approach to trust: Interpersonal perception, group membership and trusting behaviour. *European Journal of Social Psychology, 35*(3), 413–424. <http://doi.org/10.1002/ejsp.256>
- Terry, D. J., & Hogg, M. A. (1996). Group norms and the attitude-behavior relationship: A role for group identification. *Personality and Social Psychology Bulletin, 22*(8), 776–793. <http://doi.org/10.1177/0146167296228002>
- Tingley, D., Yamamoto, T., Hirose, K., Keele, L., & Imai, K. (2014). Mediation: R package for causal mediation analysis. *UCLA Statistics/American Statistical Association, 59*(5). <http://doi.org/10.18637/jss.v059.i05>
- Todorov, A., Olivola, C. Y., Dotsch, R., & Mende-Siedlecki, P. (2015). Social attributions from faces: Determinants, consequences, accuracy, and functional significance. *Annual*

Review of Psychology, 66, 519-4308. doi: 10.1146/annurev-psych-113011-143831

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S. D., & Wetherell, M. S. (1987).

Rediscovering the social group: A self-categorization theory. Oxford: Blackwell.

Turner, J. C., & Oakes, P. J. (1997). The socially structured mind. In C. McGarty & S. A.

Haslam (Eds.), *The message of social psychology* (pp. 355–373). Oxford: Blackwell.

Valenstein, M., Bair, M. J., Ganoczy, D., McCarthy, J. F., Ilgen, M. A., & Blow, F. C. (2011).

Association between opioid prescribing patterns and opioid overdose-related deaths.

JAMA: The Journal of the American Medical Association, 305(13), 1315–1321.

van Rijswijk, W., Haslam, S. A., & Ellemers, N. (2006). Who do we think we are? The

effects of social context and social identification on in-group stereotyping. *British*

Journal of Social Psychology, 45, 161–174.

Vicario, C. M., Rafal, R. D., Martino, D., & Avenanti, A. (2017). Core, social and moral

disgust are bounded: A review on behavioral and neural bases of repugnance in clinical disorders. *Neuroscience and Biobehavioral Reviews*, 80, 185–200.

<http://doi.org/10.1016/j.neubiorev.2017.05.008>

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of*

Statistical Software, 36(3), 1–48. <http://www.jstatsoft.org/v36/i03/>.

Voci, A. (2006). The link between identification and in-group favouritism: Effects of threat

to social identity and trust-related emotions. *British Journal of Social Psychology*, 45,

265–284. <https://doi.org/10.1348/014466605X52245>

Woicik, P. A., Stewart, S. H., Pihl, R. O., & Conrod, P. J. (2009). The substance use risk

profile scale: A scale measuring traits linked to reinforcement-specific substance use profiles. *Addictive Behaviors*, 34(12), 1042–1055.

<http://doi.org/10.1016/j.addbeh.2009.07.001>

Woody, S. R., & Tolin, D. F. (2002). The relationship between disgust sensitivity and

avoidant behavior: Studies of clinical and nonclinical samples. *Journal of Anxiety Disorders*, 16(5), 543–559. doi.org/10.1016/S0887-6185(02)00173-1

Ziersch, A. M., Baum, F. E., Macdougall, C., & Putland, C. (2005). Neighbourhood life and social capital: The implications for health. *Social Science & Medicine*, 60(1), 71–86. <http://doi.org/10.1016/j.socscimed.2004.04.027>

Zinn, J. O. (2017). The meaning of risk-taking – key concepts and dimensions. *Journal of Risk Research*, 9877, 1–15. <http://doi.org/10.1080/13669877.2017.1351465>

Zuckerman, M., & Kuhlman, D. M. (2000). Personality and risk-taking: Common biosocial factors. *Journal of Personality*, 68, 999–1029. <http://doi.org/10.1111/1467-6494.00124>