

Social Psychology

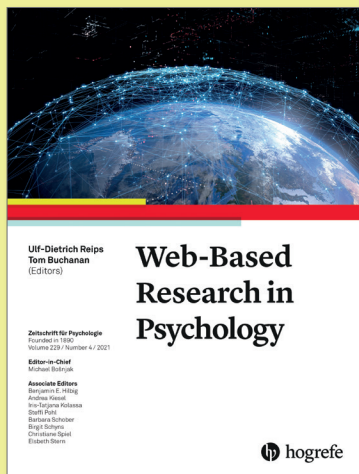
Editor-in-Chief

Anja Achtziger

Associate Editors

Anna Baumert
Wing Yee (VerBon) Cheung
Jaye L. Derrick
Kai Epstude
Adam Fetterman
Teresa Garcia-Marques
Mauro Giacomantonio
Mandy Hütter
Tali Kleiman
Toon Kuppens
Brian P. Meier
Andreas Mojzisch
Kim Peters
Lukas Thürmer
Kumar Yogeeswaran

A showcase of how the web can be used to advance psychological knowledge



Ulf-Dietrich Reips / Tom Buchanan (Editors)

Web-Based Research in Psychology

Zeitschrift für Psychologie, vol. 229/4

2021, iv + 76 pp., large format

US \$49.00 / € 34.95

ISBN 978-0-88937-597-0

Web-based research methodology has evolved since the development of the world wide web in the 1990s and has proliferated and diversified with layers upon layers of new major developments in Internet technology and life generally (e.g., search engines, social media, smartphones, Open Science). This volume presents a selection of state-of-the-art contributions on web-based research in psychology. Expert authors explore research methodology, including new methods made possible through the web or research that cannot be done without the web. In addition, research

ethics, which have special characteristics in the online research environment, are addressed to varying degrees, including deception and inclusivity. Further topics presented range from how web-based research can advance our knowledge on perception to the adoption of Open Science (including sharing data, materials, and preregistrations), as well as how behavior is observed in web-based research. This collection of contributions is a showcase of the creativity of researchers to find nifty new ways to harness the web to advance psychological research.

Contents and topics include

- From Modems to Mobile Apps: Web-Based Research in Psychology
- Personality Assessment To-Go: Formal Aspects of Unproctored Web-Based Personality Assessment in Relation to Its Content and Quality
- On the (Mis)Use of Deception in Web-Based Research
- Ebbinghaus Illusion: Relative Size as a Possible Invariant Under Technically Varied Conditions?
- Anxiety-Related Difficulties With Complex Arithmetic: A Web-Based Replication of the Anxiety-Complexity Effect
- Web Scraping: A Useful Tool to Broaden and Extend Psychological Research
- Does Peoples' Keyboard Typing Reflect Their Stress Level?
- Are Minority Opinions Shared Less?

Social Psychology

Volume 53/Number 1/2022

Editor-in-Chief	Anja Achtziger, Zeppelin University Friedrichshafen, Department of Psychology, Am Seemooser Horn 20, 88045 Friedrichshafen, Germany, Tel. +49 7541 6009 1376, Fax +49 7541 6009 1399, E-mail anja.achtziger@zu.de
Editorial Office	Julia Felfeli / Alexander Jaudas, Zeppelin University Friedrichshafen, Am Seemooser Horn 20, 88045 Friedrichshafen, Germany, Tel. +49 7541 6009 1379 / +49 7541 6009 1372, Fax +49 7541 6009 1399, E-mail SocialPsych.EditorialOffice@gmail.com
Associate Editors	<p>Anna Baumert, Max-Planck Institute for Collective Goods, Germany</p> <p>Wing Yee (VerBon) Cheung, University of Winchester, UK</p> <p>Jaye L. Derrick, University of Houston, TX, USA</p> <p>Kai Epstude, University of Groningen, The Netherlands</p> <p>Adam Fetterman, University of Houston, TX, USA</p> <p>Teresa Garcia-Marques, ISPA Instituto Universitário, Portugal</p> <p>Mauro Giacomantonio, University of Rome "La Sapienza", Italy</p> <p>Mandy Hütter, University Tübingen, Germany</p> <p>Tali Kleiman, The Hebrew University Jerusalem, Israel</p> <p>Toon Kuppens, University of Groningen, The Netherlands</p> <p>Brian P. Meier, Gettysburg College, PA, USA</p> <p>Andreas Mojzisch, University of Hildesheim, Germany</p> <p>Kim Peters, University of Exeter Business School, UK</p> <p>Lukas Thürmer, University of Salzburg, Austria</p> <p>Kumar Yogeeswaran, University of Canterbury, New Zealand</p>
Consulting Editors	<p>Susanne Abele (Oxford, OH, USA)</p> <p>Andrea Abele-Brehm (Erlangen-Nürnberg, Germany)</p> <p>Herbert Bless (Mannheim, Germany)</p> <p>Gerd Bohner (Bielefeld, Germany)</p> <p>Oliver Christ (Hagen, Germany)</p> <p>Paul Conway (Tallahassee, FL, USA)</p> <p>Katja Corcoran (Graz, Austria)</p> <p>Olivier Corneille (Louvain, Belgium)</p> <p>Amanda Diekmann (Oxford, OH, USA)</p> <p>Andrew Elliott (Rochester, NY, USA)</p> <p>Bertram Gawronski (Austin, TX, USA)</p> <p>Guido Gendolla (Geneva, Switzerland)</p> <p>Jessica Good (Davidson, NC, USA)</p> <p>Tobias Greitemeyer (Innsbruck, Austria)</p> <p>Michael Häfner (Berlin, Germany)</p> <p>Bettina Hannover (Berlin, Germany)</p> <p>Nina Hansen (Groningen, The Netherlands)</p> <p>Nicole Harth (Jena, Germany)</p> <p>S. Alexander Haslam (Brisbane, Australia)</p> <p>Ying-Yi Hong (Hong Kong, ROC)</p> <p>Joachim Hüffmeier (Dortmund, Germany)</p> <p>Roland Imhoff (Mainz, Germany)</p> <p>Eva Jonas (Salzburg, Austria)</p> <p>Franciska Krings (Lausanne, Switzerland)</p> <p>Maja Kutlaca (Osnabrueck, Germany)</p> <p>Daniel Lakens (Eindhoven, The Netherlands)</p> <p>Thomas Morton (Exeter, UK)</p> <p>Corinne Moss-Racusin (Saratoga Springs, NY, USA)</p> <p>Roland Neumann (Trier, Germany)</p> <p>Esther Papies (Glasgow, UK)</p> <p>Susanne Quadflieg (Bristol, UK)</p> <p>Marc-André Reinhard (Kassel, Germany)</p> <p>Toni Schmader (Vancouver, BC, Canada)</p> <p>Manfred Schmitt (Landau, Germany)</p> <p>Thomas W. Schubert (Oslo, Norway)</p> <p>Beate Seibt (Oslo, Norway)</p> <p>Monika Sieverding (Heidelberg, Germany)</p> <p>Dagmar Stahlberg (Mannheim, Germany)</p> <p>Fritz Strack (Würzburg, Germany)</p> <p>Christian Unkelbach (Cologne, Germany)</p> <p>Rolf van Dick (Frankfurt/Main, Germany)</p> <p>Harm Veling (Nijmegen, The Netherlands)</p> <p>Tobias Vogel (Mannheim, Germany)</p> <p>Ulrich Wagner (Marburg, Germany)</p> <p>Eva Walther (Trier, Germany)</p> <p>Michaela Wänke (Mannheim, Germany)</p> <p>Michael Wohl (Ottawa, ON, Canada)</p> <p>Bogdan Wojciszke (Warsaw, Poland)</p> <p>Vincent Yzerbyt (Louvain-la-Neuve, Belgium)</p>
Publisher	Hogrefe Publishing, Merkelstr. 3, 37085 Göttingen, Germany, Tel. +49 551 99950-0, Fax +49 551 99950-111, E-mail publishing@hogrefe.com North America: Hogrefe Publishing Corp., 44 Merrimac St., Suite 207, Newburyport, MA 01950, USA, Tel. +1 978 255 3700, E-mail publishing@hogrefe.com
Production	Regina Pinks-Freybott, Hogrefe Publishing, Merkelstr. 3, 37085 Göttingen, Germany, Tel. +49 551 99950-0, Fax +49 551 99950-111, E-mail production@hogrefe.com
Subscriptions	Hogrefe Publishing, Herbert-Quandt-Str. 4, D-37081 Göttingen, Germany, Tel. +49 551 99950-900, Fax +49 551 90050-998, zeitschriftenvertrieb@hogrefe.de
Advertising/Inserts	Melanie Beck, Hogrefe Publishing, Merkelstr. 3, D-37085 Göttingen, Germany, Tel. +49 551 99950-423, Fax +49 551 99950-111, E-mail marketing@hogrefe.com
ISSN	ISSN-L 1864-9335, ISSN-Print 1864-9335, ISSN-Online 2151-2590
Copyright Information	© 2022 Hogrefe Publishing. This journal as well as the individual contributions and illustrations contained within it are protected under international copyright law. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without prior written permission from the publisher. All rights, including translation rights, reserved.
Publication	Published in 6 issues per annual volume. <i>Social Psychology</i> is the continuation of <i>Zeitschrift für Sozialpsychologie</i> (ISSN 0044-3514), the last annual volume of which (Volume 38) was published in 2007.
Subscription Prices	Calendar year subscriptions only. Rates for 2022: Institutions – from US \$454.00/€348.00 (print only; pricing for online access can be found in the journals catalog at hgfi.io/journalcatalog); Individuals – US \$223.00/€168.00 (print & online). Postage and handling – US \$27.00/€24.00 (Germany €12.00). Single copies – US \$79.50/€62.50 + postage and handling.
Payment	Payment may be made by check, international money order, or credit card, to Hogrefe Publishing, Merkelstr. 3, D-37085 Göttingen, Germany. US and Canadian subscriptions can also be ordered from Hogrefe Publishing Corp., 44 Merrimac St., Suite 207, Newburyport, MA 01950, USA, Tel. +1 978 255 3700.
Electronic Full Text	The full text of <i>Social Psychology</i> is available online at https://econtent.hogrefe.com and in PsycARTICLES.
Abstracting Services	Abstracted/indexed in Current Contents/Social and Behavioral Sciences (CC/S&BS), Social Sciences Citation Index (SSCI), PsycINFO, PASCAL, PSYINDEX, ERIH, Scopus, and EM Care. 2020 Impact Factor 2.473, 5-year Impact Factor 2.510, <i>Journal Citation Reports</i> (Clarivate Analytics, 2021)

Contents

Original Articles	Forbid/Allow Asymmetry in Persuasion: The Forbid Frame Decreases Biased Elaboration and Increases Attitude Change <i>Paweł Koniak and Wojciech Cwalina</i>	1
	Members of Transgressor Groups Prefer Reparations to Come From Third Parties: An Extension of Interpersonal Findings on Moral Emotions <i>Maciej Sekerdej and Roger Giner-Sorolla</i>	21
	When You Are Wrong on Facebook, Just Admit It: Wrongness Admission Leads to Better Interpersonal Impressions on Social Media <i>Adam K. Fetterman, Nicole L. Muscanell, Dongjie Wu, and Kai Sassenberg</i>	34
<hr/>		
Research Report	Not Getting What You Want: Aggression, Prosocial Behaviors, and Popularity <i>Martin H. Jones, Toby J. Cooke, and Jennifer Symonds</i>	46
<hr/>		



Forbid/Allow Asymmetry in Persuasion

The Forbid Frame Decreases Biased Elaboration and Increases Attitude Change

Paweł Koniak^{ORCID} and Wojciech Cwalina

Department of Social Psychology, Maria Curie-Skłodowska University, Lublin, Poland

Abstract. Previous research showed that responses to questions about forbidding something differed from those to the seemingly equivalent questions about allowing the same object (forbid/allow asymmetry). We postulate that the effect of the forbid vs. allow framing may be also consequential for the processing of attitude related information and attitude change. The forbid frame (compared with the allow frame) may increase the impact of negative (vs. positive) arguments and/or reduce the impact of initial attitudes on the elaboration the presented information. To test these predictions we conducted three experiments (one preregistered, total $N = 655$). Participants were reading both pro and con arguments, differing in consistency with their initial attitudes, and concerning three different attitude objects: genetically modified organisms (GMOs), euthanasia, and barbecuing in public places. The results show that the forbid (vs. allow) frame decreases the tendency for generating thoughts prevalently consistent with participants' initial attitudes (Experiment 2). It also reduces bias in the evaluation and interpretation of the presented arguments and yields more similar assessments of arguments that are consistent and inconsistent with initial attitudes (Experiment 3). As a result, the attitudes are more susceptible to change within the forbid frame (they move more in the direction opposite to the initial attitude) than within the allow frame (Experiments 1-3). The results for the first time show the existence of forbid vs. allow asymmetry in persuasion. This effect has practical consequences, e.g., when designing referenda.

Keywords: forbid/allow asymmetry, persuasive communication, attitude depolarization, disconfirmation bias, resistance

In September 2018, in the Swiss canton of St. Gallen, a referendum was held about covering one's face in public (Shields, 2018). The media called it the *burka ban referendum* ("Voters approve 'burka ban' in St Gallen," 2018). Referendums are held fairly often in Switzerland, and at least some of them focus on whether something should be forbidden. For instance, in 2008, the Swiss voted on whether to forbid training flights of fighter aircraft over tourism areas in times of peace (Khetani, 2012), and in 2009, the Swiss voted on whether building minarets should be banned (Wyler, 2017). In each of these cases, the issue put to vote could have just as easily been presented within the allow frame. Instead of asking whether training flights or minarets should be banned (forbidden), the question could be whether these objects should be allowed.

The effect of wording questions in public opinion surveys is well known: Asking about forbidding something produces different responses than asking about allowing something (e.g., Chessa & Holleman, 2007; Hippler & Schwarz, 1986; Rugg, 1941). In a classic study, Rugg (1941) showed that if the question concerned allowing public speeches against democracy, 75% of the respondents chose an answer indicating

that they were against such speeches (i.e., they answered the speeches should not be allowed). However, if the question concerned forbidding public speeches against democracy, only 54% of the respondents were against such speeches (i.e., they answered they supported the ban). Subsequent studies confirmed the existence of the forbid versus allow asymmetry (Hippler & Schwarz, 1986; Holleman, 2006; Schuman & Presser, 1981): Responses to questions about forbidding something differed from those to the seemingly equivalent questions about allowing the same object. Specifically, more people were willing not to allow something than to forbid it, and vice versa – more persons were willing to not to forbid something than to allow it.

We postulate that the effect of the forbid versus allow framing is not limited to differences in responses to surveys or referendums. This effect can also concern the processes that take place when the respondent comes into contact with arguments during a referendum campaign. Instead of trying to persuade the voters that training flights or minarets should be or should not be banned (forbidden), the prereferendum campaigns could be focused on whether these objects should be allowed. However, would the attempts at persuasion have

progressed the same way in both cases (forbid vs. allow)? It is possible that framing can affect how different types of arguments and information (pro vs. con and attitude-consistent vs. attitude-inconsistent) are processed. Moreover, these potential differences in elaboration of different types of arguments can also affect the scope and direction of attitude change.

Factors Affecting Elaboration of Different Types of Arguments

During prereferendum campaign, citizens are confronted with positive and negative information (pro and con arguments) concerning the object of the referendum, but these different types of arguments or information are not evaluated and processed in the same way. As research on the negativity effect showed, people are more attentive to negative than positive information, and the negatives shape attitudes stronger than the positives (Baumeister et al., 2001). As a result, con arguments are more persuasive than pro ones (Cobb & Kuklinski, 1997).

Moreover, looking at arguments as positive or negative (pro or con) is only one way of encoding them. In addition to this descriptive encoding, these arguments can also be encoded evaluatively – as compatible or incompatible with the person's preexisting attitude (Wyer et al., 1991). As a result, influence of pro versus con arguments can be moderated by the compatibility or incompatibility of these arguments to the preexisting attitude. Arguments consistent with attitudes are accepted uncritically, while the inconsistent ones are prone to critical scrutiny and, as a result, are evaluated as weaker than consistent ones (*disconfirmation bias*; Edwards & Smith, 1996; Taber et al., 2009). This biased elaboration (*biased assimilation*; Lord et al. 1979; or *biased evaluation*; Cohen et al., 2000) leads to the perception of incoming evidence as supporting one's previous attitude, which in turn allows leaving this attitude unchanged or even to bolster it (Kuhn & Lao, 1996; Lord et al., 1979).

Both the tendency to concentrate on the negatives, as well as the tendency for a biased elaboration of evidence, may be moderated by individual-related factors, for example, by the strength of initial attitude (e.g., Pomerantz et al., 1995; Taber et al., 2009). What from our study perspective is more interesting – the aforementioned effects can be also moderated by situational or external factors (e.g., Cohen et al., 2000; Lord et al., 1984). Simple and seemingly unimportant variations in framing the question or task can change the way the object-related information is processed or weighted. This may also increase responders' propensity to correct their automatic or biased responses and affect the susceptibility of the

attitudes to change (e.g., Bizer & Petty, 2005; Sokolova & Krishna, 2016).

Here, we focus on the attribute framing, where only a single attribute in the object's description is changed, which results in differences in the evaluations of the object (Levin et al., 1998). According to Levin and Gaeth (1988), attribute-framing effects occur due to information being encoded relative to its descriptive valence. Positive frames lead to an encoding of the information that evokes favorable associations in memory, and negative frames evoke unfavorable associations. However, this is not only a matter of associations. Research on the task-type effect (which is an example of the attribute framing) showed that framing affects salience of different features of the object. In the task-type paradigm, the person has to decide which of the available options to choose or which of them to reject. For example, a person can be asked which of the two dishes offered they choose (positively framed task), but the question could also be which of the two dishes they reject (negatively framed task). With a positively (vs. negatively) framed task, people focus more on the positive (vs. negative) attributes of the objects. As a result, a rejection task, when compared with a choice task, showed a greater impact of negative information on decision-making (Ganzach, 1995; Shafir, 1993; Shafir et al., 1993; Wedell, 1997). Ganzach and Schul (1995) state that this effect may be mediated by *confirmation bias*. If the decision is framed as an acceptance decision, people are likely to test hypotheses concerning the acceptance of alternatives, and therefore, they attend to, and rely on, positive information more than negative information. On the other hand, when the decision is framed as a rejection decision, the reverse is likely to occur.

Moreover, in the case of previously existing attitudes, framing may also affect encoding and reliance on attitude consistent versus inconsistent information. It may also reduce the tendency for preferential treatment of information or arguments consistent with an initial attitude. Laran and Wilcox (2011) propose that, in a choice task, people focus more on the options consistent with the currently activated preferences and goals. In a rejection task, they focus on the preference-inconsistent alternatives. As a result, a rejection task shifts preferences toward options that are inconsistent with one's baseline preference. Other research showed that simply switching the task from choice to rejection led to more deliberative processing and increase responders' propensity to correct their automatic or biased responses (Sokolova & Krishna, 2016).

To sum up, attribute framing can change associations evoked by particular arguments and switch the salience of these arguments. This may result in changing the impact of different arguments (pro vs. con and consistent vs. inconsistent with initial attitude) on final attitude, which in

turn can affect the susceptibility of attitudes to change and determine the direction of this change.

Forbid Versus Allow Framing in Persuasion

To date, no studies have addressed susceptibility to persuasion in the forbid versus allow frame. However, for several reasons, we can expect that formulating the question within the forbid versus allow frame may lead to differences in the respondents' susceptibility to persuasion.

First, as Chessa and Holleman (2007) suggest, the forbid versus allow asymmetry in public opinion surveys is the consequence of an increased response threshold in forbid answers. If respondents are expected to take an extreme position (especially if the respondent is expected to agree to a ban), they may want "to retrieve extra arguments to be sure about their stance, or to defend it to an outsider" and, as a result, "an extra (negative) attribute of the same attitude or attitudinal network is being activated" (p. 222). This suggests the possibility that answering a question about forbidding something may be associated with a stronger susceptibility to the presented arguments than answering a question about allowing something. Moreover, it may lead to increased focus on some specific (i.e., negative) information.

Second, the forbid versus allow framing, like choosing versus rejecting framing, can be considered as an example of attribute framing. Moreover, both rejecting and forbidding, as well as choosing and allowing, have common characteristics.¹ Both rejecting and forbidding involve a lack of possibility for further action related to the object that has been rejected or forbidden (Holleman, 2006). Similarly, both choosing and allowing involve creating a possibility for further action related to the object that has been chosen or allowed. Moreover, in Polish, that is, the language in which this study was conducted, the words "reject" and "forbid" (*odrzucać* and *zakazywać*) and the words "choose" and "allow" (*wybierać* and *pozwalać*) have similar evaluative and emotional connotations (Wierzba et al., 2015). The valence of both *odrzucać* (reject) and *zakazywać* (forbid) is negative, and the level of excitement evoked by these words is similar. The valence of both *wybierać* (choose) and *pozwalać* (allow) is slightly positive, with a slightly

lower level of excitement evoked than with negative counterparts. Therefore, we can expect some analogies between the task-type effect described earlier and influence of forbid versus allow frame.

Third, the word "forbid" seems to have some specific status (especially when compared with the word "allow"; cf. Chessa & Holleman, 2007; Schuman & Presser, 1981). This word is connected with different signs informing that some actions cannot be taken in the given time or situation. Moreover, in many cases, the word "forbid" (or sign representing it graphically) implies that an ongoing activity cannot be continued or at least that the person should decide whether their activity is breaking the communicated rule. In the course of life, people are learning to associate the word "forbid" with stopping the action, an interruption of ongoing activity, an obstacle to fulfilling their goals, a disagreement with some activity, the consideration of potential fines, etc. It is possible, then, for the word "forbid" to have the ability to induce some specific state, where caution is evoked and previous actions or goals are reassessed; the state when people are motivated to attend to all relevant information more carefully and objectively (Kunda, 1990). In the context of persuasion, this can result in lowering the reliance on previously existing attitudes. As Houston and Fazio (1989) noted, when a person is found in a situation with potential serious consequences, they may be motivated to evaluate the incoming information with less attitudinally driven processes. The consistency or inconsistency of information with previous attitudes can become less important than an accurate evaluation of the situation. Persuasive consequences of this state can be similar to the effect of fear on the processing of information. Fear increases cautiousness and motivates a search for information, decreases the salience of prior beliefs and reliance on previous preferences, and encourages reconsideration of choices based on contemporary evaluations, which results in increasing the susceptibility to persuasion (Brader, 2005; Griskevicius et al., 2009; Koniak & Cwalina, 2020).

On the other side, the word "allow" is connected rather with the possibility of doing, the continuation of ongoing activity, an absence of obstacles to fulfilling actual goals, agreement, etc. In the course of life, people learn that when something is allowed, there is no need for

¹ It should be noted that most reject versus choose framing studies analyze its effect on preference choice (Ganzach & Schul, 1995), where the respondent has to choose between different objects. Research on forbid versus allow framing is mostly focused on its effect on preference judgment, where only one attitude object is being considered. Then, expressing opinion about forbidding an object is not the same as deciding which object from a larger set to reject. Similarly, expressing opinion about allowing an object is not the same as deciding which object from a larger set to choose. Thus, the forbid versus allow framing and reject versus choose framing, although both are examples of the attribute framing, are not identical.

constraining and re-evaluating their existing goals, that obstacles should not be expected, and that ongoing actions can be continued. There is no need to consider information other than one's own preferences. In the context of attitude change, this can result in lowering reliance on the presented arguments and increasing reliance on previously existing attitudes. Moreover, the impact of preexisting attitudes on the evaluation and interpretation of presented information or arguments should not be interrupted. As a result, preexisting attitudes can be kept without significant change.

Therefore, it can be expected that forbid versus allow framing will have consequences for the processes of changing attitudes toward the framed object. Specifically, we can expect that forbid versus allow frame can result in increased susceptibility of attitude to change. This change can result from how the negative versus positive (or attitude-consistent vs. attitude-inconsistent) information about the object is weighted and processed and even what of this information is seen as more persuasive.

Pro Versus Con Information

When people are considering whether an object should be forbidden, the arguments or information concerning negative aspects of the object (con arguments) should be more influential than those concerning positive ones (pro arguments). When they are considering allowing this object, the reverse is likely to occur. The differential influence of these arguments can be reflected in a change in their evaluation, in the ease of processing or memory for positive versus negative information, as well as in the pattern of eventual attitude change. From this perspective, which can be called a *descriptive valence account*, we can expect that when both pro and con arguments are presented, attitudes toward the object should change more in the negative direction when the forbid frame is used and/or more in the positive direction when the allow frame is used. Moreover, because of the word "forbid" arousing more excitement than the word "allow" (Wierzba et al., 2015), as well as for the negativity effect, we can expect the forbid frame to have a stronger ability to focus people on negatives than the allow frame would have to focus them on positives. As a result, we can expect that the absolute amount of attitude change will be higher in the case of the forbid frame than that of the allow frame.

Attitude-Consistent Versus Attitude-Inconsistent Information

From another perspective, which can be called an *evaluative encoding account*, it can be expected that forbid (vs. allow) framing would mitigate the overall tendency to rely mostly on arguments and information consistent with the previous attitude (regardless of their descriptive valence).

Moreover, it can also mitigate the tendency to discredit inconsistent arguments and/or the preferential treatment of consistent ones, that is, it can reduce the biased processing of information or arguments. This can result in weakening of tendency to perceive or interpret presented evidence as consistent or supportive of one's previous attitude. As a result, we can expect that when both arguments consistent and inconsistent with previous attitude are presented, and the forbid frame is used, attitudes toward the object should change more in a direction opposite to the initial attitude. When the allow frame is used, we can expect the attitude to remain unchanged or even to slightly polarize in a more extreme direction consistent with initial attitude (Kuhn & Lao, 1996; Lord et al., 1979).

However, it is also possible that forbid versus allow framing affects attitude change in a more direct way, where the mediating role of biased interpretation and/or evaluation of arguments is unnecessary. If the forbid frame is inducing some caution, the previous attitude can be simply withdrawn or at least mitigated because of this caution regardless of how this framing is affecting the perception and evaluation of arguments.

Study Overview

In this study, we are reporting the results of the three experiments (one preregistered) dealing with the consequences of forbid versus allow framing on changing preexisting attitudes (toward different objects: GMOs, euthanasia, and barbecuing in public places). First, in Experiments 1–3, we have tested whether the forbid versus allow framing leads to a decrease in the biased elaboration, evaluation, and interpretation of presented arguments. In Experiments 2 and 3, we have also verified whether the forbid versus allow frames simply affects processing of con versus pro arguments (descriptive valence account) or whether the relationship of these arguments to the preexisting attitude is crucial (evaluative encoding account). Next, we have tested whether forbid versus allow framing results in a different pattern of attitude change and whether this pattern is consistent with predictions of the descriptive valence account or evaluative encoding account (Experiments 1–3). We have also tested whether this attitude change is mediated by a decrease in the biased processing, evaluation, and interpretation of presented arguments (Experiments 2 and 3). Moreover, in Experiments 1 and 3, we have also tested potential alternative explanations for the proposed forbid/allow asymmetry in persuasion.

For all experiments, we report how we determined our sample size, all data exclusions (if any), all manipulations, and all measures. The experimental procedures were

approved by the Ethics Commission of SWPS University of Social Sciences and Humanities.

Experiment 1

The primary aim of Experiment 1 was to test whether the forbid versus allow framing affects susceptibility to the persuasion of attitudes toward forbidding or allowing barbecuing in public parks. Moreover, the arrangement of this study also helped to determine whether an account alternative to the ones mentioned above should be taken into consideration. According to Sokolova and Krishna (2016), the task-type effect can be explained by the deliberative processing account – they conclude that the rejection frame led to more deliberative processing. However, more deliberative processing (or elaboration, to use the persuasion studies terminology) can be decomposed as two separate processes (Petty & Cacioppo, 1986). First, the direction of processing may be affected, where more deliberative elaboration results in a decrease in biased processing (this is, in fact, equal to what we called the evaluative encoding account). Second, it can also be a matter of increased extent of elaboration, where the amount of elaboration is affected regardless of the direction of this elaboration. In this experiment, we have checked whether the forbid versus allow framing also affects the extent of the processing.

Method

Participants and Design

A total sample of 151 participants ($M_{\text{age}} = 23.05$ years; $SD = 3.46$; 97 women, 59.2%) were recruited to participate in the experiment via social media (invitations to participate were sent by email and posted on Facebook pages and other social media). A larger sample size was desired for this study; however, because our participants were recruited from the student sample, data were collected for as many participants as possible before the end of the semester. All participants completed the survey online and were not compensated. The participants were randomly assigned to a 2 (attitude question framing: forbid vs. allow) \times 2 (strength of arguments: strong vs. weak) between-participant factorial design. Participants evaluated the presented arguments, and the change of their attitudes was assessed by the repeated measures (attitudes toward the issue before vs. after the message). A sensitivity analysis conducted with G*Power (Faul et al., 2007) showed that our sample sufficed to detect the effects of $f = 0.23$ with a power of 0.80 for a

between-subject ANOVA and $f = 0.14$ for a mixed-design ANOVA.

Material and Procedure

The participants were first asked about their attitudes toward various issues: forbidding cycling on the sidewalk, allowing GMO crops, requiring every dog to be walked on a leash and aggressive breeds to wear also a muzzle, barbecuing in public parks, requiring drivers to use winter tires between November and March, forbidding the use of mobile phones in schools, and allowing the sale of home-made alcohol. The only topic of interest in this study was the attitude toward barbecuing in public parks; all the other topics were buffer questions. The participants were randomly shown one of two versions of the question about barbecuing in public parks: One group declared their attitude toward forbidding it and the other declared their attitude toward allowing it. They declared their attitude toward each issue on three 11-point items ranging from -5 (*very strongly against this proposal; this is a very bad idea; this should definitely not be introduced*) to 5 (*very strongly in favor of this proposal; this is a very good idea; this should definitely be introduced*). The reliability of this scale regarding the attitudes toward barbecuing in public parks was $\alpha = .98$.

A pilot study ($N = 26$) indicated that the population from which we planned to recruit the participants of Experiment 1 was dominated by the proponents of barbecuing in public parks. For this reason, only those participants who were in favor of barbecuing in public parks were recruited to Experiment 1 (i.e., those who answered higher than 0 to the question about allowing it and those who answered lower than 0 to the question about forbidding it).

After the initial measurement of their attitudes, the participants read a message containing only arguments that were inconsistent with their attitude, that is, arguments against barbecuing in public parks. One group read a message containing strong arguments (e.g., “Barbecuing in public places will result in these spaces getting more litter”), and another group read a message containing weak arguments (e.g., “Vegetarians, vegans, and other persons who do not eat meat may feel excluded from the public space”). Each version of the message contained three arguments, and the strength of these arguments had been tested in a pilot study ($N = 25$). The author of the message, described as Professor Andrzej Goździcki, President of the Polish City Planners Forum, explicitly declared at the beginning and end of the message that he was against barbecuing in parks owned by the city or village and that in his opinion, public parks were not the place for such entertainment as barbecuing.

Next, the participants used scales ranging from -5 to 5 to assess the strength of the arguments in the message (*very unconvincing–very convincing, makes no sense–makes total*

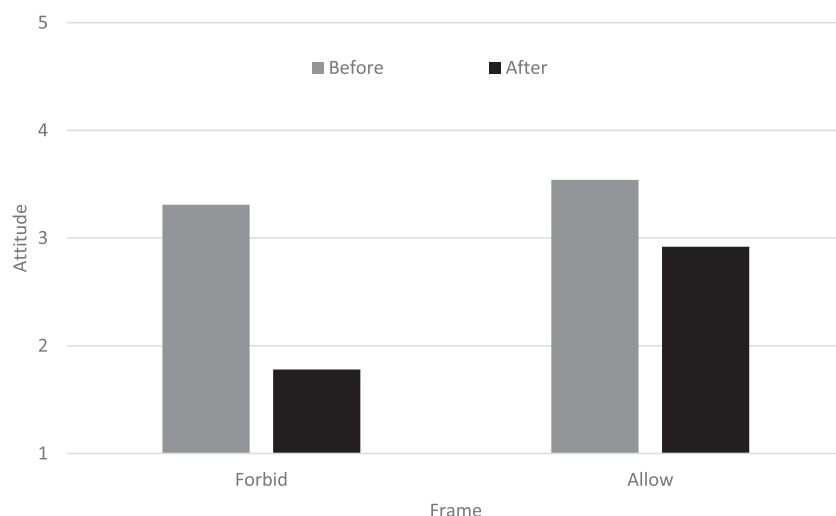


Figure 1. Attitudes toward barbecuing in public parks by frame condition and time of measurement (Experiment 1).

sense, very weak-very strong, completely unimportant-very important; $\alpha = 0.96$) and the direction of the arguments (*strongly against barbecuing in public parks-strongly in favor of barbecuing in public parks*). Last, the participants again declared their attitudes toward forbidding or allowing barbecuing in public parks using the same scale as at the beginning of the experiment ($\alpha = .98$).

Results

Evaluation of Arguments

We submitted both the perceived location of arguments on the *against-for* scale and the assessment of their strength to a two-way ANOVA. The assessment of the direction of arguments was not affected by the frame, argument strength, or the interaction between these factors ($F_s < 1$).

As intended, the weak arguments were perceived as weaker than the strong arguments ($M = -0.52$, $SD = 2.89$ and $M = 0.67$, $SD = 2.87$, respectively; $F(1, 147) = 6.44$, $p = .012$, $\eta_p^2 = .042$). The assessments of the strength of the arguments were not in any way affected by the frame ($F_s < 1$).

Attitude Change

The participants' attitudes were recoded from attitudes toward forbidding or allowing barbecuing in public parks into an index of attitudes toward barbecuing in public parks. We reversed the responses given by the participants who had been asked about forbidding barbecuing in public parks such that high values in both the forbid and allow groups indicate a more positive attitude toward barbecuing in public parks. Next, we conducted a 2 (attitude question framing: forbid vs. allow) \times 2 (strength of arguments: strong vs. weak) \times 2 (attitude measurement: before vs. after the message) ANOVA (where the last factor was

within-participants). Here, as well as in the further experiments, we report only the results directly related to our predictions and to the influence of the forbid versus allow frames. Due to the space limitations, we report the results which are of secondary importance (i.e., the effects not interacting with the frame) in the Electronic Supplementary Material, ESM 1.

We found a significant interaction between the frame and the time of attitude measurement, $F(1, 147) = 5.38$, $p = .022$, $\eta_p^2 = .035$. As shown in Figure 1, before the presentation of the message, participants in the forbid ($M = 3.31$, $SD = 1.46$) and allow ($M = 3.54$, $SD = 1.43$) groups had similar attitudes toward barbecuing in public parks ($F(1, 147) = 1.01$, $p = .317$). After the presentation of the counter-attitudinal message, attitudes in both the allow ($M = 2.92$, $SD = 2.48$; $F(1, 147) = 5.20$, $p = .024$, $\eta_p^2 = .034$) and the forbid groups shifted significantly in a more negative direction ($M = 1.78$, $SD = 2.54$; $F(1, 147) = 29.66$, $p < .001$, $\eta_p^2 = .168$). However, as a result of this shift, the final attitudes were less positive in the forbid than allow group ($F(1, 147) = 7.83$, $p = .006$, $\eta_p^2 = .051$), which indicates that the amount of change was higher within the forbid than allow group. Finally, neither of the interactions with the strength of arguments was significant ($F_s < 1$).

Discussion

Attitudes toward forbidding barbecuing in public parks were more susceptible to persuasion than attitudes toward allowing barbecuing in public parks. Moreover, we did not find any difference in the initial attitudes between the forbid and allow groups. In other words, the forbid/allow asymmetry in persuasion worked even when the forbid/allow effect known from the survey research did not occur.

Although manipulating the strength of the arguments proved to be effective, attitude change was independent from the quality of the presented arguments. Thus, in this experiment, we did not find any evidence that the forbid versus allow framing leads to a different extent of elaboration (Petty & Cacioppo, 1986). As a result, the difference in the scope of elaboration does not seem to be a plausible explanation for the observed difference in attitude change. It seems that it is the nature of this elaboration that is crucial here. The pattern of results of Experiment 1 is aligned with the overall prediction that the forbid frame increases the influence of negatives. However, because participants' preexisting attitudes were unipolar and positive, the arguments we have presented them were, at the same time, the con and the attitude-inconsistent. For this reason, the results do not allow us to indicate which of the accounts proposed (descriptive valence or evaluative encoding) is a more valid explanation of how forbid versus allow framing works. We will explore this issue in the next two experiments.

Finally, we would also expect that framing should influence the perception or evaluation of the arguments presented – an effect we did not find in this experiment. It is possible that this effect is not detectable with such direct measures as an evaluation of arguments. These processes can manifest, for example, in the fluency of processing different types of arguments, etc. It can also be detectable by tracking thoughts generated by the participant when they are thinking about forbidding versus allowing. Moreover, here only one type of argument was presented (con or inconsistent). It is possible that differences in evaluations are easier to detect when both pro and con (consistent and inconsistent) arguments are presented (see Edwards & Smith, 1996; Lord et al., 1979). It should also be stressed that framing effect can manifest itself in differential treatment of negative (con or inconsistent) or positive (pro or consistent) information, or both. In other word, it can affect the perception or evaluation of one type of information without affecting the perception or evaluation of the other one. Moreover, the perception of which type of information will be more strongly affected can be the result of the initial views held by the participants. For example, we can expect that for proponents, as the group holding weaker attitudes and thus inclined to bolster these attitudes, framing will primarily affect the perception of arguments consistent with an attitude (see Bizer et al., 2013, 2011; Clark & Wegener, 2013). It is possible, then, that by measuring the evaluation of only con (inconsistent)

arguments, we simply missed the opportunity to catch changes in the evaluation of arguments that are positive. All the above considerations have been taken into account in the following two experiments.

Experiment 2

In this experiment, we have investigated how the forbid versus allow framing affects attitude change toward different objects – GMOs and euthanasia. These issues were chosen due to the fact that they allow for a distinction between two groups of participants: the opponents and the proponents. The participants were presented with messages containing a set of arguments for and against GMOs or euthanasia. This allows us to manipulate both encoding of arguments (pro-con and consistent-inconsistent) orthogonally. Moreover, to capture eventual shifts in the focus on negative versus positive arguments, we have used the thought-listing technique.

Method

Participants and Design

A total sample of 175 participants ($M_{\text{age}} = 35.03$ years; $SD = 8.86$; 96 women, 54.9%) were recruited to participate in the experiment via social media. A larger sample size was desired for this study; however, because our participants were recruited from the student sample, data were collected for as many participants as possible before the end of the semester. All participants completed the survey online and were not compensated. They were randomly assigned to a 2 (issue: GMOs vs. euthanasia) \times 2 (attitude question framing: forbid vs. allow) between-participant factorial design. After the presentation of the message, participants listed thoughts that came to their minds when reading the message, and the change of their attitudes was assessed by the repeated measurements (attitudes toward the issue before vs. after the message). For an analytical reason, based on the initial attitude scale, we divided participants into two groups: proponents and opponents of the discussed issue (details described later). Eleven participants were excluded from further analysis due to being identified as outliers with an extremely high attitude change index (equal to a full range on the attitude scale). Their exclusion does not affect the results substantially.² A

² For the full sample-based analyses, our crucial three-way interactions remained significant for both the valence of thoughts ($F(1, 167) = 21.69$, $p < .001$, $\eta_p^2 = .115$) and the change of attitudes ($F(1, 167) = 15.64$, $p < .001$, $\eta_p^2 = .086$).

sensitivity analysis conducted with G*Power (Faul et al., 2007) showed that our final sample was sufficient for detecting the effects of $f = 0.15$ with a power of 0.80 for a mixed-design ANOVA.

Material and Procedure

The participants were randomly assigned one of two issues: Half of the participants indicated their attitudes toward GMO crops and read a message about GMOs, and the other half indicated their attitudes toward euthanasia and read a message about euthanasia. The initial measurement of the participants' attitudes toward GMOs or euthanasia was placed between questions about other attitudes: legalizing same-sex marriage, introducing the death penalty, penalizing possession of marijuana, removing religion classes from the school curriculum, and building a nuclear power plant. The issue of interest in this study was the question about GMOs or euthanasia; all other topics were buffer questions.

The participants in the GMOs group were randomly assigned one of two versions of the question about GMOs: Half of the participants declared their attitudes toward forbidding GMO crops and the other half declared their attitudes toward allowing GMO crops. Similarly, the participants in the euthanasia group were randomly assigned one of two versions of the question about euthanasia: They declared their attitudes toward either forbidding or allowing euthanasia. The participants indicated their attitudes toward each issue using an 11-point scale ranging from -5 (*strongly against*) to 5 (*strongly in favor*). Because the aim of the experiment was to measure the susceptibility of attitudes to change, only those participants who declared having an attitude toward the issue (responses between -5 and -1 and between 1 and 5) proceeded with the experiment.³

Next, the participants were informed that they will see a message about one of the aforementioned issues, in their case, the GMO crops (in one group) or euthanasia (in the second group). Participants assigned to the GMO version read a message with three pro arguments (e.g., "GMO farming provides high yields of crops even with bad soil") and three con arguments about GMOs (e.g., "The long-term effect of GMOs on human health is difficult to predict"). Similarly, participants in the euthanasia group read a message with arguments for (e.g., "A small number of medical errors or inexplicable recuperations cannot

justify the inordinate amount of pain and suffering experienced by many other, terminally ill patients") and against euthanasia (e.g., "The wish to undergo assisted suicide can be analogous to the wish to die caused by depression, and can also be eliminated using the same means as with other groups of patients with suicidal thoughts"). All arguments had been tested in pilot studies ($N = 15$ for GMOs and $N = 37$ for euthanasia). We only used the arguments that the participants of the pilot studies viewed as unambiguously for or against and did not differ in terms of understandability and strength.

In both versions of the message, the author (described as an expert in the respective field) only provided arguments for and against and did not state any opinion on the matter. Arguments for and against were provided in two separate paragraphs, preceded by the expressions "on the one hand" and "on the other hand." To control the effect of the order of arguments, the paragraphs were randomly presented in the for-against or the against-for order. Preliminary analyses showed that the effect of the forbid versus allow framing was not affected by the order of argument presentation; therefore, this factor was omitted in the subsequent analyses.

Next, the participants described the thoughts that came to their minds when reading the message. They were instructed to spend up to two minutes listing a maximum of 12 thoughts in separate boxes. Two judges, blind to the conditions in which the thoughts were generated, coded the thoughts as positive, negative, neutral, or irrelevant to GMOs or euthanasia (depending on the group). All disagreements were resolved through discussion.

Finally, the participants declared their attitudes toward forbidding or allowing GMOs (or euthanasia) again using the same scale as at the beginning of the experiment.

Results

To distinguish between effects supporting the descriptive valance account or evaluative encoding account, we need to include the valence of participants' preexisting attitudes in the analyses. For this reason, as well as for simplicity of presentation, we recoded the participants' attitudes from attitudes toward forbidding or allowing GMOs or euthanasia into a GMOs or euthanasia attitude index. The responses given by the participants who had been asked

³ Moreover, this experiment also tested the role of another variable that could be related to the effect of the frame on the susceptibility of attitudes to persuasion, that is, attitude certainty (Bizer et al., 2011). Unfortunately, due to a programming error, the initial attitude certainty scale was only administered in the euthanasia group. Since we have not found a significant influence of the forbid versus allow frame on this measure, we skip the more detailed description of it.

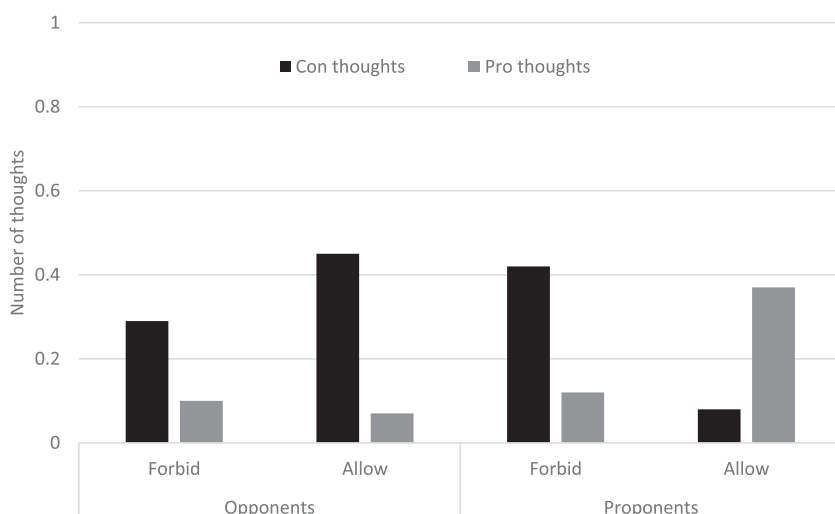


Figure 2. Numbers of pro and con thoughts reported as a function of initial attitude and the framing condition (Experiment 2).

about forbidding GMOs or euthanasia were reversed such that high values in both the forbid and allow groups mean a more positive attitude toward GMOs or euthanasia. Based on this, we divided participants into two groups: the proponents of the GMOs or euthanasia ($N = 52$) and the opponents of the GMOs or euthanasia ($N = 112$).

Valence of Thoughts

Our object of interest here was the number of pro versus con thoughts reported by participants related to the issue. To control for individual differences in the total number of thoughts reported, for each subject, we calculated ratios of each category over the total number of issue-related thoughts (pro, con, and neutral; Blankstein et al., 1989). Next, we conducted a 2 (issue: GMOs vs. euthanasia) \times 2 (attitude question framing: forbid vs. allow) \times 2 (initial attitude: pro vs. con) \times 2 (thoughts reported: pro vs. con) ANOVA (where the last factor was within-participants).

We found a significant interaction between the forbid versus allow framing and the numbers of pro versus con thoughts reported ($F(1, 156) = 5.33, p = .022, \eta_p^2 = .033$). For the forbid frame, the con thoughts prevailed over the pro thoughts ($M = 0.36, SD = 0.41$ and $M = 0.11, SD = 0.27$, respectively, $F(1, 156) = 18.51, p < .001, \eta_p^2 = .106$). In the allow frame condition, numbers of con and pro thoughts were similar ($M = 0.27, SD = 0.48$ and $M = 0.22, SD = 0.30$, respectively, $F < 1$). From the other side, pro thoughts were more numerous for the allow frame than the forbid frame ($F(1, 156) = 6.40, p = .012, \eta_p^2 = .039$), but the quantity of the con thoughts in the forbid frame was not significantly higher than in the allow frame ($F(1, 156) = 1.62, p = .205$). Overall, these results may suggest that processes predicted by the descriptive valence account were working here to some degree and that the participants were just more prone to generate more negative thoughts for the forbid

frame and/or more positive thoughts for the allow frame. However, even if the frame itself skewed the valence of reported thoughts, this effect was modified by the processes predicted by the evaluative encoding account. Specifically, the frame and thoughts interaction was qualified by a significant three-factor interaction between the forbid versus allow framing, the number of pro versus con thoughts, and the participants' initial attitudes ($F(1, 156) = 21.70, p < .001, \eta_p^2 = .122$).

As shown in Figure 2, in the allow frame condition, thoughts consistent with initial attitudes prevailed over inconsistent ones, for both proponents and opponents. Specifically, when the allow frame was applied, the opponents reported more con than pro thoughts ($M = 0.45, SD = 0.38$ and $M = 0.07, SD = 0.24$, respectively, $F(1, 156) = 45.35, p < .001, \eta_p^2 = .225$), while the proponents, conversely, reported more pro than con thoughts ($M = 0.37, SD = 0.24$ and $M = 0.08, SD = 0.39$, respectively, $F(1, 156) = 6.49, p = .012, \eta_p^2 = .040$). However, this tendency was mitigated (for the opponents) or even reversed (for the proponents) when the forbid frame was applied. In the forbid frame condition, the proponents reported more inconsistent (con) than consistent (pro) thoughts ($M = 0.42, SD = 0.41$ and $M = 0.12, SD = 0.26$, respectively, $F(1, 156) = 12.40, p < .001, \eta_p^2 = .074$). The opponents still reported more consistent (con) than inconsistent (pro) thoughts ($M = 0.29, SD = 0.39$ and $M = 0.10, SD = 0.25$, respectively, $F(1, 156) = 6.26, p = .013, \eta_p^2 = .039$). However, in the forbid condition, they reported less con (consistent) thoughts than in the allow condition ($F(1, 156) = 4.76, p = .031, \eta_p^2 = .030$). The amount of pro (inconsistent) thoughts reported by the opponents was similar for the forbid and allow frames ($F < 1$). Thus, the forbid frame decreased the generation of thoughts consistent with the preexisting attitude and, to some extent, also increased the

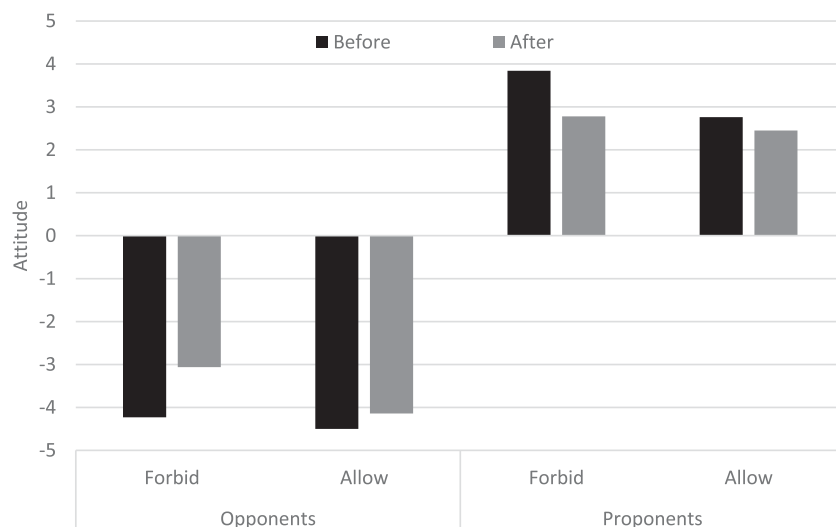


Figure 3. Before versus after the message attitudes as a function of initial attitude and framing condition (Experiment 2).

generation of thoughts inconsistent with this attitude. Overall, participants in the forbid group showed a less-biased pattern of thoughts.

Attitude Change

To explore whether forbid versus allow framing affects attitude change, we conducted a 2 (issue: GMOs vs. euthanasia) \times 2 (attitude question framing: forbid vs. allow) \times 2 (initial attitude: pro vs. con) \times 2 (attitude measurement: before vs. after the message) ANOVA (where the last factor was within-participants). We found a significant interaction between preexisting attitudes, the time of measurement of attitudes, and the forbid versus allow framing (see Figure 3; $F(1, 156) = 5.14, p = .025, \eta_p^2 = .032$). The opponents' initial attitudes were similar in the forbid and allow groups ($M = -4.23, SD = 1.11$ and $M = -4.50, SD = 1.09$, respectively, $F(1, 156) = 1.62, p = .206$). However, the proponents' reported more positive initial attitudes in the forbid than allow group ($M = 3.84, SD = 1.16$ and $M = 2.76, SD = 1.09$, respectively, $F(1, 156) = 10.91, p = .001, \eta_p^2 = .065$). In other words, in the last case (and only in this case), the forbid/allow asymmetry known from the research on the survey questions was found (see Hippler & Schwarz, 1986). What is more important, attitudes changed significantly only in the forbid frame groups. The opponents' attitudes shifted after the message in a positive direction when the forbid frame was applied ($F(1, 156) = 15.53, p < .001, \eta_p^2 = .091$) but remained unchanged in the allow frame condition ($F(1, 156) = 2.53, p = .114$). Similarly, the proponents' attitudes moved in a negative direction within the forbid frame group ($F(1, 156) = 9.17, p = .003, \eta_p^2 = .056$) but remained at the initial level within the allow frame group ($F < 1$). As a result of these shifts, the opponents' final attitudes were

less negative in the forbid than allow group ($M = -3.06, SD = 2.30$ and $M = -4.14, SD = 2.25$, respectively, $F(1, 156) = 5.87, p = .017, \eta_p^2 = .036$). The proponents' final attitudes were similar in the forbid and allow groups ($M = 2.78, SD = 2.40$ and $M = 2.45, SD = 2.26$, respectively, $F < 1$), but note that this was a consequence of the significant differences in the initial attitudes of this group of participants.

Mediational Analysis

To check whether the effect of forbid versus allow framing on attitude change is mediated by the decrease in biased elaboration, we used PROCESS model 4 (Hayes, 2018; with allow frame coded as 0, forbid frame coded as 1, continuous variables that define products mean centered, and CIs estimated using 5,000 bootstrapping samples). The overall consistency of each participant's thoughts with their initial attitude was indexed by subtracting the number of thoughts inconsistent from the number of the consistent ones and dividing this difference by the total number of issue-related thoughts. Thus, a high score indicated a dominance of thoughts that were consistent with a participant's initial attitude, or - to speak differently - thoughts more biased by the preexisting attitude. An attitude change index was created by subtracting the attitudes at Time 1 from the attitudes at Time 2 for participants with initially unfavorable attitudes and subtracting the attitudes at Time 2 from the attitudes at Time 1 for participants with initially favorable attitudes. As a result, positive scores indicated a change in the opposite direction from a participant's initial attitude (cf. Taber et al., 2009).

The forbid frame led to less biased thoughts, $b = -0.37, SE = 0.08, t(162) = -4.69, p < .001$. Decreasing the bias of

thoughts was associated with higher attitude change, $b = -0.73$, $SE = 0.29$, $t(161) = -2.55$, $p = .012$. The direct effect of a frame on attitude change was not significant, $b = 0.41$, $SE = 0.31$, $p = .182$. More importantly, the effect of the forbid versus allow frame on attitude change was fully mediated by the thought consistency index, effect = 0.27, boot $SE = 0.12$, 95% CI [0.06, 0.55].

Discussion

The results of Experiment 2 confirm that attitudes are more susceptible to change when the forbid frame is used than when the allow frame is used. Specifically, the forbid frame led to attitude change in the direction opposite to the initial attitude (which confirms the predictions of the evaluative encoding account), not in the direction of a more negative attitude (as the descriptive valence account predicts). Moreover, the change was induced solely by presenting a set of arguments for and against GMOs or euthanasia to the participants, with no clear favoring of either viewpoint. We can assume, then, that this change reflects the focus of participants on different types of arguments. Analyses of the thoughts generated by participants seem to confirm this assumption – the forbid frame decreased the generation of thoughts consistent with the preexisting attitude and to some extent also increased the generation of thoughts inconsistent with this attitude. Thus, the forbid frame may decrease the biased elaboration of the presented information (Edwards & Smith, 1996; Taber et al., 2009) and the tendency to confirm preexisting attitudes, or even to lead to a greater acceptance of information that is inconsistent with the initial attitudes.

Experiment 3

The aim of Experiment 3 was to replicate the results concerning the impact of the forbid versus allow frame on attitude change and biased processing of information. Moreover, this time we have checked whether framing can affect not only focus on the specific arguments but also how they are evaluated and interpreted. We have also administered some auxiliary measures, which can help to specify the exact nature of the forbid/allow asymmetry in persuasion. Primarily, we have measured associations evoked by the words “forbid” and “allow” (Wierzbica et al., 2015). Furthermore, a series of other measures was administered to check for possible alternative explanations and antecedents of the forbid/allow asymmetry in persuasion (e.g., attitude certainty, perceived risk, psychological distance).

Because none of these variables proved to serve as a reliable alternative explanation of the results obtained, and due to space limitations, we present the rationale behind all of them and their description and results in ESM 1. This study was preregistered at <https://aspredicted.org/2c56v.pdf>.

Method

Participants and Design

A sample of 329 participants ($M_{age} = 23.01$ years; $SD = 5.08$; 266 women, 78.9%) were recruited to participate in the experiment via four Polish university mailing lists and social media. All participants completed the survey online and were not compensated. Participants were randomly assigned to one of two experimental groups using the forbid or the allow frames. Next, pro and con arguments were presented and participants were asked to evaluate each argument using the scales provided. Moreover, initial and final attitudes were measured to assess the attitude change. Basing on the initial attitude scale, participants were divided into two groups: proponents and opponents of the discussed issue (details described later). A sensitivity analysis conducted with G*Power (Faul et al., 2007) showed that the final sample was sufficient for detecting the effects of $f = 0.09$ with a power of 0.80 for a mixed-design ANOVA.

Material and Procedure

At the beginning of the study, the participants declared their attitudes toward forbidding or allowing GMOs using an 11-point scale ranging from -5 (*strongly against*) to 5 (*strongly in favor*). Because the aim of the experiment was to measure the susceptibility of attitudes to change, only those participants who declared having an attitude toward the issue (responses between -5 and -1 and between 1 and 5) proceeded with the experiment. Based on this question, similar to Experiment 2, we divided participants into two groups: the proponents of GMOs ($N = 188$) and the opponents of GMOs ($N = 141$). Next, a battery of auxiliary measures was administered (see ESM 1).

After finishing answering the auxiliary measures, participants were informed that their task would be to indicate whether GMO crops should be allowed (in one group) or whether they should be forbidden (in the second group), and that to make their task easier, they would be shown several arguments about GMOs. Next, three pro and three con GMOs arguments were presented in random order (the arguments were the same as in Experiment 2). The participants answered questions concerning each argument's interpretation (*against GMOs-for GMOs*) and evaluation (*unpersuasive-persuasive*), scales ranging from -5 to 5 . Overall indices of interpretation and evaluation of the

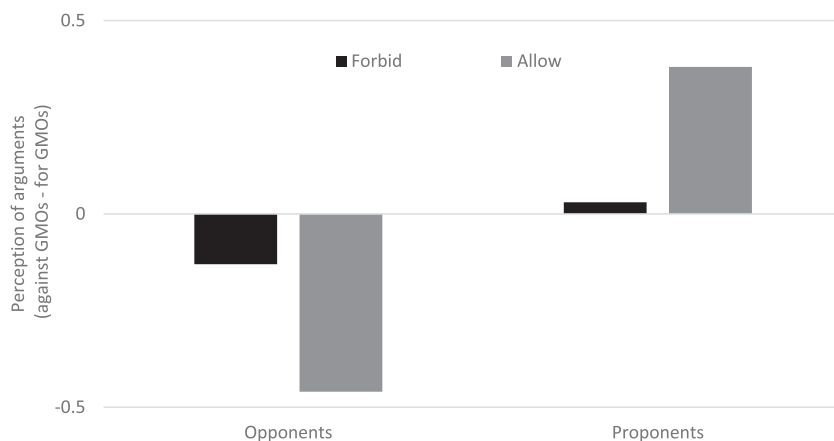


Figure 4. The perceived direction of arguments as a function of initial attitude and framing condition (Experiment 3).

pro and con arguments were created. Specifically, the assessments of all three pro arguments were averaged for interpretation ($\alpha = .86$) and evaluation ($\alpha = .76$) scales. A similar procedure was followed for the assessments of the three con arguments (interpretation index $\alpha = .83$ and evaluation index $\alpha = .64$).

Next, associations with the words “allow” (in the first group) and “forbid” (in the second group) were assessed. Participants indicated their associations using a series of bipolar adjective scales (ranging from -5 to 5). Associations with blocking the action and goals were indexed by averaging answers on the following scales: *difficulty-facilitation*, *obstacle-incentive*, *stopping action-continuing action*, *limitation-increase of available options*, *discord-consent* ($\alpha = .94$). To index associations with cautiousness and consideration of costs, answers on six scales were averaged: *caution-confidence*, *danger-safety*, *prevention-encouragement*, *receding-approaching*, *loss-gain*, *penalty-reward* ($\alpha = .87$).

Finally, the participants again declared their attitudes toward forbidding or allowing GMOs using the same scale as at the beginning of the experiment.

Results

Forbid Versus Allow Associations

Our analyses confirm that the word “forbid” brings up associations related to blocking an action and goals ($M = -3.24$, $SD = 1.64$), while the word “allow” is instead connected with the smooth continuation of an action and the realization of a goal ($M = 2.52$, $SD = 1.62$, $t(327) = 31.98$, $p < .001$, $d = 3.535$). Moreover, while the word “forbid” is associated with caution and consideration of costs ($M = -2.65$, $SD = 1.47$), the word “allow” is related to safety and orientation about gains ($M = 1.75$, $SD = 1.50$, $t(327) = 26.78$, $p < .001$, $d = 2.961$).

Interpretation of Arguments

One consequence of biased elaboration can be the tendency to perceive the direction of presented arguments or information as skewed in the direction of a preexisting attitude (cf. Houston & Fazio, 1989; Kunda, 1990). For example, pro arguments can be perceived as more pro by proponents, but in the eyes of the opponents, the same arguments may not support the pro stance so much. To check whether the forbid versus allow framing affects the perceived direction of an argument, we conducted a 2 (attitude question framing: forbid vs. allow) \times 2 (initial attitude: pro vs. con) \times 2 (arguments: pro vs. con) ANOVA, where the last factor was within-participants and the dependent variable was the assessment of arguments on the *against GMOs-for GMOs* scales.

None of the interactions with the framing and the type of arguments were significant. However, rather than differences in the perception of the pro and con arguments, we are interested in whether framing mitigates the overall tendency to interpret arguments or information as confirming a preexisting point of view. Overall, the proponents’ perception of all presented arguments was skewed in the direction for GMOs ($M = 0.20$, $SD = 0.99$), and the opponents perceived the same arguments as speaking rather against GMOs ($M = -0.29$, $SD = 0.98$, $F(1, 325) = 20.68$, $p < .001$, $\eta_p^2 = .060$). This tendency to perceive the arguments in accordance with one’s point of view was lowered in the forbid frame condition (see Figure 4; $F(1, 325) = 9.50$, $p = .002$, $\eta_p^2 = .028$). Specifically, the opponents assessed the presented arguments as less anti-GMOs in the forbid condition than they did in the allow condition ($M = -0.13$, $SD = 0.98$ and $M = -0.46$, $SD = 0.98$, respectively, $F(1, 325) = 4.10$, $p = .044$, $\eta_p^2 = .012$). The proponents assessed the presented arguments as less for GMOs when the forbid rather than allow frame was applied ($M = 0.03$,

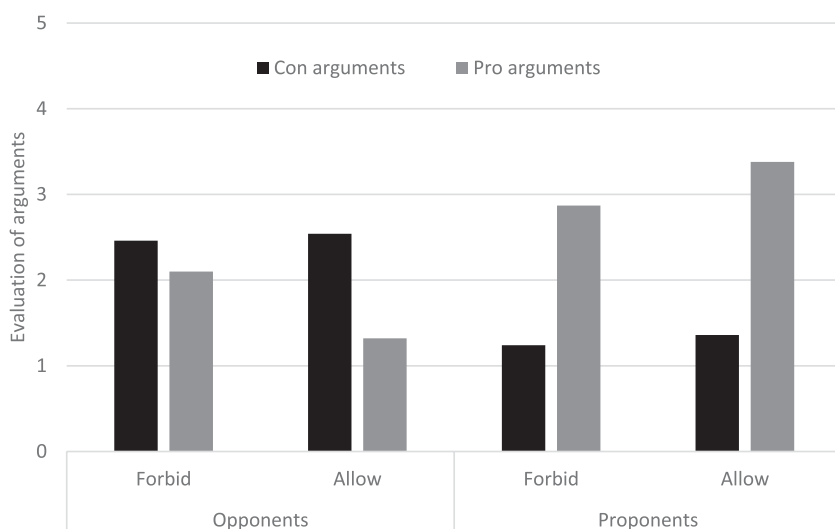


Figure 5. The assessment of the persuasiveness of con and pro arguments as a function of initial attitude and framing condition (Experiment 3).

$SD = 0.97$ and $M = 0.38$, $SD = 0.98$, respectively, $F(1, 325) = 5.60$, $p = .019$, $\eta_p^2 = .017$).

Evaluation of Arguments

To check whether forbid versus allow framing affects the evaluation of an argument, we conducted a 2 (attitude question framing: forbid vs. allow) \times 2 (initial attitude: pro vs. con) \times 2 (arguments: pro vs. con) ANOVA, where the last factor was within-participants and the dependent variable was the assessment of arguments on the *unpersuasive–persuasive* scales. We found a significant interaction between the type of arguments, the initial attitude, and the framing (see Figure 5; $F(1, 325) = 5.01$, $p = .026$, $\eta_p^2 = .015$). Framing did not affect the evaluation of the con arguments. The proponents of GMOs evaluated these arguments as equally unpersuasive in the forbid and allow conditions ($M = 1.24$, $SD = 2.05$ and $M = 1.36$, $SD = 2.05$,

respectively, $F < 1$). The opponents of the GMOs assessed the persuasiveness of these arguments as similarly high when the forbid versus allow frames were used ($M = 2.46$, $SD = 2.05$ and $M = 2.54$, $SD = 2.05$, respectively, $F < 1$). However, the forbid versus allow framing affected the evaluation of pro GMOs arguments. Specifically, the proponents evaluated the pro (attitude consistent) arguments as less persuasive in the forbid than allow condition ($M = 2.87$, $SD = 1.74$ and $M = 3.38$, $SD = 1.74$, respectively, $F(1, 325) = 3.99$, $p = .047$, $\eta_p^2 = .012$). The reverse was true for the opponents – they assessed the pro (attitude inconsistent) arguments as more persuasive when the forbid frame was used rather than allow ($M = 2.10$, $SD = 1.74$ and $M = 1.32$, $SD = 1.74$, respectively, $F(1, 325) = 7.15$, $p = .008$, $\eta_p^2 = .022$).

As a result, the overall tendency toward a preferential evaluation of arguments consistent versus inconsistent with preexisting attitudes (measured as the difference between

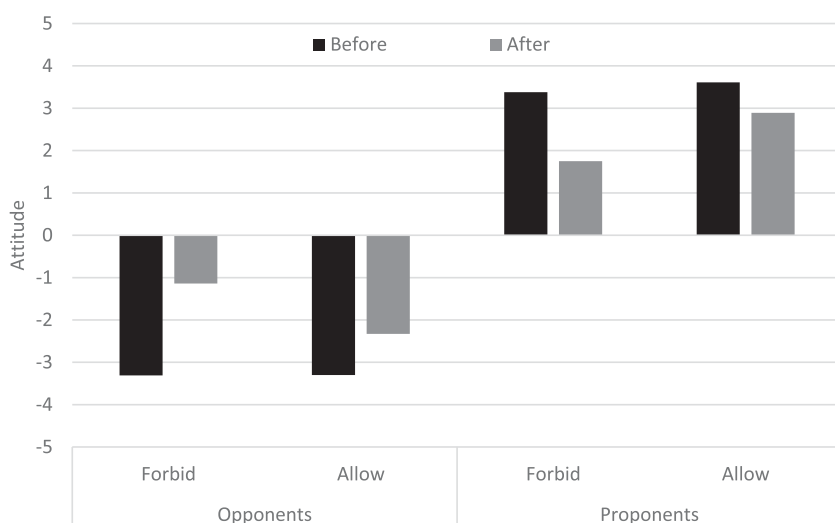


Figure 6. Attitudes before versus after the experiment as a function of initial attitude and framing condition (Experiment 3).

the evaluation of consistent and inconsistent arguments) was lower in the forbid than allow frame ($M = 0.99$, $SD = 2.50$ and $M = 1.62$, $SD = 2.56$, respectively, $F(1, 325) = 5.01$, $p = .026$, $\eta_p^2 = .015$).

Attitude Change

Similar to Experiment 2, we recoded the participants' attitudes from attitudes toward forbidding or allowing GMOs into a GMOs attitude index. The responses given by the participants who had been asked about forbidding GMOs were reversed such that high values in both the forbid and allow groups mean a more positive attitude toward GMOs. To explore whether the forbid versus allow framing affects attitude change, we conducted a 2 (attitude question framing: forbid vs. allow) \times 2 (initial attitude: pro vs. con) \times 2 (attitude measurement: before vs. after the message) ANOVA (where the last factor was within-participants).

We found a significant interaction between preexisting attitudes, the time of measurement, and the forbid versus allow framing (see Figure 6; $F(1, 325) = 14.76$, $p < .001$, $\eta_p^2 = .043$). The opponents' initial attitudes were similar in the forbid and allow groups ($M = -3.31$, $SD = 1.22$ and $M = -3.30$, $SD = 1.22$, respectively, $F < 1$). The same was true for the proponents – their initial attitudes were similar in the forbid and allow groups ($M = 3.38$, $SD = 1.22$ and $M = 3.61$, $SD = 1.22$, respectively, $F(1, 325) = 1.71$, $p = .192$). The opponents' attitudes shifted after the presentation of arguments in a positive direction, both when the forbid frame ($F(1, 325) = 56.79$, $p < .001$, $\eta_p^2 = .149$) and the allow frame were applied ($F(1, 325) = 10.93$, $p = .001$, $\eta_p^2 = .033$). Similarly, the proponents' attitudes moved in a negative direction both within the forbid frame group ($F(1, 325) = 35.49$, $p < .001$, $\eta_p^2 = .098$) and the allow frame group ($F(1, 325) = 9.46$, $p = .002$, $\eta_p^2 = .028$). However, as a result of these shifts, the opponents' final attitudes were less negative in the forbid than allow group ($M = -1.14$, $SD = 2.61$ and $M = -2.33$, $SD = 2.61$, respectively, $F(1, 325) = 7.41$, $p = .007$, $\eta_p^2 = .022$). Analogically, the proponents' final attitudes were less positive in the forbid than allow group ($M = 1.75$, $SD = 2.61$ and $M = 2.89$, $SD = 2.61$, respectively, $F(1, 325) = 8.78$, $p = .003$, $\eta_p^2 = .026$). Thus, although attitudes changed significantly in both the allow and forbid groups, the change was higher when the forbid frame was applied.

Mediation Analysis

To check whether the effect of the forbid versus allow framing on attitude change is mediated by the decrease in biased interpretation and/or biased evaluation of arguments, we ran a mediation analysis. Attitude change was indexed by subtracting the attitudes at Time 1 from the attitudes at Time 2 for participants with initially unfavorable attitudes and subtracting the attitudes at Time 2 from the attitudes at Time 1 for participants with initially favorable attitudes. As a result,

positive scores indicated a change in the opposite direction from a participant's initial attitude. To create an index of the biased perception of arguments, we reversed the opponents' assessments of the direction of the arguments such that high results mean the arguments are perceived as consistent with the initial attitude. The biased evaluation of arguments was indexed by subtracting the evaluation of arguments inconsistent with the participant's initial attitude from the evaluation of consistent ones. As a result, positive scores indicated that arguments consistent with initial attitudes were evaluated as more persuasive than inconsistent ones.

We used PROCESS model 4 (Hayes, 2018), with biased interpretation and the biased evaluation of the arguments serving as two mediators (allow frame coded as 0, forbid frame coded as 1, continuous variables that define products mean centered, and CIs estimated using 5,000 bootstrapping samples). The forbid frame led to a less-biased interpretation of the arguments, $b = -0.33$, $SE = 0.11$, $t(327) = -3.06$, $p = .002$, as well as a less-biased evaluation of them, $b = -0.68$, $SE = 0.28$, $t(327) = -2.42$, $p = .016$. The decreased bias in the interpretation of the arguments was associated with higher attitude change, $b = -0.49$, $SE = 0.13$, $t(325) = -3.67$, $p < .001$. The same was true for the decreased bias in evaluation, $b = -0.27$, $SE = 0.05$, $t(325) = -5.34$, $p < .001$. The direct effect of the frame on attitude change was significant, $b = 0.71$, $SE = 0.25$, $t(325) = 2.81$, $p = .005$. However, more importantly, the effect of the forbid versus allow frame on attitude change was mediated by both the biased interpretation index, effect = 0.16, boot $SE = 0.07$, 95% CI [0.05, 0.33], and the biased evaluation index, effect = 0.19, boot $SE = 0.08$, 95% CI [0.04, 0.37].

Discussion

Overall, the pattern of results in this study confirms the predictions of the evaluative encoding account. First, the interpretation of arguments, regardless of whether they were pro or con, was less biased in the forbid than allow frame – results not predicted by the descriptive valence account. Second, the pattern of results for the assessment of arguments' persuasiveness also seems to indicate that the consistency versus inconsistency of arguments with initial attitude was crucial here. For the proponents of GMOs, the forbid frame mitigated the tendency toward a higher evaluation of arguments consistent with attitudes. For the opponents, the forbid frame reduced the tendency toward a low evaluation of arguments inconsistent with attitudes (we will come back to this asymmetry between proponents and opponents in the general discussion). Despite these shifts in evaluation not concerning both types of arguments, they still led to an overall decrease in the biased evaluation of the information presented.

Finally, decreased bias in the interpretation and evaluation of arguments seems at least a partial reason for the greater susceptibility to change of attitudes within the forbid versus allow frame.

General Discussion

In three experiments (one preregistered), we showed that the forbid/allow asymmetry affects the effectiveness of persuasion and may impact public discourse. Asking citizens about forbidding GMOs, rather than allowing GMOs, may not only yield different survey results (e.g., Chessa & Holleman, 2007; Hippler & Schwarz, 1986), but it may also increase citizens' susceptibility to persuasion. As the results of our experiments show, the forbid frame (compared with the allow frame) reduced the impact of initial attitudes on the direction of processing and evaluating the presented information. As a result, the forbid frame caused people to defend the arguments or thoughts consistent with preexisting attitudes less strongly and/or less strongly discredit the inconsistent ones. Specifically, the forbid frame decreases the tendency for generating thoughts prevalently consistent with participants' initial attitudes (Experiment 2). Moreover, when compared to the allow frame, the forbid frame also reduces bias in the evaluation and interpretation of the presented arguments and yields more similar assessments of arguments that are consistent and inconsistent with initial attitudes (Experiment 3). This contributes to a relative increase in the influence of inconsistent arguments on final attitude. This, in turn, led to a change of attitude in the direction opposite to the initial attitude. Overall, attitudes are more susceptible to change within the forbid frame than within the allow frame (Experiments 1–3).

These results indicate the existence of new, previously undescribed effects. The forbid versus allow framing seems to be more than simply choosing between two alternative and compatible formulations of an issue. It can be very impactful in the case of prereferendum campaigns or in the case of campaigns aimed at changing social attitudes. Inducing people to think about forbidding (instead of allowing) some object, for example, GMO or euthanasia, leads them to process the information more objectively and can result in attitude change.

Limitations and Further Research

Our results can also shed some light on the potential limitations of the effect described. Further research could

help to specify the conditions in which the forbid versus allow framing works as well as to describe the exact mechanism of its influence.

Issue Importance and Scope of Elaboration

First, it should be noted that the issues used in Experiments 2 and 3 were fairly relevant to the participants (GMOs and euthanasia). Our two other supplementary studies asked the participants about how relevant these topics were to them (the participants assessed the topics on scales ranging from -5 to 5 , i.e., *completely irrelevant*–*very relevant*). The results show that both GMOs ($M = 2.00$, $SD = 2.62$, $N = 251$) and euthanasia ($M = 2.79$, $SD = 2.29$, $N = 104$) were assessed as fairly important. This suggests the possibility that to initiate the process described in the evaluative encoding account, not only is the preexisting attitude necessary, but this attitude should also be important or relevant enough to motivate the biased processing of arguments (the processing, perception, and evaluation of information inconsistent vs. consistent with initial attitudes). In the case of less important attitudes, or when the object is novel and previous attitudes do not exist, the predictions of the descriptive valence account could be more applicable and forbid versus allow framing may simply change the influence of positive (pro) versus negative (con) information. Moreover, in this case, attitude change can also be the simple result of vigilance – people can withdraw their initial assessments without considering the available information at all.

Furthermore, research on valence framing shows it is limited to situations of high cognitive elaboration: Differences in the strength of positively and negatively framed attitudes disappear when attitudes are of low relevance or when the participants are under cognitive load (Bizer et al., 2013). Similarly, the focus on preference-consistent information in a choice task and on preference-inconsistent information in a rejection task is eliminated or even reversed when the information is processed under cognitive load (Laran & Wilcox, 2011). Our study involved no factors that could have led to cognitive load. The initial level of motivation for elaboration was also not manipulated. However, the aforementioned results of the supplementary studies suggest that the issues used in Experiment 2 and Experiment 3 were relevant enough to motivate participants to process the information. With the issue used in Experiment 1, it should be noted that participants were informed at the beginning of the study that the issues they would be presented with could concern their neighborhoods, which may have increased the initial scope of processing (Petty & Cacioppo, 1986). Therefore, it is possible that, similar to the valence framing effect and the task-type effect, the precise mechanism behind the

forbid versus allow effect or even the existence of this effect is related to participants' motivation or ability to process.⁴

Attitude Strength

Second, it is also worth noting that we found some asymmetry between the evaluations by proponents and opponents. In Experiment 2, the forbid frame affected both the consistent and inconsistent thoughts for the proponents only. Specifically, the forbid frame led these participants to generate less consistent and more inconsistent thoughts. For the opponents, we found only a forbid frame-related decrease in the generation of thoughts consistent with the initial attitude. This asymmetry may be explained with the help of the previous research, which showed that the attitudes of opponents are more resistant and stronger than the attitudes of proponents (Bizer & Petty, 2005; Bizer et al., 2011, 2013). Switching the focus of less resistant participants from the arguments consistent to inconsistent may be relatively easy. However, when participants are more resistant, undermining what they accept may be easier than making them think the way that they initially opposed (see Clark & Wegener, 2013). Furthermore, in Experiment 3, the forbid frame affected (decreased) the evaluation of arguments consistent with the attitudes of the proponents, while for the opponents the evaluation of inconsistent ones was changed (increased). This asymmetry could be explained by the different motives prevailing within the proponents and opponents. The proponents may be more inclined to bolster their attitudes, that is, to look for information confirming them. The opponents may be oriented toward defending their attitude, which can be achieved by attacking information inconsistent with this attitude (Clark & Wegener, 2013). Thus, it seems that for both the proponents and the opponents, the forbid frame decreased the predominant motivation. As motives involving bolstering are more probable for relatively weak attitudes, and defense motives for relatively strong ones, these results suggest that the attitudes of proponents were weaker than the attitudes of opponents, which once again seems to be in accord with Bizer and Petty's results (2005; see also Bizer et al., 2011, 2013). Overall, this asymmetry indicates that the frame is not the only factor affecting the influence of different types of arguments. Factors such as strength of initial attitude may be important here – frame and the

attitude strength may conjointly shape the processing of arguments and scope and direction of attitude change.

Moreover, forbid versus allow framing can also be consequential for attitude strength. In two of our studies (see footnote 3 and ESM 1), we did not find evidence that the forbid versus allow asymmetry (we found for other measures) is mediated by the frame-related shift in the certainty of initial attitudes. In fact, participants answered questions about forbidding versus allowing an object with a similar level of certainty. However, it does not mean that forbid versus allow framing has no consequences for the attitude certainty at all. We measured certainty of attitudes at the beginning of the study, just after the initial question about attitude toward forbidding versus allowing, when no other attitude object-related information was presented. Thus, subtle difference in the initial framing of the situation (asking about forbidding vs. allowing) can be not enough to change the certainty of attitudes. In other studies, where impact of different frames on attitude certainty was found, the certainty measure was applied after participants have an occasion to process some attitude object-related information (Bizer et al., 2013, 2011; Rucker et al., 2008). It is possible, then, that forbid versus allow framing can affect the certainty of attitudes at later stages of persuasion process when people encounter some argumentation and when they start thinking about the object of persuasion.

Furthermore, basing on the other research findings, we can even expect that frame can affect attitude change via two separate routes: affecting consistency of the thoughts with initial attitude and affecting the confidence of these thoughts (Clarkson et al., 2011). Since we have found that forbid versus allow framing differently affects processing and evaluation of thoughts consistent versus inconsistent with initial attitude, especially interesting would be checking how these frames affect the certainty of these two types of thoughts. As Clarkson et al. (2013) showed, attitude-consistent thought confidence and attitude-inconsistent thought confidence can play an independent role in mediating attitude change (see also Koniak & Cwalina, 2020).

Object of Attitude

Finally, it is also worth taking into account the other characteristics of the object of an attitude. This study tested how the forbid versus allow frame affected the susceptibility of attitudes toward GMOs, euthanasia, and

⁴ Although it should be noted that research on risky choice framing found exactly opposite effect, the way in which a decision was framed had a weaker influence among people for whom the issue was more relevant (McElroy & Seta, 2003). This discrepancy between the role of the relevance in the attribute framing and the risky choice framing research may confirm that these are distinctive types of framing and that the effects found for the one type of framing cannot be easily extrapolated on the second type of framing (Levin et al., 1998).

barbecuing in public parks to persuasion. It seems that, at least in Poland, where this study took place, the population is mostly unfamiliar with the legal status of these issues; neither GMOs nor euthanasia is persistently associated with the status of a forbidden or allowed object (CBOS, 2013; Szadowska-Szlachetka et al., 2019). Familiarity with the legal regulations concerning barbecuing in public places is also uncommon; moreover, the issue itself is not specifically regulated. With such *indeterminate* objects, both the forbid and the allow frames can easily be used. Asking about forbidding GMOs or euthanasia comes as naturally as asking about allowing them. However, the forbid versus allow asymmetry in persuasion may be reduced if the status of an object is commonly known. For instance, in Poland, asking respondents about allowing marijuana (which is currently an illegal substance) may seem more natural than asking about forbidding it (as the ban is currently in effect). Moreover, the actual status of the object, whether it is forbidden or allowed, or rather how this status is perceived by people, can be an important mediator of the effect. For the objects already allowed, when people know that they are free to use these objects, asking about forbidding them can rather rise a reactance than create an opportunity for the less-biased processing of object-related information (cf. Koniak & Cwalina, 2020). This would suggest that for such objects, the forbid frame could even lead to a polarization of attitudes. However, this effect could probably be modified depending on whether the proposed ban or restriction would be seen as an absolute (i.e., complete, certain, or permanent) or nonabsolute (i.e., incomplete, uncertain, or temporally or spatially limited; see Laurin et al., 2012).

Another possibly important factor is how an issue is discussed in the public space. For example, most respondents have likely an established attitude toward GMOs, an attitude that will translate into their attitude toward forbidding or allowing GMOs (Aerni et al., 2011; Rzymiski & Królczyk, 2016). However, certain issues are immediately discussed with either frame, as with the ban on trading on Sundays in Poland (Sierpińska, 2007). The object of attitudes in this case was more likely *forbidding trading on Sundays*, rather than *trading on Sundays*. It is possible that an increased susceptibility of attitudes to change did not occur in response to this ban.

Relation to Other Research

Finally, some notes should be made about the relationship of our results to other studies. First, our results could be seen as contradicting those obtained by Bizer and Petty (2005), where negative framing resulted in attitudes more resistant to change. We believe that this is only an apparent

inconsistency. Note that in the case of the valence framing described by Bizer and Petty (2005), the frame changed the valence of attitude, that is, it led people to think about themselves as opponents versus proponents. If a person is to choose between two options, one they like and one they do not (e.g., between liked candidate A and disliked candidate B), the attitude is framed positively by asking the person about their attitude toward the option they like. For example, a person asked what she thought of liked candidate A being elected would indicate that she supports this result. To frame the attitude negatively, the person needs to be asked about the attitude toward the option she dislikes (e.g., a person asked what she thought of disliked candidate B being elected would indicate that she opposes this result). In other words, in the case of the valence framing, frame concerns the attitude of the person and affects how people “conceptualize their own preferences” (Bizer et al., 2011, p. 64). However, in case of the forbid versus allow framing, as well as in case of the task-type effect and the attribute framing generally, framing does not determine the valence of attitudes. The forbid frame does not necessarily lead to negative attitudes, and the allow frame does not necessarily lead to positive attitudes. These frames concern the attitude object and change the way the object is perceived and evaluated.

It is possible, then, that framing primarily impacts processing, evaluation, or weighting a framed subject-related information. In the case of an attitude framing, frame changes the way this attitude is weighted, how it affects the persuasion process, etc. In the case of framing of an object of attitude, frame changes how the characteristics of the object or object-related information are used, etc. As a result, in the case of an attitude framing, the attitudes seen as negative are more impactful on later evaluation, cognition, and behavior than positive attitudes are. In the case of framing of an attitude object, the negative frame increase impact of negative information concerned the object (as the arguments inconsistent with one's attitude can be seen from the point of view of the evaluative encoding account). It could be said, then, that negative frame in the case of both the valence framing and the forbid versus allow framing results in an increase in the influence of negative information. What is changing is whether these are primarily an attitude itself or an attitude object-related information.

Moreover, the effects connected with these two cases of framing can work simultaneously toward shaping final attitudes. We could expect that even with the forbid frame, the opponents should still be less susceptible to attitude change than the proponents. We have not found such asymmetry in our research, but the valence of initial attitude was not systematically manipulated here. However, as we mentioned earlier, some of our results confirm that the attitudes held by the opponents were stronger than the

proponents' attitudes, which is in accord with Bizer and Petty's findings (2005; see also Bizer et al., 2011, 2013).

Second, our results can shed some light on the potential mediators of the forbid/allow asymmetry known from the studies concerning the effect of forbid versus allow questions on answers in surveys (e.g., Chessa & Holleman, 2007; Hippler & Schwarz, 1986). In these survey research, participants usually answer the question relying only on their own attitudes. No additional information about the surveyed issue is provided for them. Our results suggest that providing participants with additional information about the surveyed issue should lower the extremity of the answers when the question about forbidding an object is answered. However, to our knowledge, this possibility has not been studied so far. It is thus a potential area for future research. It should also be noted that in our research, we found only one case of the forbid/allow asymmetry for the initial measurement of attitudes. The initial measurement of attitudes is a case similar to how this asymmetry was studied in the survey research. This result confirms that the forbid/allow asymmetry in survey answers is a quite elusive phenomenon (Holleman, 1999).

Third, our studies, especially Experiment 2 and Experiment 3, can be treated as an extension of research on attitude polarization. Some of the previous research found that when people are confronted with mixed evidence on some issue, their attitudes toward this issue tend to polarize (Lord et al., 1979). However, other studies found that this effect is not a very frequent phenomenon (Kuhn & Lao, 1996), and in some cases, attitudes can even depolarize (Miller et al., 1993). Stimuli materials presented to our participants in Experiment 2 and Experiment 3 were in fact a case of mixed evidence – a set of pro and con arguments without clear support for any of them. What we have found is in fact an attitude depolarization. Thus, our findings can add something to the understanding of the determinants of the phenomenon of attitude polarization versus depolarization. Specifically, the results showed in what conditions attitudes are prone to depolarization. It is possible that the forbid frame is a relatively easy way to make people *consider the opposite*, the strategy that was proven to be effective in reducing the biased evaluation of evidence and in preventing attitude polarization (Lord et al., 1984). This could be an important finding nowadays, when the growing polarization of attitudes in society is a serious problem. It looks like one way of depolarizing attitudes is to make people think about what should be or should not be forbidden, rather than making them think about what should be or should not be allowed.

Finally, we believe that the results obtained here are a consequence of the ability of the word “forbid” to induce some specific state where caution is evoked and previous actions or goals are reassessed. Some similarities between this state and the prevention focus can be observed (Higgins,

1998). This suggests the possibility that forbid versus allow frames are able to induce the prevention versus promotion focus. This, in turn, leads to the hypothesis that persuasive consequences of the prevention versus promotion focus are similar to those induced by the forbid versus allow framing. It is possible that the prevention (vs. promotion) focus leads people to process information more objectively and can result in attitude change. Similarly, Clark and Wegener (2013) suggested that a promotion orientation can lead to enhanced processing of proattitudinal information and prevention orientation can lead to enhanced processing of counterattitudinal ones. To our knowledge, this aspect of regulatory focus has not yet been studied empirically. It seems to be a promising area for future research. The practical importance of this new possibility is much larger than the present one stemming from our research.

Electronic Supplementary Materials

The electronic supplementary material is available with the online version of the article at <https://doi.org/10.1027/1864-9335/a000469>

ESM 1. Results of secondary importance (Experiments 1–3) and auxiliary measures (Experiment 3).

References

- Aerni, P., Scholderer, J., & Ermen, D. (2011). How would Swiss consumers decide if they had freedom of choice? Evidence from a field study with organic, conventional and GM corn bread. *Food Policy*, 36(6), 830–838. <https://doi.org/10.1016/j.foodpol.2011.08.002>
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology*, 5(4), 323–370. <https://doi.org/10.1037/1089-2680.5.4.323>
- Bizer, G. Y., Larsen, J. T., & Petty, R. E. (2011). Exploring the valence-framing effect: Negative framing enhances attitude strength. *Political Psychology*, 32(1), 59–80. <https://doi.org/10.1111/j.1467-9221.2010.00795.x>
- Bizer, G. Y., & Petty, R. E. (2005). How we conceptualize our attitudes matters: The effects of valence framing on the resistance of political attitudes. *Political Psychology*, 26(4), 553–568. <https://doi.org/10.1111/j.1467-9221.2005.00431.x>
- Bizer, G. Y., Žeželj, I. L., & Luguri, J. B. (2013). When are support and opposition not opposites? Depth of processing as a moderator of the valence-framing effect. *British Journal of Social Psychology*, 52(1), 191–201. <https://doi.org/10.1111/bjso.12014>
- Blankstein, K. R., Toner, B. B., & Flett, G. L. (1989). Test anxiety and the contents of consciousness: Thought-listing and endorsement measures. *Journal of Research in Personality*, 23(3), 269–286. [https://doi.org/10.1016/0092-6566\(89\)90001-9](https://doi.org/10.1016/0092-6566(89)90001-9)
- Brader, T. (2005). Striking a responsive chord: How political ads motivate and persuade voters by appealing to emotions.

- American Journal of Political Science*, 49(2), 388–405. <https://doi.org/10.1111/j.0092-5853.2005.00130.x>
- CBOS – Centrum Badań Opinii Społecznej. (2013). Polacy o bezpieczeństwie żywności i GMO [Poles on food safety and GMOs; Survey report]. https://www.cbos.pl/SPISKOM.POL/2013/K_002_13.PDF
- Chessa, A. G., & Holleman, B. C. (2007). Answering attitudinal questions: Modelling the response process underlying contrastive questions. *Applied Cognitive Psychology*, 21(2), 203–225. <https://doi.org/10.1002/acp.1337>
- Clark, J. K., & Wegener, D. T. (2013). Message position, information processing, and persuasion: The discrepancy motives model. *Advances in Experimental Social Psychology*, 47, 189–232. <https://doi.org/10.1016/B978-0-12-407236-7.00004-8>
- Clarkson, J. J., Tormala, Z. L., & Leone, C. (2011). A self-validation perspective on the mere thought effect. *Journal of Experimental Social Psychology*, 47(2), 449–454. <https://doi.org/10.1016/j.jesp.2010.12.003>
- Clarkson, J. J., Valente, M. J., Leone, C., & Tormala, Z. L. (2013). Motivated reflection on attitude-inconsistent information: An exploration of the role of fear of invalidity in self-persuasion. *Personality and Social Psychology Bulletin*, 39(12), 1559–1570. <https://doi.org/10.1177/0146167213497983>
- Cobb, M. D., & Kuklinski, J. H. (1997). Changing minds: Political arguments and political persuasion. *American Journal of Political Science*, 41(1), 88–121. <https://doi.org/10.2307/2111710>
- Cohen, G. L., Aronson, J., & Steele, C. M. (2000). When beliefs yield to evidence: Reducing biased evaluation by affirming the self. *Personality and Social Psychology Bulletin*, 26(9), 1151–1164. <https://doi.org/10.1177/01461672002611011>
- Edwards, K., & Smith, E. E. (1996). A disconfirmation bias in the evaluation of arguments. *Journal of Personality and Social Psychology*, 71(1), 5–24. <https://doi.org/10.1037/0022-3514.71.1.5>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/bf03193146>
- Ganzach, Y. (1995). Attribute scatter and decision outcome: Judgment versus choice. *Organizational Behavior and Human Decision Processes*, 62(1), 113–122. <https://doi.org/10.1006/obhd.1995.1036>
- Ganzach, Y., & Schul, Y. (1995). The influence of quantity of information and goal framing on decision. *Acta Psychologica*, 89(1), 23–36. [https://doi.org/10.1016/0001-6918\(94\)00004-Z](https://doi.org/10.1016/0001-6918(94)00004-Z)
- Griskevicius, V., Goldstein, N. J., Mortensen, C. R., Sundie, J. M., Cialdini, R. B., & Kenrick, D. T. (2009). Fear and loving in Las Vegas: Evolution, emotion, and persuasion. *Journal of Marketing Research*, 46(3), 384–395. <https://doi.org/10.1509/jmkr.46.3.384>
- Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Press.
- Higgins, E. T. (1998). Promotion and prevention: Regulatory focus as a motivational principle. *Advances in Experimental Social Psychology*, 30, 1–46. [https://doi.org/10.1016/S0065-2601\(08\)60381-0](https://doi.org/10.1016/S0065-2601(08)60381-0)
- Hippler, H.-J., & Schwarz, N. (1986). Not forbidding isn't allowing: The cognitive basis of the forbid-allow asymmetry. *Public Opinion Quarterly*, 50(1), 87–96. <https://doi.org/10.1086/268961>
- Holleman, B. (1999). Wording effects in survey research using meta-analysis to explain the forbid/allow asymmetry. *Journal of Quantitative Linguistics*, 6(1), 29–40. <https://doi.org/10.1076/jqul.6.1.29.4145>
- Holleman, B. C. (2006). The meanings of 'yes' and 'no'. An explanation for the forbid/allow asymmetry. *Quality and Quantity*, 40(1), 1–38. <https://doi.org/10.1007/s11135-005-4479-6>
- Houston, D. A., & Fazio, R. H. (1989). Biased processing as a function of attitude accessibility: Making objective judgments subjectively. *Social Cognition*, 7(1), 51–66. <https://doi.org/10.1521/soco.1989.7.1.51>
- Khetani, S. (2012, February 21). *Democracy in action? 10 weird examples of the Swiss referendum system in action*. <https://www.businessinsider.com/democracy-in-action-10-weird-examples-of-the-swiss-referendum-system-in-action-2012-2?IR=T>
- Koniak, P., & Cwalina, W. (2020). Fear of coronavirus and forbid/allow asymmetry as determinants of acceptance of COVID-19 pandemic related restrictions and persistence of attitudes towards these regulations. *Social Psychological Bulletin*, 15(4), 1–13. <https://doi.org/10.32872/spb.4421>
- Koniak, P., & Cwalina, W. (2022). *Data sheet file for the paper "Forbid/allow asymmetry in persuasion: The forbid frame decreases biased elaboration and increases attitude change"*. <https://doi.org/10.5281/zenodo.6091442>
- Kuhn, D., & Lao, J. (1996). Effects of evidence on attitudes: Is polarization the norm? *Psychological Science*, 7(2), 115–120. <https://doi.org/10.1111/j.1467-9280.1996.tb00340.x>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Laran, J., & Wilcox, K. (2011). Choice, rejection, and elaboration on preference-inconsistent alternatives. *Journal of Consumer Research*, 38(2), 229–241. <https://doi.org/10.1086/659040>
- Laurin, K., Kay, A. C., & Fitzsimons, G. J. (2012). Reactance versus rationalization: Divergent responses to policies that constrain freedom. *Psychological Science*, 23(2), 205–209. <https://doi.org/10.1177/0956797611429468>
- Levin, I. P., & Gaeth, G. J. (1988). How consumers are affected by the framing of attribute information before and after consuming the product. *Journal of Consumer Research*, 15(3), 374–378. <https://doi.org/10.1086/209174>
- Levin, I. P., Schneider, S. L., & Gaeth, G. J. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes*, 76(2), 149–188. <https://doi.org/10.1006/obhd.1998.2804>
- Lord, C. G., Lepper, M. R., & Preston, E. (1984). Considering the opposite: A corrective strategy for social judgment. *Journal of Personality and Social Psychology*, 47(6), 1231–1243. <https://doi.org/10.1037/0022-3514.47.6.1231>
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, 37(11), 2098–2109. <https://doi.org/10.1037/0022-3514.37.11.2098>
- McElroy, T., & Seta, J. J. (2003). Framing effects: An analytic-holistic perspective. *Journal of Experimental Social Psychology*, 39(6), 610–617. [https://doi.org/10.1016/S0022-1031\(03\)00036-2](https://doi.org/10.1016/S0022-1031(03)00036-2)
- Miller, A. G., McHoskey, J. W., Bane, C. M., & Dowd, T. G. (1993). The attitude polarization phenomenon: Role of response measure, attitude extremity, and behavioral consequences of reported attitude change. *Journal of Personality and Social Psychology*, 64(4), 561–574. <https://doi.org/10.1037/0022-3514.64.4.561>
- Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. Springer.
- Pomerantz, E. M., Chaiken, S., & Tordesillas, R. S. (1995). Attitude strength and resistance processes. *Journal of Personality and Social Psychology*, 69(3), 408–419. <https://doi.org/10.1037/0022-3514.69.3.408>
- Rucker, D. D., Petty, R. E., & Briñol, P. (2008). What's in a frame anyway? A meta-cognitive analysis of the impact of one versus two sided message framing on attitude certainty. *Journal of Consumer Psychology*, 18(2), 137–149. <https://doi.org/10.1016/j.jcps.2008.01.008>
- Rugg, D. (1941). Experiments in wording questions: II. *Public Opinion Quarterly*, 5(1), 91–92. <https://doi.org/10.1086/265467>
- Rzyski, P., & Królczyk, A. (2016). Attitudes toward genetically modified organisms in Poland: To GMO or not to GMO? *Food Security*, 8(3), 689–697. <https://doi.org/10.1007/s12571-016-0572-z>

- Schuman, H., & Presser, S. (1981). *Questions and answers in attitude surveys: Experiments on question form, wording and context*. Sage Publications.
- Shafir, E. (1993). Choosing versus rejecting: Why some options are both better and worse than others. *Memory & Cognition*, 21(4), 546–556. <https://doi.org/10.3758/bf03197186>
- Shafir, E., Simonson, I., & Tversky, A. (1993). Reason-based choice. *Cognition*, 49(1–2), 11–36. [https://doi.org/10.1016/0010-0277\(93\)90034-S](https://doi.org/10.1016/0010-0277(93)90034-S)
- Shields, M. (2018, September 23). *Swiss canton becomes second to ban burqas in public*. <https://www.reuters.com/article/us-swiss-burqas/swiss-canton-becomes-second-to-ban-burqas-in-public-idUSKCN1M30J2>
- Sierpińska, A. M. (2007). *Zakaz handlu po raz pierwszy 1 listopada 2007 r* [Trading ban for the first time November 1, 2007]. <http://www.prawo.egospodarka.pl/25608,Zakaz-handlu-po-raz-pierwszy-1-listopada-2007-r,1,34,3.html>
- Sokolova, T., & Krishna, A. (2016). Take it or leave it: How choosing versus rejecting alternatives affects information processing. *Journal of Consumer Research*, 43(4), 614–635. <https://doi.org/10.1093/jcr/ucw049>
- Szadowska-Szlachetka, Z. C., Antoniuk, K., Łuczyk, M., Ślusarska, B., Stanisławek, A., Nowicki, G., & Muraczyńska, B. (2019). Knowledge and attitudes of Medical University students with regard to euthanasia. *Palliative Medicine*, 11(2), 73–80. <https://doi.org/10.5114/pm.2019.86531>
- Taber, C. S., Cann, D., & Kucsova, S. (2009). The motivated processing of political arguments. *Political Behavior*, 31(2), 137–155. <https://doi.org/10.1007/s11109-008-9075-8>
- Voters approve 'burka ban' in St Gallen. (2018, September 23). https://www.swissinfo.ch/eng/directdemocracy/face-look_voters-approve-burka-ban-in-st-gallen/44419680
- Wedell, D. H. (1997). Another look at reasons for choosing and rejecting. *Memory & Cognition*, 25(6), 873–887. <https://doi.org/10.3758/BF03211332>
- Wierzbica, M., Riegel, M., Wypych, M., Jednoróg, K., Turnau, P., Grabowska, A., & Marchewka, A. (2015). Basic emotions in the nencki affective word list (NAWL BE): New method of classifying emotional stimuli. *PLoS One*, 10(7), e0132305–16. <https://doi.org/10.1371/journal.pone.0132305>
- Wyer, R. S., Budesheim, T. L., Shavitt, S., Riggall, E. D., Melton, R. J., & Kuklinski, J. H. (1991). Image, issues, and ideology: The processing of information about political candidates. *Journal of Personality and Social Psychology*, 61(4), 533–545. <https://doi.org/10.1037/0022-3514.61.4.533>
- Wyler, D. (2017). The Swiss minaret ban referendum and Switzerland's international reputation: A vote with an impact. *Journal of Muslim Minority Affairs*, 37(4), 413–425. <https://doi.org/10.1080/13602004.2017.1405506>

History

Received October 20, 2020

Revision received December 2, 2021

Accepted December 12, 2021

Published online February 28, 2022

Acknowledgments

The authors thank M. Drzewiecka, A. Błachnio, A. Przepiórka, P. Próchniak, P. Długoński, M. Płaczkowski, A. Milanowska, S. Niedźwiedz, A. Pałac, K. Piotrowska, K. Rosół, J. Siejka, I. Skoczek, W. Strzałkowska, and J. Tomczyk for help with recruiting participants.

Conflict of Interest

The authors declare that there are no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Publication Ethics

The authors confirm that the manuscript adheres to ethical guidelines specified in the APA Code of Conduct as well as authors' national ethics guidelines.

Authorship

Both authors were involved in all parts of the research.

Open Data

Data for Studies 1–3 are available at <https://doi.org/10.5281/zenodo.6091442>. Study 3 was preregistered at <https://aspredicted.org/2c56v.pdf>.

Funding

This work was supported by the Narodowe Centrum Nauki (Grant 2013/09/D/HS6/02823 to Paweł Koniak).

ORCID

Paweł Koniak

 <https://orcid.org/0000-0001-7795-8549>

Paweł Koniak

Maria Curie-Skłodowska University
Department of Social Psychology
ul. Głęboka 45
20-612 Lublin
Poland
pawel.koniak@mail.umcs.pl



Members of Transgressor Groups Prefer Reparations to Come From Third Parties

An Extension of Interpersonal Findings on Moral Emotions

Maciej Sekerdej¹ and Roger Giner-Sorolla²

¹Institute of Psychology, Jagiellonian University, Krakow, Poland

²School of Psychology, University of Kent, Canterbury, UK

Abstract: Interpersonal research has shown that guilt motivates perpetrators to compensate victims at the expense of a third party, indicating that the emotion's goal involves reparative outcomes rather than self-mortification. However, this motivation has yet to be tested in an intergroup context. Based on findings about ingroup wrongdoing, compensation was expected to draw primarily on shame rather than guilt. Three experiments (total $N = 617$) showed that participants only allocated more to the victims versus their own group when compensation was assigned from a third party's rather than their own group's resources. There was also evidence that shame was felt, and related to compensating victims, more strongly than guilt, whether at the expense of the ingroup or the third party.

Keywords: shame, guilt, moral emotions, prosocial behavior, intergroup relations

Although self-critical emotions such as guilt and shame are not pleasant, they can at times increase prosocial behavior. Experimental studies have shown that guilt feelings in particular increase cooperation between persons (de Hooge et al., 2007; Nelissen et al., 2007; Regan et al., 1972), promote ethical behavior in economic decisions (Cohen et al., 2012), and motivate people to compensate for their own wrongs (Yu et al., 2014; Zeelenberg & Breugelmans, 2008).

More recent studies, however, have shown that the prosocial consequences of interpersonal guilt do not always entail taking a personal disadvantage and may end up displacing rather than alleviating wrongs on a larger scale (de Hooge et al., 2011). In these studies, induced guilt motivated repairing the harm done to the victim. But, if perpetrators had an opportunity to compensate the victim at the expense of others, they willingly took it, in preference to giving compensation from their own resources. In other words, prosocial actions motivated by guilt focused on the victim rather than on the larger picture, without the necessity of self-deprivation from paying reparations personally. An action that remedies one injustice merely to alleviate guilt could thus plausibly create another injustice.

Because guilt often involves self-punishment (Griffin et al., 2016; Vel-Palumbo et al., 2018), it may seem at first glance

that repayment without suffering would not be the preferred response to guilt. Even if self-punishment depletes one's own resources, it may be chosen to send a signal that one is a morally motivated member of the community. But instead, de Hooge et al. showed that self-sacrifice is often passed over when restitution can be made at another's expense. Thus, the goal of guilt in this context was revealed to be compensation of the victim by any means, rather than self-mortification *per se*.

Our main question of interest in the present research was whether this preference for third-party restitution, until now only demonstrated in interpersonal contexts, could also be shown in collective situations involving the rights and wrongs of nations. Emotions can be felt on an *intergroup* level when a person appraises a collectively self-relevant situation involving larger groups, such as nations (Bonnot & Krauth-Gruber, 2018; de Rivera et al., 2007; Doosje et al., 1998; Reysen et al., 2014). It is therefore possible that, in an intergroup situation of historical or ongoing injustice perpetrated by one's own group, people would likewise prefer to compensate the victim group at the expense of a third party, when given an opportunity. Such a finding would be particularly relevant to international situations in which allied states or transnational groups, such as the European Union (EU) or United Nations (UN), can relieve a perpetrator nation of the costs of compensation.

Shame as a Prosocial Agent in Intergroup Context

Our secondary question of interest asked which emotions in particular might be felt most strongly and predict behavior in such a group-based situation. de Hooge et al. (2011, Experiment 4) found that third-party compensation followed an induction of guilt, but not one of shame. Shifting contexts from personal to group-based culpability, however, there is a reason to believe that shame would take on more importance than guilt. This prediction follows from the literature on shame in intergroup reconciliation and prosocial attitudes.

Overall, shame has been identified as an individual prosocial motivator, but not as consistently as for guilt. Many findings suggest that, for individuals, shame can fuel avoidance tendencies rather than reparation (e.g., Tangney, 1995; Tangney & Fischer, 1995; Wicker et al., 1983). Nevertheless, shame can also motivate people to act prosocially (Gausel et al., 2012, 2017; Goldberg, 1991) and enhance social commitment (de Hooge et al., 2008). The prosocial versus antisocial consequences of shame may depend on such factors as the cultural value placed upon shame (Sheikh, 2014) and the possibility to take effective action (Leach & Cidam, 2015).

We suggest that shame is also more viable as a morally relevant emotion when felt in a group versus individual context. A retrospective analysis by Nelissen et al. (2013) argued that guilt is more likely to motivate prosocial behavior in a context of direct reciprocity, when it is likely that the victim and perpetrator will meet again. Shame, on the other hand, is more likely to motivate prosocial behavior in indirect reciprocity, when it is much less likely that the victim and perpetrator will meet again. Indirect reciprocity is thus more characteristic of the intergroup context, in which the approval of a relevant audience, but not necessarily of specific victims, is sought. Another line of research has shown that in intergroup contexts, when negative behavior is exposed to other groups triggering collective blame (in terms of collective guilt and shame, this time treated as a single factor), group members behave more prosocially toward outgroup members (Kardos et al., 2019). However, if an individual, as part of a nation-level ingroup, sees a collective failure in reciprocity to another outgroup by harming or disadvantaging it, shame would play an even larger role than guilt because the people harmed are even more remotely and indirectly connected to the individual.

Indeed, studies of emotions in social context show this predicted link between shame and higher-level social relations. Lickel et al. (2005), in studies of vicarious guilt and shame for another person's wrongdoings, showed that people feel guilty when they have a highly interdependent association with the perpetrator but felt ashamed when the

perpetrator's actions were relevant to the social identity that they shared with them and could tarnish the identity's image. In another study, Johns et al. (2005) found that feeling vicarious shame was associated with stronger ingroup identification: Participants highly identified with US nationality felt more ashamed at examples of US anti-Arab prejudice.

Research on collective emotions felt toward the wrongdoing of one's own larger group (e.g., nation) also tends to find that shame is both more prevalent and more effective than guilt (Rees et al., 2013; Shepherd et al., 2013a). Some studies show that shame lends stronger support than guilt does to collective actions against a prospective ingroup transgression (Shepherd et al., 2013b). Also, higher expectations and satisfaction result from perpetrators' expressions of shame versus guilt among victim group members (Giner-Sorolla et al., 2008; Kamau et al., 2013). Moreover, in the national context, shame has turned out to be stronger than guilt in instigating support for collective political actions aimed at unjust conflict (Iyer et al., 2007), collective apology and victim compensation (Allpress et al., 2010), and motivation for ingroup change (Gausel & Brown, 2012).

Furthermore, shame has also been shown to play an important role in shaping reparative attitudes for the damage done to others. For example, it has been found that both collective guilt and shame are associated with support for reparations, although the underlying motivations were somewhat different: The effects of shame were mediated by self-pity and empathy for the outgroup, while the effects of guilt were mediated by empathy, but not self-pity (Brown & Čehajić, 2008). Notably, in other intergroup research, guilt predicted stable, longitudinal reparation attitudes, while shame predicted rather short-term, cross-sectional prosocial tendencies (Brown et al., 2008). In conclusion, these authors speculated that shame leads people to *take the path of least resistance* to the goal of improving the ingroup's reputation. This prediction in particular further leads us to expect that group members, when reminded of responsibility for collective wrongs, would prefer reparations to be paid by a third party. Avoiding the depletion of ingroup resources by using another entity's resources would be consistent with the *path of least resistance* typical of intergroup shame.

The Present Research

To restate, this research tested two hypotheses:

Hypothesis 1 (H1): As in interpersonal research, perpetrator group members will be more strongly motivated to support reparative actions at the expense of a third party, rather than sending a signal at their own group's expense, when the situation allows.

Hypothesis 2 (H2): In the intergroup context, prosocial behavior toward victims regardless of the source of funds will be related to ingroup-critical emotions, and where these can be distinguished, mainly to shame rather than guilt. This can be shown in two ways:

- a. There should be higher mean levels of shame versus guilt, showing shame to be the predominant feeling. This advantage of shame should be especially strong in conditions where the ingroup was responsible for a misdeed, showing sensitivity to ingroup-caused wrongs.
- b. Reparations funded by an outgroup should be correlated with levels of moral emotions just as much as the literature has shown for reparations funded by the ingroup, especially when the ingroup is responsible.

We conducted three experiments, the first two in the context of Polish collective responsibility for contemporary injustices and the third in the context of US collective responsibility for military intervention in the Dominican Republic (DR). In Experiment 1, we show initial evidence that ingroup members are more motivated to compensate outgroup victims if provided with the opportunity to use someone else's resources, drawing on a larger collective – the resources of the EU. We also show that collective emotions still play a role in these motivations, but that there is a more prominent role for shame rather than guilt. In Experiment 2, we replicated those findings and extended them to a fuller experimental design. Experiment 3 further extended the findings, modeled after Experiment 2 but using a US sample and issue, and the UNs as the third-party option.

Experiment 1

Method

Participants and Design

The participants were 140 Polish citizens (103 female, 37 male) who were recruited via community advertisements and social portals (September 2016). Their age ranged from 18 to 56 years ($M = 25.31$, $SD = 4.94$). All participants signed a written informed consent form before starting the experiment. Sample size was determined prior to analysis.

For analyses involving the comparison of guilt and shame, the design was mixed factorial, 2 (Condition, between: EU vs. Polish source of reparations) \times 2 (Emotion, within: guilt vs. shame). For analyses involving distribution of compensation, the design was mixed factorial, 2 (Condition, as before) \times 3 (Compensation Recipient, within: Poland, victims, global institutions).

In sensitivity power analyses, the emotion design with $N = 140$ had 80% power to detect a medium effect size of $f = 0.24$. The compensation design had 80% power to detect a medium-to-small effect size of $f = 0.19$ based on the nonsphericity correction factor (epsilon) in the data, which was .91 (nonsphericity being relevant only to a design with three or more levels in a factor). All data reported in the paper along with the Supplemental Material are made publicly available at the Open Science Framework (https://osf.io/ag2dv/?view_only=d65f6434d510443ba2990c).

Procedure

The experiment took place online. Participants were asked if they would like to fill out a questionnaire on social relations. Having given informed consent, participants read a short newspaper article, ostensibly sourced from Polish Press Agency (see Supplemental Material, Sekerdej & Giner-Sorolla, 2021). The article gave an account of the situation of refugees in Poland, describing Poles' negative opinions on refugees and some discriminatory acts against them. The text was followed by measures of guilt, moral shame, and image shame. Finally, the participants were asked to distribute money, ostensibly provided by Poland or the EU, depending on condition.

Materials

The measures of group-based shame and guilt were adapted from Allpress et al. (2014). Participants indicated on scales ranging from 1 (*strongly disagree*) to 6 (*strongly agree*) how much they agreed or disagreed with the statements.

These items distinguish between a more internalized “moral shame” and a more externally concerned “image shame” felt toward an ingroup that has committed intergroup transgressions. In two out of three studies in Allpress et al. (2014), moral shame (e.g., believing that the group is fundamentally bad as a result of the misdeeds) related positively to support for compensation and apology, while image shame (e.g., believing that the group looks bad to others as a result of the misdeeds) had a negative or null relation to compensation and apology, controlling for moral shame. These findings initially suggested to us that moral shame items, if found to be statistically distinct from image shame, could more specifically relate to support for compensation from third parties as well as one's own group.

Moral shame was measured by three items: “I do feel ashamed to be Polish for the way we have treated other people,” “I feel ashamed for the opinions about immigrants uttered by Polish people, because they are immoral,” and “I feel ashamed for the damage resulting from the attitudes of Poles towards immigrants” ($M = 3.63$, $SD = 1.45$, $\alpha = .87$).

Image shame was measured by three items: “I feel disgraced because the behaviour of Polish people toward

immigrants has created a *bad image* of Poland among other EU countries,” “I feel humiliated when I think of how Poland is *seen* negatively by the rest of the world for how it has treated immigrants,” and “To think how Poles are *seen* for their treatment of immigrants makes me feel ashamed” ($M = 3.55$, $SD = 1.50$, $\alpha = .90$).

Guilt was measured by three items: “I feel guilty for the manner in which Poles have treated immigrants,” “Even if I have done nothing bad, I feel guilty for the behaviour of Poles in this situation,” and “I feel guilty for the bad living conditions of immigrants” ($M = 2.51$, $SD = 1.32$, $\alpha = .93$).

Because the two shame variables were correlated so highly as to be almost indistinguishable ($r = .79$) and had very close means, in further analyses, moral shame and image shame were collapsed into one variable, *shame* ($M = 3.59$, $SD = 1.40$, $\alpha = .93$).

Compensation Task

Participants were given three spaces to fill in. They read: “Because of the current wave of immigrants, the European Union [Polish government] assigned € 1,000,000 to be allocated to three purposes. Please write what percent of this sum you would give to: (1) victims (the refugees that experienced harm from Poles); (2) Poland, to establish a special fund to secure future financial obligations; (3) Global institutions that help refugees.”

Results and Discussion

To first test the main hypothesis in the novel intergroup setting, we ran a mixed model 2×3 ANOVA (Compensation Source, between: Poland vs. EU \times Recipient, within: victims vs. Poland vs. global institutions).

Descriptive statistics for money distribution to victims, Poland, and global institutions in each source condition are provided in Table 1. The main effect of source condition was not significant, $F(1, 138) = 0.37$, $p = .55$, $\eta_p^2 = .003$, but we found significant effects of recipient, $F(2, 138) = 6.84$, $p < .01$, $\eta_p^2 = .05$, and a significant interaction between recipient and compensation source, $F(2, 138) = 8.97$, $p < .001$, $\eta_p^2 = .06$. As expected, pairwise comparisons within the estimated marginal means test showed that when the money came from the EU versus Poland, participants gave significantly more to the victims ($M = 35.53$, $SE = 2.41$ vs. $M = 24.29$, $SE = 2.45$, $p = .001$, Cohen’s $d = 0.55$), suggesting that ingroup members were willing to pay off people the ingroup has harmed but would rather do it at someone else’s expense. By the same token, when the money came from Poland versus

Table 1. Amounts of money assigned to victims, Poland, and global institutions in the two funding source conditions in Experiment 1

		Mean	95% CI for mean		SD
			Lower bound	Upper bound	
Source: EU	Victims	35.53	30.76	40.31	20.88
	Poland	32.65	26.77	38.51	22.94
	Global institutions	31.80	27.21	36.39	20.27
Source: Poland	Victims	24.29	19.45	29.13	19.79
	Poland	47.61	41.66	53.55	26.91
	Global institutions	28.07	23.41	32.73	19.59

Note. All figures are percentages of total amount.

EU, participants gave significantly more to Poland ($M = 47.61$, $SE = 3.01$ vs. $M = 32.65$, $SE = 2.96$, $p = .001$, Cohen’s $d = 0.60$); the amount of money given to global institutions did not differ across conditions ($M = 31.80$, $SE = 2.32$ vs. $M = 28.07$, $SE = 2.36$, $p = .26$, Cohen’s $d = 0.19$). Overall, in the EU condition, the distribution of money between Poland, victims, and global institutions was not different (ps between means from .31 to .86). However, when the money came from Poland, participants kept significantly more for Poland than they gave either to victims or to global institutions (all $ps < .001$). Note that in this situation, allocation of money to victims and Poland was not completely confounded due to the third option of global institutions. But due to the nature of the task, there would of course be a natural inverse relationship between money given to one recipient and another.

Next, we ran a mixed model 2×2 ANOVA comparing the two self-conscious emotions of interest (Compensation Source, between: Poland vs. EU \times Emotion, within: Guilt vs. Shame). This revealed that participants, having all read the text that put their ingroup in a bad light, felt shame more strongly than guilt, $M = 3.59$ versus $M = 2.51$, $F(1, 138) = 107.36$, $p < .001$, $\eta_p^2 = .43$. However, the overall level of emotions was not significantly different between conditions, $F(1, 138) = 0.25$, $p = .62$, $\eta_p^2 = .002$, and the interaction was not significant, $F(1, 138) = 2.50$, $p = .12$, $\eta_p^2 = .02$.¹

Second, we checked the correlations between shame, guilt, and the compensation outcomes, that is, the percent of money given to the victims, Poland, and global institutions in each condition (see Table 2). When the EU was paying, shame correlated positively with the percentage of money assigned to victims and significantly more strongly than guilt did: test between correlations, $t(68) = 2.49$, $p = .02$. Also when the EU was paying, shame correlated

¹ We repeated the same analyses for image and moral shame separately obtaining similar results, respectively, $F(1, 138) = 0.39$, $p = .53$, $\eta_p^2 = .003$ and $F(1, 138) = 0.12$, $p = .73$, $\eta_p^2 = .001$ (condition) and $F(1, 138) = 2.60$, $p = .11$, $\eta_p^2 = .02$ and $F(1, 138) = 1.77$, $p = .19$, $\eta_p^2 = .01$ (interaction).

Table 2. Experiment 1 correlations between shame, guilt, and percent of money given to the victims, Poland, and global institutions, divided by the source of money: EU (above diagonal) and Poland (below)

	1	2	3	4	5
1. Shame		.67**	.30*	-.28*	.02
2. Guilt	.49**		.07	-.10	.04
3. Money assigned to victims	.48**	.24*		-.58**	-.38**
4. Money assigned to Poland	-.33*	-.23+	-.72**		-.54**
5. Money assigned to global institutions	-.03	.07	-.03	-.68**	

Note.* $p < .05$, ** $p < .01$, + $p = .06$.

negatively with money assigned to Poland and marginally more strongly than guilt did, $t(68) = 1.90$, $p = .059$. When Poland was paying, shame also correlated more strongly than guilt with the money assigned to victims, $t(66) = 2.18$, $p = .03$, but there was no difference between shame and guilt in (negatively) predicting the money given back to Poland, $t(66) = 0.85$, $p = .39$. Overall, regardless of the source, shame more so than guilt was associated with intention to pay back victims.

In summary, Experiment 1 showed that that group members were motivated to compensate victims harmed by their ingroup, but they preferred to do it while avoiding strain on ingroup resources by calling on a superordinate group to pay. Indeed, if another entity was paying, they supported paying the victims more than if the reparations would have been paid from their own group resources. However, even when the other entity was paying, they still decided to keep a decent share for their own group, while also allocating a similar sum to global institutions described as helping refugees.

The situation changed when reparations had to be paid from the group's own resources. Here, the group members decided to keep a plurality of the money for the ingroup. The rest was distributed approximately equally between the victims and the third party (global institutions), which was described as ultimately helping refugees in general. Such a distribution suggests that when group members are deprived of the possibility to use external resources and forced to rely on their own, they are less prone to choose the reparation option.

Finally, in this *indirect reciprocity* context (Nelissen et al., 2013), shame more so than guilt was related to reparative decisions. Levels of shame were higher overall than guilt, against the background of ingroup wrongdoing. Moreover, shame, significantly and more strongly than guilt, was related to assigning more money to victims in both conditions, and less money to the ingroup in the condition where the EU paid.

In Experiment 2, we replicated these results in a fuller experimental plan. We added a between-participants factor that introduced control conditions in which the ingroup was not responsible for wrongdoing to test

whether moral emotions would play a similar role. If the preference for others to pay reparations were shown to be the same in both conditions, this would further clarify that the effect simply consists of people applying their selfishness even to situations where they are collectively responsible. This would mean that heightened responsibility did nothing to encourage people to punish the ingroup, above and beyond helping the victims. However, if we were to observe a relative increase in using third-party versus own resources when group responsibility was not salient, this would show that there was some tendency for reparation preferences to reflect collective self-sanctioning.

Furthermore, in conditions without ingroup responsibility, we expected no link between self-critical emotions and outgroup giving, unlike in Experiment 1, when all participants read about ingroup responsibility. This is because, without the involvement of the ingroup as perpetrators, there would be no reason for guilt or shame to motivate any kind of compensation.

Experiment 2

Method

Participants and Procedure

The participants were 129 Polish students (110 female, 37 male) recruited via community advertisements and social portals. Their age ranged from 18 to 30 years ($M = 21.21$, $SD = 2.29$). All participants signed a written informed consent before starting the experiment. The design was 2 (Responsibility: Ingroup Responsibility vs. Control) \times 2 (Compensation Source: Poland vs. EU) between-subjects, to which could be added an additional within-subjects variable as in Experiment 1, compensation recipient (victims, Poland, global institutions). In a sensitivity post hoc power analysis, our design had 80% sensitivity given the observed nonsphericity correction of .84 to detect a medium-sized effect ($f = 0.21$) smaller than the Condition \times Recipient interaction in Experiment 1 ($f = 0.25$, converted from partial $\eta^2 = .06$).

Manipulation

In the *ingroup responsibility* experimental condition, the participants read a newspaper article (ca. 1,500 words), ostensibly sourced from a Polish Press Agency, that gave an account of the situation of refugees in Poland, including Poles' negative opinions on refugees and descriptions of discriminatory acts against them. In the *neutral* control condition, the participants read an article of the same length (based on excerpts retrieved from Eurostat, 2018), attributed to the same press agency, that reported general statistics on migration in Europe (see ESM 1). The text was concerned with different European countries, including Poland, and was neutral in tone. Both texts were followed by the measure of emotions toward the situation of immigrants in Poland (manipulation check) and the other dependent variables specifically measuring guilt, moral shame, and image shame. Finally, as in Experiment 1, the participants were asked to distribute money coming from either Poland or the EU among the victims, Poland, and global institutions.

Materials

Following the manipulation of ingroup responsibility, we measured on a 5-point scale, 10 emotions toward the situation of immigrants in Poland – six negative: regret, guilt, disgust, shame, anger, and embarrassment; and four positive: rapture, happiness, pride, and enthusiasm.

More specific emotions were measured by the same items as in Experiment 1 on 6-point scales ranging from 1 (*strongly disagree*) to 6 (*strongly agree*): *image shame* ($M = 3.23$, $SD = 1.67$, $\alpha = .95$), *moral shame* ($M = 3.49$, $SD = 1.56$, $\alpha = .89$), and *guilt* ($M = 2.73$, $SD = 1.53$, $\alpha = .94$). The *distribution task* was also the same as in Experiment 1 and came last.

Results and Discussion

To again test the main hypothesis about compensation, we conducted a mixed model $2 \times 2 \times \text{three}$ ANOVA (Responsibility, between: Ingroup Responsibility vs. Control; \times Compensation Source, between: Poland vs. EU; \times Recipient, within: victims vs. Poland vs. global institutions). The main effects of responsibility, compensation source, and the interaction between them were nonsignificant, $F(1, 125) = 1.17$, $p = .28$, $\eta_p^2 = .009$, $F(1, 125) = 1.22$, $p = .27$, $\eta_p^2 = .01$, and $F(1, 125) = 0.70$, $p = .40$, $\eta_p^2 = .006$. Moreover, exactly as in Experiment 1, we found a significant main effect of recipient, $F(2, 125) = 8.79$, $p < .001$, $\eta_p^2 = .07$, and a significant interaction between recipient and compensation source, $F(2, 125) = 7.15$, $p = .001$, $\eta_p^2 = .05$. The three-way interaction between ingroup responsibility, compensation source, and recipient turned out to be nonsignificant $F(2, 125) = 1.01$, $p = .37$, $\eta_p^2 = .008$.

As expected, pairwise comparisons within the estimated marginal means test showed that when the money came from the EU versus Poland, participants gave significantly more to the victims ($M = 41.60$, $SE = 2.34$ vs. $M = 28.55$, $SE = 2.29$, $p < .001$, Cohen's $d = 0.70$), which confirmed the findings from Experiment 1. Similarly, when the money came from Poland versus EU, participants gave more to Poland ($M = 41.62$, $SE = 2.80$ vs. $M = 34.43$, $SE = 2.87$, $p = .075$, Cohen's $d = 0.36$); the amount of money given to global institutions did not differ across conditions ($M = 28.53$, $SE = 1.90$ vs. $M = 24.13$, $SE = 1.95$, $p = .11$, Cohen's $d = 0.29$). Overall, in the EU condition, victims received a larger amount of money than Poland received (although not significantly, $p = .14$). In turn, when the money came from Poland, participants kept significantly more for Poland than they gave to both victims and global institutions (all $ps < .01$).

Specifically, pairwise comparisons between the money assigned to victims and Poland in all four conditions of the 2×2 (Responsibility, Ingroup Responsibility vs. Control \times Compensation Source, Poland vs. EU) design revealed that the only condition where victims received substantially more money than Poland was the combination where Poles were described as responsible for the outgroup's suffering and the money came from the EU (Table 3). The preference for victims in this condition was marginally significant, $p = .052$.

The findings replicated the findings of Experiment 1 in the responsibility conditions. However, they also showed a weaker source effect in the control conditions, where Poland had been described with no particular responsibility, but participants still relatively preferred to help outgroups with EU, not Polish, money. To further clarify the predicted motivational differences between these giving effects in the responsibility and control conditions, we looked at mean levels of emotions and their correlation with the giving variables.

All negative emotions were significantly higher, and all positive emotions were significantly lower in ingroup responsibility conditions than in control conditions (see Table 4).

Table 5 and Table 6 show correlations among the analyzed variables by condition. Due to overly high correlations between image shame, moral shame, and guilt ranging from .74 to .87 (i.e., showing an overlap of over 50% of variance), it was not considered advisable to look at them separately in correlational analyses, so for those purposes, they were collapsed into one variable, *shame/guilt* ($M = 3.15$, $SD = 1.47$, $\alpha = .92$). However, because correlations are based on standardized scores, mean differences between (for example) shame and guilt can exist despite these overly high correlations, and so could be compared in a repeated-measures ANOVA.

Table 3. Pairwise comparisons between means of the money distribution to victims and Poland in the four conditions of Experiment 2

Condition	Recipient				
	Victims		Poland		p
	M	SE	M	SE	
Poles responsible/money from the EU	45.78	3.29	32.50	4.03	.052
Poles responsible/money from Poland	28.49	3.24	38.85	3.40	.123
Control condition/money from the EU	37.42	3.34	36.36	4.09	.877
Control condition/money from Poland	28.61	3.24	44.39	3.97	.019

To test this effect on the means, we conducted a mixed model ANOVA, 2 (Responsibility, between: Ingroup Responsibility vs. Control) \times 2 (Compensation Source, between: Poland vs. EU), \times 2 (Emotion, within: Guilt vs. Shame, moral and image collapsed into one measure, as in Experiment 1). The main effect of responsibility was marginally significant, $F(1, 125) = 3.60, p = .06, \eta_p^2 = .03$, the trend being to feel more guilt and shame when the ingroup was responsible, which is in line with findings from previous research showing higher negative ingroup-focused emotions when collective responsibility is accepted. The effect of source and the interactions were nonsignificant. Nevertheless, the difference between felt shame and guilt (within-subject test across all conditions) turned out to be significant, $F(1, 125) = 44.57, p < .001, \eta_p^2 = .26$. As in Experiment 1, shame was the overall stronger emotion than guilt.

For emotion correlations, shame and guilt were averaged into one index. In the responsibility condition, shame/guilt was correlated strongly with both the money given to victims (positively) and the money given to Poland (negatively), which confirms the overall important role of moral emotions in reparative actions. This was true regardless of the source of funds, supporting the idea that

compensation from other groups' money is also related to guilt and shame. In control conditions, moral emotions were not related to money distribution to victims, supporting the idea that responsibility is necessary to motivate compensation via emotions. The negative correlations among distribution scores, as in Experiment 1, generally reflect the zero-sum nature of the task.

Experiment 3

To extend these findings to a different intergroup context and participant population, we conducted a conceptual replication of Experiment 2 among US citizen participants. This new experiment also tightened the manipulation of ingroup responsibility by eliminating some elements in the Experiment 2 manipulation that might have introduced confounds not relevant to collective responsibility, such as the appeal to empathy for the refugees in the responsible condition, or the contrast between personal narratives in the responsible condition and dry statistics in the control. Also, we pre-registered our analytic method and increased

Table 4. Means of positive and negative emotions in ingroup responsibility and control conditions, Experiment 2

	Ingroup responsibility condition		Control condition			
Emotions	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>F</i>	η_p^2
Negative emotions						
Regret	3.48	1.10	2.59	1.09	21.06**	.14
Guilt	2.55	1.26	1.90	1.12	9.46*	.07
Disgust	3.21	1.36	2.17	1.14	21.83**	.15
Shame	3.25	1.36	1.98	1.21	30.71**	.20
Anger	2.94	1.26	2.21	1.12	12.0**	.09
Embarrassment	3.23	1.33	2.03	1.12	30.27**	.19
Positive emotions						
Rapture	1.40	0.61	1.75	0.78	7.84*	.06
Happiness	1.41	0.58	1.94	0.75	20.52**	.14
Pride	1.29	0.49	1.76	0.78	16.81**	.12
Enthusiasm	1.52	0.75	2.09	0.96	14.09**	.10

Note. * $p < .01$, ** $p < .001$.

Table 5. Experiment 2, ingroup responsibility condition: correlations among shame/guilt, and percent of money given to the victims, Poland, and global institutions, by the source of money: EU (above diagonal) and Poland (below)

	1	2	3	4
1. Shame and guilt		.57**	-.52**	-.04
2. Money assigned to victims	.53**		-.87**	-.18
3. Money assigned to Poland	-.46**	-.65**		-.32
4. Money assigned to global institutions	.09	.17	-.60**	

Note. ** $p < .01$.

the statistical power of this study relative to Experiment 2, nearly tripling the number of participants to reach an *a priori* power of 90% to detect the smallest key effect found in Experiment 2. Because of the higher power, this study had the potential to more definitively follow up on effects that were supported in Experiment 2, such as the tendency for outgroups to be assigned more compensation than ingroups only in the responsibility/third-party source condition, and the tendency to show more shame than guilt primarily in the responsibility condition.

The issue in this new experiment was the history of US military interference in the democratic politics of the DR. In the responsibility condition, participants read about the history of US military intervention to overthrow elected leaders in the DR on two occasions in the 20th century. In the control condition, they also read about a threat to democracy in the DR, but this was presented as a description of internal corruption in DR politics, with no mention of US intervention. The reparation option in the funding task was described as contributions from either the United States or the UN to help the DR stabilize its democratic institutions, with the alternatives being funding of US interests in the Caribbean (equivalent to the money going to Poland in Experiments 1–2) and funding democratic institutions in other parts of the world (equivalent to the money going to international institutions in Experiments 1–2). Design and analyses of the dependent variables were similar to Experiment 2, except that we pre-registered a more inclusive approach to the emotion measures in which the *shame*, *guilt*, and near-synonym items from the single-item emotion questionnaire were included together with the more group-based emotion items as indices of shame and guilt.

Method

Participants and Procedure

Participants were recruited via the Amazon Mechanical Turk crowdsourcing platform, with further selectivity via CloudResearch filtering of invalid IP addresses. The design and analysis plans for the study were pre-registered at <https://aspredicted.org/blind.php?x=vb22wc>. The sample was limited to adult US citizens with a past study approval rate of 95% or more from 50 or more studies. The pre-registered analysis for 90% power to detect partial eta squared of .03 (smallest effect of interest found in Experiment 2) in the study's design yielded an *a priori* sample size of 344. The design, as in Experiment 2, was 2 (Responsibility: Ingroup Responsibility vs. Control, between-subjects) \times 2 (Compensation Source: United States vs. UNs, between-subjects) \times compensation recipient (DR, United States, or international aid fund; within-subjects).

The total sample of participants who filled out at least one question was 408. Having excluded people, as pre-registered, who failed either of the attention checks (e.g., "I feel that this is a question to check attention and I should answer agree.") or completed the study under 90 seconds, the final sample consisted of 348 participants (48.3% female, two persons indicated nonbinary identities and two preferred not to say; $M_{\text{age}} = 41.89.8$, $SD = 11.9$). All participants signed a written informed consent before starting the experiment.

Manipulation

In the *ingroup responsibility* experimental condition, the participants read a short, four-paragraph text telling of historical relations between the DR and the United States,

Table 6. Experiment 2, control condition: correlations among shame, guilt, and percent of money given to the victims, Poland, and global institutions, by the source of money: EU (above diagonal) and Poland (below)

	1	2	3	4
1. Shame and guilt		.03	-.40*	.40*
2. Money assigned to victims	-.08		-.55**	.30
3. Money assigned to Poland	-.04	.72**		-.63**
4. Money assigned to global institutions	.14	-.04	-.66**	

Note. * $p < .05$ ** $p < .01$.

which described the responsibility of the United States for anti-democratic military and political interventions in the DR. The participants were instructed that, having read the text, they would be asked a number of questions about their reaction to the situation and to hypothetical actions following it. In the control condition, the participants read a text of about the same length that dealt with internal corruption in the DR without implicating the United States (see ESM 1). Both texts were followed by the measure of emotions toward the situation in the DR and a short questionnaire on guilt, moral shame, and image shame. Finally, as in Experiment 1 and Experiment 2, the participants were asked to distribute money coming from either the United States or the UNs among the DR, the United States, and a general international aid fund.

Materials

Following the manipulation of ingroup responsibility, we measured, on 5-point scales, 10 emotions toward the situation in the DR – six negative: regret, guilt, disgust, shame, anger, and embarrassment; and four positive: joy, happiness, pride, and enthusiasm. Next, as in Experiment 1 and Experiment 2, participants filled in a short questionnaire including nine items on image shame, moral shame, and guilt, where two attention check questions were embedded. The *distribution task* was also the same as in Experiment 1 and Experiment 2, and came last. As pre-registered, *shame index* was calculated from the means of *shame* and *embarrassment* items and all six group-based shame items ($M = 3.04$, $SD = 1.32$, $\alpha = .96$), and guilt index from the means of *guilt*, *regret*, and all three group-based guilt items ($M = 2.72$, $SD = 1.26$, $\alpha = .92$). This was a slightly different plan from the method followed in Experiment 2, in which only the group-based shame and guilt items were analyzed.

Results and Discussion

To test the main hypothesis about source and responsibility, we conducted again a mixed model $2 \times 2 \times 3$ ANOVA (Responsibility, between: Ingroup Responsibility vs. Control; \times Compensation Source, between: UN vs. the United States; \times Recipient, within: DR vs. the United States vs. international aid fund). Similar to Experiment 2, the main effects of responsibility and source and the interaction between them were nonsignificant and close to zero. Moreover, exactly as in Experiment 1 and Experiment 2, we found a significant main effect of recipient, $F(2, 344) = 20.86$, $p < .001$, $\eta_p^2 = .06$, and a significant interaction between recipient and compensation source, $F(2, 344) = 12.05$, $p < .001$, $\eta_p^2 = .03$, and this time also a significant interaction between recipient and responsibility, $F(2, 344) = 3.94$, $p = .02$, $\eta_p^2 = .01$. The three-way

interaction between ingroup responsibility, compensation source, and recipient again turned out to be nonsignificant, $F(2, 344) = 1.42$, $p = .23$, $\eta_p^2 = .004$.

As expected, pairwise comparisons between source conditions for each of the three funding recipients showed that when the money came from the UN versus the United States (across responsibility conditions), participants gave significantly more to the DR ($M = 31.08$, $SE = 1.84$ vs. $M = 25.80$, $SE = 1.80$, $p = .041$, Cohen's $d = 0.22$), which confirmed the findings from Experiment 1 and Experiment 2. Similarly, when the money came from the United States versus UN, participants gave more to the United States ($M = 35.96$, $SE = 1.94$ vs. $M = 22.11$, $SE = 1.99$, $p < .001$, Cohen's $d = 0.53$). The amount of money given to the third-party international aid fund this time was larger when the money came from the UN versus the United States ($M = 46.81$, $SE = 2.26$ vs. $M = 38.24$, $SE = 2.21$, $p = .007$, Cohen's $d = 0.29$). Overall, in the UN condition, the DR received significantly larger amount of money than the United States ($p = .004$) received. In turn, when the money came from the United States, the participants kept significantly more for the United States than they gave to DR ($p = .001$); the difference between the amount given to the United States and the international aid fund was not significant ($p < .54$).

The next point of interest was whether these results were especially pronounced in the ingroup responsibility condition, reflecting the stacking effects of the two-way interactions and main effects. Specifically, pairwise comparisons between the money assigned to DR and the United States in all four conditions of the 2×2 design (Responsibility, Ingroup Responsibility vs. Control \times Compensation Source, the United States vs. UN) revealed that the only condition where DR received more money than the United States was the combination where Americans were perpetrators, and the money came from the UN (Table 7). The preference for DR in this condition was this time highly significant, $p < .001$, while in all other conditions the trend was for the United States to be assigned more money than DR. Put another way, when the United States was presented as responsible for the DR's problems and had to pay reparations itself, the proportion of money assigned to the DR ($M = 26.72\%$) was quite similar to the amount of money assigned when it was not presented as responsible for the DR's problems ($M = 26.41\%$). Only when the UN was paying, and the United States was presented as responsible, the amount of money assigned to the DR was appreciably more generous ($M = 34.22\%$).

We conducted a maximum likelihood factor analysis as part of the pre-registered plan to see whether the different shame items separated out from each other and from guilt. The analysis entered the single-word items shame, guilt, embarrassment, and regret, and all nine of

the detailed shame/guilt items. One factor emerged as clearly dominant with no others having eigenvalue greater than 1, and this had eigenvalue = 9.27, accounting for 71.3% of variance and including all items (loadings > .50). Looked at another way, the pre-registered shame index (mean of *shame* and *embarrassed* items and all six detailed shame items) correlated with the pre-registered guilt index (mean of *guilt*, *regret*, and all three detailed guilt items) at $r = .87$. As in Experiment 2 where similar conditions were obtained, and again following the pre-registered plan, we involved only the average of the shame and guilt indices ($M = 2.88$, $SD = 1.25$, $\alpha = .97$) in all correlations (Table 8 and Table 9). However, we were able to use the two separate index scores in analyses comparing their means, which would not be affected by the high correlation.

To test the comparative strength of shame and guilt means, we conducted a mixed model ANOVA, 2 (Responsibility, between: Ingroup Responsibility vs. Control) \times 2 (Compensation Source, between: the United States vs. UN), \times 2 (Emotion, within: Guilt vs. Shame index). As in Experiment 2, the mean levels of shame were higher than guilt overall, $F(1, 344) = 81.85$, $p < .001$, $\eta_p^2 = .19$. Also, the main effect of responsibility was significant, $F(1, 344) = 66.95$, $p < .001$, $\eta_p^2 = .16$, indicating more guilt and shame when the ingroup was responsible. The Emotion \times Responsibility interaction, $F(1, 344) = 13.50$, $p < .001$, $\eta_p^2 = .04$, showed a greater advantage for shame over guilt in the responsibility condition ($M = 3.59$ vs. $M = 3.15$, mean difference 0.44) than in the control ($M = 2.45$ vs. $M = 2.27$, mean difference 0.18). Source did not show any significant main effects or interactions, all $p > .41$.

Table 8 and Table 9 show correlations between shame/guilt and compensation amounts to each recipient, separately for the four conditions. As in Experiment 2, in the responsibility condition, shame/guilt was correlated with both the money given to DR (positively) and the money given to the United States (negatively), which confirms the overall important role of moral emotions in these reparative actions. Similarly, the negative correlations among distribution scores generally reflect the zero-sum nature of the task. Unexpectedly, there were also significant correlations

between shame/guilt and DR money assignment in the control conditions, possibly due to general awareness of unequal relations between the United States and DR even when no specific US misdeed had been singled out.

General Discussion

The present research investigated whether fellow group members of perpetrators are more strongly motivated to engage in reparative actions if they have an opportunity to source the money for this purpose from a third party, as recently shown by de Hooge et al. in an individual context (de Hooge et al., 2011). Moreover, we predicted that in the intergroup context, *prosocial* behavior, even at the expense of a third party, would be characterized by shame more so than guilt.

The findings confirmed our hypotheses. All three experiments demonstrated that when the negative behavior of ingroup members toward a victimized group was made salient, and participants as ingroup members envisioned the opportunity to compensate the victims at the expense of a third party, they allocated substantially more to the victims than when the money was assigned from their own group's resources. It is worth noticing that they did this although they were also allowed to allocate externally sourced money to the ingroup. Thus, it is not a mere case of maximizing outcomes for the ingroup. Besides, although there was a negative dependency among the targets, the presence of a third party (global institutions or an international aid fund) meant that giving to the outgroup was not exactly the same as holding back money from the ingroup. There appears to have been a guilt-and-shame-related motivation in both conditions to compensate victims, but participants found it easier to do when the money came from another entity than their own country.

Furthermore, we demonstrated that in this intergroup context, shame had overall a better case than guilt to be the emotion that motivated perpetrators to compensate the victims at the expense of the third party. First, shame in all experiments was felt more strongly than guilt. Second, in

Table 7. Pairwise comparisons between means of the money distribution (in percentage of \$100 million) between Dominican Republic (DR) and the United States in the respective conditions in Experiment 3

Condition	Recipient				
	DR		The United States		<i>p</i>
	<i>M</i>	<i>SE</i>	<i>M</i>	<i>SE</i>	
The United States responsible/money from UN	34.22	2.51	15.71	2.71	<.001
The United States responsible/money from the United States	26.72	2.54	34.36	2.74	.074
Control condition/money from UN	27.94	2.70	28.51	2.91	.900
Control condition/money from the United States	24.89	2.54	37.55	3.12	.003

Table 8. Experiment 3, ingroup responsibility condition: correlations among shame/guilt, and money given to the DR, the United States, and international aid fund (IAF), by the source of money: UN (above diagonal) and the United States (below)

	1	2	3	4
1. Shame and guilt		.27**	-.28*	-.06
2. Money assigned to DR	.39**		-.19	-.77**
3. Money assigned to the United States	-.31**	-.41**		-.47**
4. Money assigned to IAF	-.02	-.43**	-.65**	

Note. ** $p < .01$.

Experiment 1 where shame and guilt were statistically distinct from each other (unlike Experiment 2 and Experiment 3), shame showed stronger correlations with patterns of outgroup-supporting allocation regardless of where funds came from. These findings support the assumptions in Nelissen et al. (2013) that guilt is more likely to motivate prosocial behavior in direct reciprocity whereas shame is more likely to motivate prosocial behavior in indirect reciprocity. Given that the participants did not know the victims or third parties and had little likelihood of meeting them face-to-face, they were more plausibly motivated by indirect reciprocity. In contrast to previous findings suggesting that shame had much to do with reputation management and self-pitying, which mediated reparation attitudes separately from empathy for the outgroup (Brown & Čehajić, 2008), we can surmise from these studies that shame might motivate repairing damage to victims regardless of whether the group itself is seen to take responsibility by paying from its own resources.

Both experiments that manipulated responsibility found that victims were given substantially higher amounts of compensation, compared to the ingroup, only if the ingroup was presented as responsible and a third party – not the ingroup itself – was paying. In Experiment 2, however, when the ingroup was not presented as responsible, people were also significantly more likely to use the third party's money to pay the victims than to use ingroup money; in Experiment 3, the difference was nonsignificant but in the same direction (27.94% when the UN paid and 24.89% when the United States paid). This raises the possibility that in Experiment 2, people could think that EU was in fact responsible for the refugee crisis, more so than the UN bearing responsibility for the democracy crisis in the DR. Thus, shifting responsibility completely away from Poland

and onto the EU could be taken as a possible explanation for why people in Experiment 2 would want to make the EU pay even without Polish responsibility.

Another way to see the outcomes is to note that greater giving to outgroup victims relative to the ingroup or a third party, as an effect of the responsibility manipulation, emerged only when the EU or UN supplied money for these reparations, and not when Poland or the United States themselves had to. Although this outcome was not supported by significant three-way interactions, there is some support in the more detailed analyses: The only combination of factors that led to a significant simple giving pattern such that more was given to the victims than to the ingroup was when Poland and the United States had responsibility, but the EU or the UN gave money, respectively. These findings, then, support the view that to increase help to victim groups, mere responsibility was not enough; there had to also be a lack of negative consequences, including no resource loss for the ingroup.

In any case, the findings suggest that shame specifically motivates ingroup members to support reparative actions (vs. withdrawal) when they are provided with a *good opportunity*, that is, when the actions are not too risky (de Hooij et al., 2010), or can be done in a cost-efficient manner (Nelissen et al., 2013). Moreover, they shed new light on the literature suggesting that in the intergroup context shame leads the members of a perpetrator group to take actions aimed at improving the ingroup's reputation at the lowest cost possible or to withdraw from the embarrassing situation (Brown et al., 2008). Indeed, when the money for victim compensation came from the ingroup's resources, the participants decided to keep the lion's share of it for the group and pay substantially less to the victims, evidently choosing withdrawal rather than improving the

Table 9. Experiment 2, control condition: correlations among shame/guilt, and money given to the DR, the United States, and IAF, by the source of money: UN (above diagonal) and the United States (below)

	1	2	3	4
1. Shame and guilt		.23*	-.20	.01
2. Money assigned to DR	.22*		-.32**	-.47**
3. Money assigned to the United States	-.30*	-.25*		-.69**
4. Money assigned to IAF	.10	-.52**	-.70**	

Note. * $p < .05$, ** $p < .01$.

group image through costly apology. However, we should also note that shame correlated with victim allocations about equally in conditions of responsibility, regardless of where the money came from. Shame in our studies thus looked more like a general motivator for reparation at a collective level when the ingroup is responsible, rather than an emotion that has a particular influence in situations where a collectively painless solution is available. Nevertheless, it is still an open question – calling for a repeated-measures study design – whether outsourcing reparation is just a cheaper solution or actually has a different impact on alleviating shame or guilt.

While these findings refer to limited contexts of reparations for bad treatment of ingroup's victims, we would expect them to generalize to similar group situations in which people have a similar understanding of shame and guilt. We should also note that minor context differences appear to have played a part in some differences between findings. For example, although presenting the United States as explicitly responsible increased reparation to victims and feelings of shame and guilt, we found correlations between those emotions and reparation tendencies to be significant even in the control, no responsibility condition. As mentioned, this may have been due to some participants being generally aware of the history of US intervention in Latin America and thus feeling some responsibility even for problems presented as internal to the DR. In the Polish manipulation, a parallel awareness of refugee issues might not have been present, especially given the less parallel construction of those manipulations.

Our main caveat for any future research would be that the right topic is necessary to find these effects. That is, one would have to start with an issue that at least some participants would see as an example of ingroup wrongdoing, rather than an issue that arouses unanimous denial or defensiveness. By the same token, it would be worth measuring multidimensional group identification as a possible moderator of admittance or denial of the group wrongdoing. All the same, we think that once the bridge is crossed and ingroup members accept collective wrongdoing, they would find it a very easy detour to avoid the self-punishing aspect of reparations, and more gladly support paying with someone else's money, than with their own.

References

- Allpress, J. A., Barlow, F. K., Brown, R., & Louis, W. R. (2010). Atoning for colonial injustices: Group-based shame and guilt motivate support for reparation. *International Journal of Conflict and Violence*, 4(1), 75–88.
- Allpress, J. A., Brown, R., Giner-Sorolla, R., Deonna, J. A., & Teroni, F. (2014). Two faces of group-based shame: Moral shame and image shame differentially predict positive and negative orientations to ingroup wrongdoing. *Personality and Social Psychology Bulletin*, 40(10), 1270–1284. <https://doi.org/10.1177/0146167214540724>
- Bonnot, V., & Krauth-Gruber, S. (2018). Guilt norms regarding historical violence and implications for intergroup relations in France. *International Journal of Intercultural Relations*, 62, 43–54. <https://doi.org/10.1016/j.jintrel.2017.05.003>
- Brown, R., & Čehajić, S. (2008). Dealing with the past and facing the future: Mediators of the effects of collective guilt and shame in Bosnia and Herzegovina. *European Journal of Social Psychology*, 38(4), 669–684. <https://doi.org/10.1002/ejsp.466>
- Brown, R., González, R., Zagefka, H., Manzi, J., & Čehajić, S. (2008). Nuestra culpa: Collective guilt and shame as predictors of reparation for historical wrongdoing. *Journal of Personality and Social Psychology*, 94(1), 75–90. <https://doi.org/10.1037/0022-3514.94.1.75>
- Cohen, T. R., Panter, A. T., & Turan, N. (2012). Guilt proneness and moral character. *Current Directions in Psychological Science*, 21(5), 355–359. <https://doi.org/10.1177/0963721412454874>
- de Hooge, I. E., Breugelmans, S. M., & Zeelenberg, M. (2008). Not so ugly after all: When shame acts as a commitment device. *Journal of Personality and Social Psychology*, 95(4), 933–943. <https://doi.org/10.1037/a0011991>
- de Hooge, I. E., Nelissen, R. M. A., Breugelmans, S. M., & Zeelenberg, M. (2011). What is moral about guilt? Acting “prosocially” at the disadvantage of others. *Journal of Personality and Social Psychology*, 100(3), 462–473. <https://doi.org/10.1037/a0021459>
- de Hooge, I. E., Zeelenberg, M., & Breugelmans, S. M. (2007). Moral sentiments and cooperation: Differential influences of shame and guilt. *Cognition and Emotion*, 21(5), 1025–1042. <https://doi.org/10.1080/02699930600980874>
- de Hooge, I. E., Zeelenberg, M., & Breugelmans, S. M. (2010). Restore and protect motivations following shame. *Cognition & Emotion*, 24(1), 111–127. <https://doi.org/10.1080/02699930802584466>
- de Rivera, J., Kurrien, R., & Olsen, N. (2007). The emotional climate of nations and their culture of peace. *Journal of Social Issues*, 63(2), 255–271. <https://doi.org/10.1111/j.1540-4560.2007.00507.x>
- Doosje, B., Branscombe, N. R., Spears, R., & Manstead, A. S. R. (1998). Guilty by association: When one's group has a negative history. *Journal of Personality and Social Psychology*, 75(4), 872–886. <https://doi.org/10.1037/0022-3514.75.4.872>
- Eurostat. (July 2, 2018). *Statystyka dotycząca migracji i populacji migrantów*. http://ec.europa.eu/eurostat/statistics-explained/index.php/Migration_and_migrant_population_statistics/pl
- Gausel, N., & Brown, R. (2012). Shame and guilt-do they really differ in their focus of evaluation? Wanting to change the self and behavior in response to ingroup immorality. *The Journal of Social Psychology*, 152(5), 547–567. <https://doi.org/10.1080/00224545.2012.657265>
- Gausel, N., Leach, C. W., Vignoles, V. L., & Brown, R. (2012). Defend or repair? Explaining responses to in-group moral failure by disentangling feelings of shame, rejection, and inferiority. *Journal of Personality and Social Psychology*, 102(5), 941–960. <https://doi.org/10.1037/a0027233>
- Gausel, N., Leach, C. W., Mazzitotta, A., & Feuchte, F. (2017). Seeking revenge or seeking reconciliation? How concern for social-image and felt shame helps explain responses in reciprocal intergroup conflict. *European Journal of Social Psychology*, 48(1), 062–072. <https://doi.org/10.1002/ejsp.2295>
- Giner-Sorolla, R., Castano, E., Espinosa, P., & Brown, R. (2008). Shame expressions reduce the recipient's insult from outgroup reparations. *Journal of Experimental Social Psychology*, 44(3), 519–526. <https://doi.org/10.1016/j.jesp.2007.08.003>

- Goldberg, C. (1991). *Understanding shame*. Jason Aronson.
- Griffin, B. J., Moloney, J. M., Green, J. D., Worthington, Jr., E. L., Cork, B., Tangney, J. P., Van Tongeren, D. R., Davis, D. E., & Hook, J. N. (2016). Perpetrators' reactions to perceived interpersonal wrongdoing: The associations of guilt and shame with forgiving, punishing, and excusing oneself. *Self and Identity*, 15(6), 650–661. <https://doi.org/10.1080/15298868.2016.1187669>
- Iyer, A., Schmader, T., & Lickel, B. (2007). Why individuals protest the perceived transgressions of their country: The role of anger, shame, and guilt. *Personality and Social Psychology Bulletin*, 33(4), 572–587. <https://doi.org/10.1177/0146167206297402>
- Johns, M., Schmader, T., & Lickel, B. (2005). Ashamed to be an American? The role of identification in predicting vicarious shame for anti-Arab prejudice after 9–11. *Self and Identity*, 4(4), 331–348. <https://doi.org/10.1080/15298860500145822>
- Kamau, C., Giner-Sorolla, R., & Zebel, S. (2013). Reconciliation responses, blame, and expressions of guilt or shame. *Journal of Applied Social Psychology*, 43(suppl.2), E287–E292. <https://doi.org/10.1111/jasp.12048>
- Kardos, P., Leidner, B., Castano, E., & Lickel, B. (2019). The benefits of collective responsibility: How ingroup reputation concern motivates prosociality in intergroup contexts. *European Journal of Social Psychology*, 49(1), 93–109. <https://doi.org/10.1002/ejsp.2506>
- Leach, C. W., & Cidam, A. (2015). When is shame linked to constructive approach orientation? A meta-analysis. *Journal of personality and social psychology*, 109(6), 983–1002. <https://doi.org/10.1037/pspa0000037>
- Lickel, B., Schmader, T., Curtis, M., Scarnier, M., & Ames, D. R. (2005). Vicarious shame and guilt. *Group Processes & Intergroup Relations*, 8(2), 145–157. <https://doi.org/10.1177/1368430205051064>
- Nelissen, R. M. A., Breugelmans, S. M., & Zeelenberg, M. (2013). Reappraising the moral nature of emotions in decision making: The case of shame and guilt. *Social and Personality Psychology Compass*, 7(6), 355–365. <https://doi.org/10.1111/spc3.12030>
- Nelissen, R. M. A., Dijk, A. J. M., & de Vries, N. K. (2007). How to turn a hawk into a dove and vice versa: Interactions between emotions and goals in a give-some dilemma game. *Journal of Experimental Social Psychology*, 43(2), 280–286. <https://doi.org/10.1016/j.jesp.2006.01.009>
- Rees, J. H., Allpress, J. A., & Brown, R. (2013). Niewieder: Group-based emotions for in-group wrongdoing affect attitudes toward unrelated minorities. *Political Psychology*, 34(3), 387–407. <https://doi.org/10.1111/pops.12003>
- Regan, D. T., Williams, M., & Sparling, S. (1972). Voluntary expiation of guilt: A field experiment. *Journal of Personality and Social Psychology*, 24(1), 42–45. <https://doi.org/10.1037/h0033553>
- Reysen, S., Puryear, C., Katzarska-Miller, I., Kamble, S. V., & Vithoji, N. (2014). Socio-structural intergroup characteristics and group-based emotions in three countries. *International Journal of Intercultural Relations*, 43(part B), 239–252. <https://doi.org/10.1016/j.ijintrel.2014.09.001>
- Sekerdej, M., & Giner-Sorolla, R. (2021). Data and supplementary material for “Members of transgressor groups prefer reparations to come from third parties: An extension of interpersonal findings on moral emotions.” https://osf.io/ag2dv/?view_only=d65f6434d510443ba2990c
- Sheikh, S. (2014). Cultural variations in Shame's responses. *Personality and Social Psychology Review*, 18(4), 387–403. <https://doi.org/10.1177/1088868314540810>
- Shepherd, L., Spears, R., & Manstead, A. S. (2013a). The self-regulatory role of anticipated group-based shame and guilt in inhibiting in-group favoritism. *European Journal of Social Psychology*, 43(6), 493–504. <https://doi.org/10.1002/ejsp.1971>
- Shepherd, L., Spears, R., & Manstead, A. S. R. (2013b). 'This will bring shame on our nation': The role of anticipated group-based emotions on collective action. *Journal of Experimental Social Psychology*, 49(1), 42–57. <https://doi.org/10.1016/j.jesp.2012.07.011>
- Tangney, J. P. (1995). Shame and guilt in interpersonal relationships. In J. Tangney & K. W. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride* (pp. 114–139). Guilford Press.
- Tangney, J. P., & Fischer, K. W. (1995). In W. J. P. Tangney & K. W. Fischer (Eds.), *Self-conscious emotions: The psychology of shame, guilt, embarrassment, and pride*. Guilford Press.
- Vel-Palumbo, M., Woodyatt, L., & Wenzel, M. (2018). Why do we self-punish? Perceptions of the motives and impact of self-punishment outside the laboratory. *European Journal of Social Psychology*, 48(6), 756–768. <https://doi.org/10.1002/ejsp.2368>
- Wicker, F. W., Payne, G. C., & Morgan, R. D. (1983). Participant descriptions of guilt and shame. *Motivation and Emotion*, 7(1), 25–39. <https://doi.org/10.1007/bf00992963>
- Yu, H., Hu, J., Hu, L., & Zhou, X. (2014). The voice of conscience: Neural bases of interpersonal guilt and compensation. *Social Cognitive and Affective Neuroscience*, 9(8), 1150–1158. <https://doi.org/10.1093/scan/nst090>
- Zeelenberg, M., & Breugelmans, S. M. (2008). The role of interpersonal harm in distinguishing regret from guilt. *Emotion*, 8(5), 589–596. <https://doi.org/10.1037/a0012894>

History

Received May 15, 2020

Revision received January 31, 2022

Accepted February 1, 2022

Published online March 18, 2022

Conflict of Interest

The authors do not have any conflicts of interest to declare.

Publication Ethics

The authors confirm that this article adheres to ethical guidelines specified in the APA Code of Conduct as the authors' national ethics guidelines.

Authorship

Maciej Sekerdej and Roger Giner-Sorolla designed the research. M. Sekerdej performed the research, analyzed the data, and wrote the manuscript. M. Sekerdej and R. Giner-Sorolla edited the final version of the manuscript. Both authors approved the final manuscript for submission.

Open Data

All data reported in the paper along with Supplemental Material are made publicly available at the Open Science Framework, https://osf.io/ag2dv/?view_only=d65f6434d510443ba2990c. The design and analysis plans for study 3 were pre-registered at <https://aspredicted.org/blind.php?x=vb22wc>.

Funding

This work was supported by the National Science Centre (Grant No. 2014/14/M/HS6/00917) awarded to the first author.

ORCID

Maciej Sekerdej

 <https://orcid.org/0000-0001-7031-1523>

Maciej Sekerdej

Institute of Psychology
Jagiellonian University
ul. Ingardena 6
30-060 Kraków
Poland
maciek@apple.phils.uj.edu.pl



When You Are Wrong on Facebook, Just Admit It

Wrongness Admission Leads to Better Interpersonal Impressions on Social Media

Adam K. Fetterman^{1,2} , Nicole L. Muscanell^{2,3}, Dongjie Wu¹, and Kai Sassenberg^{2,4}

¹Department of Psychology, University of Houston, Houston, TX, USA

²Social Processes Lab, Leibniz-Institut für Wissensmedien, Tübingen, Germany

³Defense Equal Opportunity Management Institute, JHT, Incorporated, USA

⁴Department of Psychology, University of Tübingen, Germany

Abstract: Intellectually humble behavior, like admitting when you are wrong, leads to better impression formation. However, online social networks (OSNs) have changed the impression formation process. We investigated the impact of wrongness admission on impression formation during an OSN argument. In four experiments ($N = 679$), participants witnessed a user engage in wrongness admission, refuse to admit, or not respond, in an argument on a Facebook wall. Participants reported their impressions of whether they would be willing to interact with the (non) admitting user. User reputation ratings and interaction intentions were higher in the admission (vs. nonadmission) condition. The latter effect was mediated by user reputation ratings. Wrongness admission appears to have a positive impact on impression formation on OSNs.

Keywords: intellectual humility, impression formation, social media, reputation, wrongness admission

Arguments on online social networks (OSNs), such as Facebook and Twitter, have become a part of everyday life (Anderson et al., 2018; Neubaum et al., 2021). This is not surprising because OSNs create a space for people to easily express their opinions, which can encourage open dialog. However, OSNs have also changed the impression formation process (Walther & Whitty, 2021). Instead of forming impressions based on face-to-face interactions, OSNs allow people to look at people's online behavior as cues to their personalities (Utz, 2010). Furthermore, people using OSNs often form impressions through passive observation (Qin et al., 2021). Therefore, an OSN user's behavior during an online argument might have an impact on the impressions passive viewers form of the user (Orben & Dunbar, 2017). One such behavior that can occur during an online argument is a user's response when they discover that their viewpoint is factually wrong. With potentially hundreds (or more, depending on their privacy settings) of passive witnesses, the user can admit that they are wrong or avoid doing so. Based on the literature of impression formation on OSNs and intellectual humility, the OSN user's best course of action, here, is to publicly admit that they are wrong. The current set of experiments investigated whether this is indeed the case by testing

whether such *wrongness admission* during a Facebook argument impacts passive viewers' impression formation.

Intellectual Humility and Wrongness Admission

Intellectual humility is defined by Leary et al. (2017) as "recognizing that a particular personal belief may be fallible, accompanied by an attentiveness to limitations in the evidentiary basis of that belief and to one's own limitations in obtaining and evaluating relevant information" (p. 793). Those who score high on broad measures of intellectual humility, based on this definition, are more agreeable and open-minded (Leary et al., 2017). More recently, Porter et al. (2021) refined intellectual humility by developing a classification framework introducing different forms of intellectual humility. This classification is based on whether the intellectual humility is focused on the self or others – that is, about one's own or another person's beliefs – and whether it is internal or expressed – that is, about self-reflective awareness and cognitions compared to behavioral manifestations. In relation to the self/expressed quadrant of their framework, these observable

behaviors can be public or private. It is within public component of the self/expressed quadrant of intellectual humility that we locate a concept we call “wrongness admission.”

Wrongness admission is a form of public attitude change that involves expressing that one has changed a previously held attitude or belief (Fetterman et al., 2019). In this sense, wrongness admission is about recognizing and publicly acknowledging that a belief is *inaccurate*. Wrongness admission is not about being morally wrong or about behavioral errors. Willingness to engage in wrongness admission, as we define here, is positively correlated with agreeableness, openness to experience, honesty/humility, and emotional intelligence (Fetterman et al., 2019). Furthermore, those who score high in intellectual humility are more likely to engage in wrongness admission during an argument (Rodríguez et al., 2019). Wrongness admission, then, is an observable behavior that cues not only intellectual humility but also positive interpersonal traits (e.g., trustworthiness and friendliness).

If intellectual humility and wrongness admission, more specifically, are cues to positive interpersonal traits, then they should be reflected in observer reports of intellectually humble people. Indeed, Meagher et al. (2021) found that those who scored higher on intellectual humility, in general, tended to be rated by interaction partners as more agreeable and open-minded, less arrogant, and dominant, and that interaction partners felt more satisfaction when conversing with intellectually humble others (Meagher et al., 2021). Based on this work, intellectual humility appears to have a positive impact on impression formation, notably due to its cue of positive interpersonal traits.

There has been relatively little work looking more specifically at the impact of wrongness admission on impression formation. There are two notable exceptions. First, a study by Fetterman and Sassenberg (2015) had published scientists read one of two scenarios. In one scenario, after learning about a convincing failed replication of one of their research findings, a hypothetical scientist admits that they were wrong about the finding. In the other condition, the scientist refuses to admit that they were wrong about the finding. The results indicated that participants in the admission condition rated the target scientist as more competent and trustworthy than participants in the refuse condition. This was the first study to suggest that wrongness admission has a positive impact on impression formation.

In the second notable exception, John et al. (2019) investigated wrongness admission in business settings. They had participants give an entrepreneurial pitch, during which a panel contradicted some of the participants' statements. The results showed that those who backed down (i.e., wrongness admitters) were more likely to be

advanced to a final competition round and were rated as more competent by observers. Furthermore, these competence ratings impacted, in a positive way, whether the observers wanted to hire the admitter or invest in their ideas. Wrongness admission, again, appears to have a positive impact on impression formation.

Intellectual humility and wrongness admission appear to cue positive interpersonal traits for observers forming impressions. Yet, OSNs are unique environments in which people manage and form impressions (Walther & Whitty, 2021). Passive observers on OSNs form spontaneous impressions, without much context, of strangers as they scroll through their posts (Levordashka & Utz, 2017). Therefore, it is important to understand how wrongness admission on OSNs impacts impression formation.

Impression Formation on OSNs

Humans have an inherent need to form and maintain relationships due to their incredible social nature (Baumeister & Leary, 1995). To ensure a mutually beneficial relationship, people rely on reputation to predict the likely behavior of potential interaction partners, even on OSNs (Tennie et al., 2010). To get a sense of someone's reputation, people make inferences about that person's personality (i.e., impression formation). People form impressions of others along two fundamental dimensions: communion and agency (Abele et al., 2016; Abele & Wojciszke, 2014; Asch, 1946; Cuddy et al., 2008; Fiske et al., 2007). The agency dimension includes traits related to achievement and competence, while the communion dimension includes social variables, such as trust and friendliness. People often form impressions about these traits quickly, sometimes within seconds, based on non-verbal cues (Ambady & Rosenthal, 1992), which are remarkably accurate (Funder, 2012). However, given the rise of OSNs, impression formation has changed and to some extent moved online.

Since OSNs have become an everyday feature of most people's social lives (Seidman, 2013), much research has focused on impression formation and management on OSNs (Bacev-Giles & Haji, 2017). OSNs, however, have changed impression formation. For example, much OSN-based impression formation is passive in that the people are not interacting with those for whom they are forming impressions (Orben & Dunbar, 2017). Due to this lack of interaction, according to the hyperpersonal model (Walther & Whitty, 2021), people often form impressions based on textual and pictorial information posted by users. Of course, this material can be highly curated by the users (Krämer & Winter, 2008). Therefore, according to the warranting principle (Walther et al., 2009), viewers prefer

information that is less subject to user manipulation. Indeed, prior work shows that viewers form fewer positive impressions from user-generated information, likely due to the perceived untrustworthiness of this information (Utz, 2010).

Prior work investigating impression formation on OSNs has revealed specific traits and online behaviors that lead to more positive impressions, or liking. Much of this work has shown that communal behavior tends to lead to the most positive impressions. For example, Buffardi and Campbell (2008) found that self-promotion led to narcissistic impressions, which was related to lower ratings of communion. Furthermore, Stopfer et al. (2014) found that communal and open-minded people were more liked on OSNs. More recently, Qin et al. (2021) found that positive self-disclosure on OSNs led to higher ratings of trustworthiness and likeability. As noted, viewers might not form as many positive impressions from user-generated information (Utz, 2010). However, the negative impact of user-generated information can be overridden if the user engages in self-deprecation, a cue of humility (Austin et al., 2021; Bareket-Bojmel et al., 2016).

Overall, it seems that impression formation on OSNs is common and spontaneous (Levordashka & Utz, 2017). Although impression formation is different on OSNs than offline situations (Walther & Whitty, 2021), these impressions tend to be accurate (Stopfer et al., 2013). Furthermore, people tend to form the most positive impressions for those on OSNs who display communal, open, and humble online behaviors. We noted above that intellectual humility and wrongness admission tend to cue such positive traits and lead to positive impressions offline. Intellectual humility and wrongness admission could lead to positive impression formation online as well.

Wrongness Admission on OSNs

Due to (a) the commonality in arguments on OSNs (Anderson et al., 2018; Neubaum et al., 2021), (b) people's tendency to form spontaneous impressions of strangers while passively scrolling through their posts (Levordashka & Utz, 2017), and (c) the impact of passive judgments in impression formation in general (Quadflieg & Penton-Voak, 2017) and on OSNs (Waggoner et al., 2009), it seems important to test how wrongness admission during OSN arguments impacts the impression formation process. In fact, Utz (2010), drawing from the Brunswik lens model (Brunswik, 1956; Gigerenzer & Kurz, 2001), suggests that the impressions formed from cues on OSNs create a lens through which viewers make inferences about individuals. Intellectually humble behaviors, such as

wrongness admission, appear to cue positive interpersonal traits, which could shape a viewer's lens of an online wrongness admitter. This leads to our first hypothesis.

Hypothesis 1 (H1): Scores on user reputation will be more positive for participants who witness a stranger in an OSN argument engage in wrongness admission than who witness a stranger refusing to admit.

Among the primary uses of OSNs is the forming of social connections (Alhabash & Ma, 2017). Impression formation serves the function of informing the viewer of whether they should engage in future interactions with the target (Gable & Reis, 2001). In the context of OSNs, people form impressions with less contextual information than in face-to-face interactions (Walther & Whitty, 2021). Since (a) people are primarily looking for social connections and thus forming impressions based on interpersonal traits on OSNs (Tennie et al., 2010), (b) friendships often form based on judgments of agreeableness (Harris & Vazire, 2016), and (c) wrongness admission appears related to such traits, wrongness admission in an OSN argument might be a cue to passive viewers that the user is someone to interact with in the future. In fact, Hagá and Olson (2017) found that people were more likely to want to interact with people who displayed outward intellectual humility, such as wrongness admission, due to their increased perceptions of niceness. This leads to our second and third hypotheses.

Hypothesis 2a (H2a): Participants who witness a stranger in an OSN argument will indicate that they are more willing to interact with the user in the future if that user engages in wrongness admission compared to a user who refuses to do so.

Hypothesis 2b (H2b): Participants who witness a stranger in an OSN argument will indicate that they are more willing to interact with the user in the future if that user engages in wrongness admission compared to a user who refuses to do so and this effect is mediated through increases in positive impression formation.

Current Experiments

Across four experiments, we tested our broad hypothesis that wrongness admission on OSNs would have a positive impact on impression formation. We focused all of our experiments on Facebook. According to statista.com, Facebook is still the most used OSN as of October 2021, with 2.9 billion active users worldwide. Even so, the studies

were run between 2014 and 2016, and during that time, Facebook use was ubiquitous (Duggan & Smith, 2016).

We designed our OSN materials in a similar manner to that of previous investigations investigating the impact of minimal information on impression formation on OSNs (e.g., Austin et al., 2021; Kaye et al., 2020) and based on the idea that viewers tend to discredit user-curated OSN profile information (Utz, 2010) and form strong impressions when passively observing interactions (Quadflieg & Penton-Voak, 2017). Therefore, we provided only the textual information and removed all identifying characteristics and profile information. For the impression formation outcome measures, we created measure of general reputation, which includes items that reflect communal and competence traits, similar to that of previous investigations of impression formation on OSNs (e.g., Austin et al., 2021; Bacev-Giles & Haji, 2017; Batenburg & Bartels, 2017). In all four experiments, participants read a staged argument on a Facebook wall – a section of Facebook that allows users to post information and engage in discussion – between two users. The argument was focused on a made-up food additive and reflected common themes of health food discourse on OSNs at the time of the study (Munro et al., 2015) to increase realism. The differences between the experiments are provided in Table 1.

Open Science Disclosures

The current experiments contain all but two studies that we have conducted on the topic of wrongness admission on Facebook. The two studies not included consisted of undergraduate thesis projects that do not meet the standards of publishable research. The data from Experiment 2B included questionnaires relevant, but not related to the

current hypotheses. This questionnaire data, but not the data we report here, were published in Fetterman et al. (2019).

We initially intended to measure general user reputation. However, as the project and our theorizing progressed, we thought that it might be more informative to split the general reputation items into separate communion and competence scores. When we did so, the two scores were highly positively correlated (all *rs* > .70). Furthermore, a parallel analysis on the data from all four experiments suggested that a single reputation factor was most appropriate. Therefore, we returned to our original general user reputation measure.

Beyond these disclosures, we report how we determined our sample sizes, all data exclusions, all manipulations, and all measures in the studies. We did not pre-register our hypotheses. Data, code, and full materials are available on the Open Science Framework (<https://osf.io/a4tpg/>).

Experiment 1A and Experiment 1B

Method

Participants

In Experiment 1A, we recruited 100 participants on Amazon Mechanical Turk based on the assumption that 50 per cell would be sufficient. We removed participants who did not accurately respond to an attention check or did not complete the experiment. Of the participants who completed the task, 88 (41 female; $M_{age} = 31.00$, $SD_{age} = 7.84$) answered the attention check correctly and 91% indicated that they use Facebook. For Experiment 1B, we performed an incorrect a priori power analysis, based on which we recruited 250 participants on Amazon Mechanical Turk. Of those participants, 238 participants (100 female) accurately responded to the attention check and 92% indicated that they use Facebook. A sensitivity analysis suggested that we had the power to detect a minimum effect size of $\eta^2_{1A} = .08$ and $\eta^2_{1B} = .04$ when $\alpha = .05$ (two-tailed) and power = .80.

Materials and Procedures

Participants were randomly assigned to an admission or a refuse (Experiment 1A and Experiment 1B) or a no-response (Experiment 1B only) condition. After providing informed consent, participants read instructions and said that we were interested in people’s responses to Facebook arguments, that we had access to a corpus of Facebook arguments, and that we would randomly select one of the arguments for them to read. In reality, all participants read the same argument. Only the final post from the target user differed and served as the

Table 1. Breakdown of the differences across experiments and final sample sizes

Experiment	Conditions present	Wording of condition	Final <i>N</i>
Experiment 1a	1. Admission	Wording 1	88
	2. No admission		
Experiment 1b	1. Admission	Wording 1	238
	2. No admission		
	3. No response		
Experiment 2a	1. Admission	Wording 2	171
	2. No admission		
	3. No response		
Experiment 2b	1. Admission	Wording 2	183
	2. No admission		
	3. No response		

Note. Wording of condition = differences in the wording of the final post of the target Facebook user based on feedback from participants.

manipulation. Participants also read that they were to pay close attention to Participant A, but that they should pay attention to all users. They could not advance until 2 minutes had past.

The Argument

Participants read a screenshot of a fabricated argument between two Facebook users. Two of us staged the argument on our own Facebook pages to increase realism. To further increase realism, we examined real arguments on OSNs and modeled the same argument tactics. For example, the *users* called-out biases, utilized and called-out logical fallacies, cited blogs as sources, and resorted to ad hominem attacks. However, we preplanned the argument

such that, while there was room for debate, Participant B clearly had the facts on their side and had the more compelling argument. Participant B also cited scientific papers, instead of blogs. We did this to avoid too many participants thinking that the admitter was inappropriately engaging in wrongness admission. We removed the identities of the actors and labeled them Participant A and Participant B. As Participant A was the focus of the experiment, we highlighted their posts with a red border. See Figure 1 for the full argument stimulus.

The final post by Participant A contained the manipulation. In one (admission) condition, Participant A ended the conversation by posting, "Thanks for the info and conversation. I was wrong and you were right." In the



Figure 1. Argument stimulus viewed by all participants in all experiments.

other (refuse) condition, Participant A ended the conversation by posting, “Thanks for the info and conversation. I guess we agree to disagree.” Therefore, the only thing that differed in the conditions was the second sentence in the final post. In Experiment 1B, participants in the no-response condition read, instead of a final post, “There were no further responses by Participant A.”

We conducted a pilot study ($N = 50$) to confirm that most people thought that Participant B was more correct. Participants read a neutral (i.e., no final post) version of the conversation and judged whether they thought Participant A or B was correct. Indeed, 70.59% of participants agreed that Participant B was correct. This was significantly different than 50%, $t(48) = 3.20$, $p = .002$, $d = 0.90$, indicating that perceptions of correctness were not random.

User Reputation Ratings

Participants, in both studies, rated their agreement (1 = *strongly disagree*; 5 = *strongly agree*) with 18 statements about Participant A with items related to communion (e.g., “Participant A is a nice person”) and to competence (e.g., “Participant A has little education”). We reverse scored the negatively worded items and averaged across all items to create a user reputation score (1A: $M = 3.06$, $SD = 0.77$, $\alpha = .94$; 1B: $M = 2.83$, $SD = 0.87$, $\alpha = .94$).

Interaction Intentions

To measure whether participants would be willing to interact, online or offline, with the target user in the future (i.e., testing *Hypotheses 2a* and *2b*), participants indicated their level of agreement (1 = *strongly disagree* to 5 = *strongly agree*) with four statements (e.g., “I would like to meet this person”). We reverse scored the negatively worded items and averaged across these items to create an interaction intentions score (1A: $M = 2.88$, $SD = 1.03$, $\alpha = .84$; 1B: $M = 2.81$, $SD = 0.87$, $\alpha = .80$).

Attention Check

To assess whether participants were aware of the *admission* or *refusal* (Study 1A and Study 1B) or *no response* (Experiment 1B), we administered a memory test at the end of the experiments. Participants indicated whether

Participant A engaged in wrongness admission. As noted, we also used this memory test as an attention check to exclude participants.

Results

We tested our hypotheses by conducting ANOVA with Admission Condition as the independent variable and user reputation and interaction intentions as the dependent variables (for inferential statistics, see Table 2). In Experiment 1A, there were significant Admission Condition effects on user reputation ratings and interaction intentions. Supporting Hypothesis 1, user reputation ratings were significantly higher in the admission condition than the refuse condition. Supporting Hypothesis 2a, interaction intentions were higher in the admission condition than the refuse condition. To test Hypothesis 2b, we performed a mediation analysis, with 10,000 bootstrapping samples, and found a significant indirect effect of Admission Condition on interaction intentions through user reputation ratings (see Table 4).

For Experiment 1B, failing to replicate Experiment 1A, none of our hypotheses were supported.

Discussion and Experiment 2A and Experiment 2B

Our hypotheses were supported in Experiment 1A. It appears that wrongness admission on OSNs serves as a cue of communal and competence traits and leads observers to be willing to interact with the wrongness admitting user. However, we were unable to replicate these findings in Experiment 1B. Importantly, however, wrongness admission did not lead to negative impression formation in this experiment, as user reputation scores were descriptively higher in admission condition than the nonadmission conditions.

We were surprised that we did not find a significant difference in user reputation ratings between the admission and nonadmission conditions in Experiment 1B.

Table 2. Effect of condition on impression formation and means and SDs by condition, Experiment 1A and Experiment 1B

	Experiment 1A (1, 87)						Experiment 1B (2, 235)					
	<i>F</i>	η^2	90% CI	<i>M</i> (<i>SD</i>) _{Admit}	<i>M</i> (<i>SD</i>) _{Refuse}	<i>M</i> (<i>SD</i>) _{NR}	<i>F</i>	η^2	90% CI	<i>M</i> (<i>SD</i>) _{Admit}	<i>M</i> (<i>SD</i>) _{Refuse}	<i>M</i> (<i>SD</i>) _{NR}
User reputation ratings	6.10*	.07	0.01, 0.17	3.28 (0.75)	2.88 (0.75)	NA	2.12	0.02	0.00, 0.05	2.93 (0.71)	2.70 (0.71)	2.85 (0.73)
Interaction intentions	8.69**	.09	0.02, 0.20	3.22 (1.01)	2.59 (0.88)	NA	0.63	0.01	0.00, 0.03	2.79 (0.86)	2.89 (0.85)	2.74 (0.92)

Note. *** $p < .001$; ** $p < .01$; * $p < .05$. DF = Degrees of Freedom, NA = Not applicable, NR = no-response condition.

Table 3. Effect of condition on impression formation and means and SDs by condition, Experiment 2A and Experiment 2B

Study (DF)	Experiment 2A (2, 168)						Experiment 2B (2, 180)					
	<i>F</i>	η^2	90% CI	<i>M</i> (SD) _{Admit}	<i>M</i> (SD) _{Refuse}	<i>M</i> (SD) _{NR}	<i>F</i>	η^2	90% CI	<i>M</i> (SD) _{Admit}	<i>M</i> (SD) _{Refuse}	<i>M</i> (SD) _{NR}
User reputation rating	8.21***	.09	0.03, 0.16	3.24 (0.71)	2.75 (.65)	2.92 (0.64)	5.32**	0.06	0.01, 0.11	3.17 (0.72)	3.05 (0.80)	2.95 (0.17)
Interaction intentions	3.45*	.04	0.00, 0.09	3.14 (0.88)	2.86 (.83)	2.75 (0.82)	1.40	0.02	0.000, 0.051	3.04 (0.72)	2.95 (0.69)	3.05 (0.80)

Note. *** $p < .001$; ** $p < .01$; * $p < .05$. NR = no-response condition.

Therefore, we reviewed the responses to an optional feedback question at the end of both experiments to determine what, if anything, led to differences in the results. Notably, some participants in Experiment 1B indicated that the wrongness admission response of the user seemed sarcastic and insincere. Therefore, in Experiment 2A and Experiment 2B, we edited the final post in the admission and refuse conditions. Participants rated the politeness and sincerity of the final post in Experiment 2A and Experiment 2B. These ratings were well above the midpoint (>6.00 for politeness and >5.50 for sincerity, both out of 7) in the admission condition. We do not report further results with these ratings as we cannot compare them to the participants' reactions to the final posts in the previous experiments and because the ratings were in regard to the post, not the OSN user.

All of our experiments, including Experiment 2A, relied on Amazon Mechanical Turk's participant pool. While this participant pool is more diverse and attentive (Buhrmester et al., 2011), there are problems, too (Arechar et al., 2017). Therefore, the purpose of Experiment 2B was to replicate Experiment 2A in a nononline sample.

Method

Participants

We recruited 191 participants in Experiment 2A from Amazon Mechanical Turk. Of the 191 participants, 171 participants (100 female) completed the experiment and passed the attention check and 98% reported using Facebook. For Experiment 2B, we created 226 slots via a psychology department's online SONA system. Participants were undergraduates at a large university in the Southwest United States. Of the 226 undergraduate students who participated, 182 (116 female) completed the experiment and passed the attention check and 85% reported using Facebook. These sample sizes were determined by our ability to recruit as close to 200 participants as possible. According to a sensitivity analysis, we had the power to detect a minimum effect size of $\eta^2 = .05$ in both studies.

Materials and Procedures

The materials and procedures were nearly identical to those of Experiment 1B. However, we modified the final posts in the admission and refuse conditions. The final post

in the admission condition was as follows: "So, I read through the sources you posted . . . The evidence is pretty strong. I guess I am wrong and you are right on this. Thanks for posting those links and thanks for the conversation!" The final post in the refuse condition was identical except that the third sentence read, "I still think I am right and you are wrong."

We created scores for user reputation ratings (2A: $M = 2.99$, $SD = 0.70$, $\alpha = .94$; 2B: $M = 3.03$, $SD = 0.61$, $\alpha = .92$) and interaction intentions (2A: $M = 2.93$, $SD = 0.86$, $\alpha = .77$; 2B: $M = 3.06$, $SD = 0.74$, $\alpha = .69$), as in the previous experiments. Participants also responded to the same attention check question, which we used to exclude participants who were not paying attention.

Results

We tested our hypotheses by conducting ANOVAs with Admission Condition as the independent variable and user reputation ratings and interaction intentions as the dependent variables (for inferential statistics, see Table 3). In Experiment 2A, there were significant Admission Condition effects on user reputation ratings and interaction intentions. To test whether this condition effect supported Hypothesis 1 and Hypothesis 2a, we ran contrast analyses comparing the admission condition to the two nonadmission conditions. Supporting Hypothesis 1, user reputation ratings were significantly higher in the admission condition than the two nonadmission conditions, $b = 0.13$, $t(168) = 3.81$, $p < .001$. Supporting Hypothesis 2a, interaction intentions were significantly higher in the admission condition than the two nonadmission conditions, $b = 0.11$, $t(168) = 2.52$, $p = .013$. To test Hypothesis 2b, we performed a mediation analysis, with 10,000 bootstrapping samples, with the admit condition contrast coded as 1 and the nonadmit conditions as 0, and found a significant indirect effect of Admission Condition on interaction intentions through user reputation ratings (see Table 4).

In Experiment 2B, there was a significant Admission Condition effect on user reputation ratings, but not interaction intentions. To test whether this condition effect supported Hypothesis 1, we ran contrast analyses comparing the admission condition to the two nonadmission

Table 4. Mediation analyses of condition effects on interaction intentions through user reputation ratings

Experiment	A path	B path	C path	C' path	95% CI for the indirect effects
1A	0.26**	0.85***	0.30**	0.09	0.09, 0.79
1B	—	—	—	—	—
2A	0.28***	0.79***	0.19*	−0.03	0.19, 0.62
2B	0.23**	0.55***	0.11	−0.01	0.07, 0.34

Note. Standardized regression coefficients are depicted for each path (** $p < .01$; *** $p < .001$). A path = condition to mediator; B path = mediator to outcome; C path = condition to outcome; C' path = condition to outcome controlling for the mediator.

conditions. Supporting Hypothesis 1, user reputation ratings were significantly higher in the admission condition than the two nonadmission conditions, $b = 0.09$, $t(180) = 3.12$, $p = .002$. To test Hypothesis 2b, we performed a mediation analysis and found a significant indirect effect of Admission Condition on interaction intentions through user reputation ratings (see Table 4).

Mini Meta-Analysis

Because there were variations in the materials, sample sizes, and effect sizes across the experiments, we conducted a meta-analysis of the effect of admission condition compared to the nonadmission conditions on user reputation scores and interaction intentions to get a truer estimate of the actual effect sizes. We used Goh et al.'s (2016) mini meta-analytic strategy, in which we converted the effect size estimates across studies to r s and then calculated a weighted (by N) mean effect size. The meta-analytic effect size was moderate for reputation, $r = .20$ (95% CI [.14, .26]), and small for interaction intentions, $r = .11$ (90% CI [.05, .17]).

Discussion

Our hypotheses were partially supported in Experiment 2A and Experiment 2B. Wrongness admission in an OSN argument served as a cue of communal and competence traits and led observers to indicate a willingness to interact with the admitting user. However, in Experiment 2B, there was no direct impact of wrongness admission on interaction intentions. Even so, wrongness admission on OSNs appears to lead to positive impression formation outcomes.

General Discussion

With the ubiquitous use of OSNs and the growing presence of arguments on these sites, it is important to understand how online behavior in these online situations impacts impression formation. Here, we investigated the impact of

a form of intellectual humility: wrongness admission. In four studies, we found general support for our hypotheses. Those who witnessed an OSN user engage in wrongness admission rated that user as higher in communion and competence traits compared witnessing a user not engaging in wrongness admission, supporting Hypothesis 1. Furthermore, we found that those in the wrongness admission condition were more likely to indicate interest in interacting with the admitting user compared to those in the nonadmission conditions, supporting Hypothesis 2a. However, this latter effect was weaker, and all effects on interaction intentions across experiments were likely due to participants' increased communion and competence ratings of the user, supporting Hypothesis 2b.

Implications and Theoretical Considerations

It seems obvious that wrongness admission, as opposed to refusing to admit, is a better strategy for those having arguments and managing their impressions on OSNs. Intellectual humility is considered a virtue (Chancellor & Lyubomirsky, 2013; Roberts & Wood, 2007), and people seem to like intellectually humble people (Meagher et al., 2021). The results of Fetterman and Sassenberg (2015) and John et al. (2019) suggest that people know that wrongness admission is the best option. Yet, this behavior is not as common as it should be. In fact, Fetterman et al. (2019) found in a daily diary study that, when given a chance to admit, 43% of participants refused to engage in wrongness admission. When asked why they refuse to admit, most people indicate impression management concerns (John et al., 2019).

If people are hesitant to engage in wrongness admission in face-to-face situations, they may be even more hesitant on OSNs, since many passive viewers can witness the admission and OSNs are a place where people spend considerable time managing their impressions (Krämer & Winter, 2008; Lee & Jang, 2019; Ranzini & Hoek, 2017; Seidman, 2013). Yet, our results suggest that, because wrongness admission serves as a cue of communion and competence, those who might refuse to engage in wrongness admission during an OSN argument might be making an impression management mistake. Wrongness

admission, in these situations, appears to *improve* the impressions formed of users by passive viewers.

It is understandable that these findings might seem counterintuitive to some, especially since the act of wrongness admission is itself an admission of incompetence. We suggest a similar mechanism for wrongness admission that Brooks et al. (2015) suggested for advice seeking. Wrongness admission serves as a cue of intellectual humility, communion, and competence. Although the admitter (like the advice seeker) is telling onlookers that they have been incompetent in this instance, it suggests that they are willing to work together and that they are competent enough to recognize faulty knowledge and change it. Because OSN users are seeking social connections (Alhabash & Ma, 2017) and, perhaps, future friends (Bacev-Giles & Haji, 2017), they are forming impressions online (Tennie et al., 2010) with relatively little contextual information (Walther & Whitty, 2021). Any behavior that serves as a cue of these communal and competence traits should lead to positive impression formation. Wrongness admission, as we have noted, is one such cue.

Wrongness admission on OSNs might be important for reasons beyond impression formation and management, however. As more people engage in debate on the internet, the more OSNs have an impact on behaviors and the spread of misinformation (Allcott & Gentzkow, 2017). OSN-based debates do not shy away from controversial and consequential topics such as politics, religion, or other important societal issues (Anderson et al., 2018; Neubaum et al., 2021), and these debates can be equally frustrating and engaging (Wang & Silva, 2018). If people avoid wrongness admission in these debates, there will likely be no, or worse, resolution to these issues. It could further affect the widening division in the world of politics (Van Bavel & Pereira, 2018) and inspire violent reactions to false information – particularly for those who become more extreme in their faulty beliefs instead of engaging in wrongness admission (i.e., the Backfire Effect; Nyhan & Reifler, 2010). Further research is underway to investigate ways to encourage wrongness admission on OSNs.

Limitations and Future Directions

In the current experiments, we focused on internal, rather than on external, validity. We wanted to limit the impact of other information readily available on OSNs (e.g., profiles, pictures, and demographics). This decision is in line with prior research investigating the effects of minimal information on OSNs (e.g., Austin et al., 2021; Kaye et al., 2020). However, it limits the current investigation because it is unclear how this extra information might impact

impression formation in relation to wrongness admission on OSNs. For example, it could be that the identity of the admitting user moderates the effects we found here in some way. Future research is needed, which subtly adds in information to the online stimuli to test this and other hypotheses. Indeed, we consider the current investigation as a launching point for much more research on the outcomes of wrongness admission and intellectual humility on OSNs.

In a similar vein, our studies were limited to supposedly real, but fake, topics and an argument between strangers. This means that the participants did not have prior attitudes toward the topic of conversation, nor the users. It is likely that prior attitudes toward both affect the impressions formed by passive viewers of OSN users who engage in wrongness admission. For example, it may be that Donald Trump engaging in wrongness admission on OSNs is seen positively by his supporters and negatively by his detractors – or vice versa, given the reactions to his late 2021 support of COVID-19 vaccinations. It may be that a user engaging in wrongness admission about the benefits of intermittent fasting is seen negatively by intermittent fasters, but positively by nonfasters. Future research should investigate these possibilities.

We presented mediation models suggesting that wrongness admission (vs. nonadmission) led to better user reputation ratings, which then related to interaction intentions. From a theoretical perspective, this makes sense. The purpose of impression formation is to determine who one wants to interact with, even on OSNs (Tennie et al., 2010). However, we did not test the full causal model, and many have cast doubts on the type of cross-sectional mediation analyses we conducted (e.g., Maxwell & Cole, 2007; Rohrer, 2018). Therefore, we suggest due caution in interpreting these results, consider them preliminary, and encourage further research investigating the full causal chain.

Conclusion

People are wrong a lot and for a variety of reasons (Schulz, 2010). However, the ratio of instances of wrongness to admissions is far from equal (Fetterman et al., 2019). It is likely that people do not want to broadcast their incompetence on OSNs. However, wrongness admission on OSNs not only allows people to change their attitudes to become more factual (being competent) but also leads to the formation of more positive impressions (appearing communal and competent). Therefore, wrongness admission on OSNs appears to lead to better impression formation outcomes than not admitting. At least, that is

what we can conclude until someone provides evidence that we are wrong. If such a time comes, we will never admit it.

References

- Abele, A. E., Hauke, N., Peters, K., Louvet, E., Szymkow, A., & Duan, Y. (2016). Facets of the fundamental content dimensions: Agency with competence and assertiveness-communion with warmth and morality. *Frontiers in Psychology*, 7, 1810. <https://doi.org/10.3389/fpsyg.2016.01810>
- Abele, A. E., & Wojciszke, B. (2014). Communal and agentic content in social cognition: A dual perspective model. *Advances in experimental social psychology* (Vol. 50, pp. 195–255). Academic Press. <https://doi.org/10.1016/b978-0-12-800284-1.00004-7>
- Alhabash, S., & Ma, M. (2017). A tale of four platforms: Motivations and uses of Facebook, Twitter, Instagram, and Snapchat among college students? *Social Media + Society*, 3(1), 2056305117691544. <https://doi.org/10.1177/2056305117691544>
- Allcott, H., & Gentzkow, M. (2017). Social media and fake news in the 2016 election. *Journal of Economic Perspectives*, 31(2), 211–236. <https://doi.org/10.1257/jep.31.2.211>
- Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological bulletin*, 111(2), 256–274. <https://doi.org/10.1037/0033-2909.111.2.256>
- Anderson, M., Toor, S., Ranie, L., & Smith, A. (2018). *Activism in the social media age*. Pew Research Center. <http://www.pewinternet.org/2018/07/11/activism-in-the-social-media-age/>
- Arechar, A. A., Kraft-Todd, G. T., & Rand, D. G. (2017). Turing overtime: How participant characteristics and behavior vary over time and day on Amazon mechanical Turk. *Journal of the Economic Science Association*, 3(1), 1–11. <https://doi.org/10.1007/s40881-017-0035-0>
- Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, 41, 258–290. <https://doi.org/10.1037/h0055756>
- Austin, A. B., Costabile, K. A., & Smith, L. (2021). Social judgments, social media, and self-deprecation: Role of information source and valence on trait and favorability judgments. *Journal of Media Psychology: Theories, Methods, and Applications*. <https://doi.org/10.1027/1864-1105/a000299>
- Bacev-Giles, C., & Haji, R. (2017). Online first impressions: Person perception in social media profiles. *Computers in Human Behavior*, 75, 50–57. <https://doi.org/10.1016/j.chb.2017.04.056>
- Bareket-Bojmel, L., Moran, S., & Shahar, G. (2016). Strategic self-presentation on Facebook: Personal motives and audience response to online behavior. *Computers in Human Behavior*, 55, 788–795. <https://doi.org/10.1016/j.chb.2015.10.033>
- Batenburg, A., & Bartels, J. (2017). Keeping up online appearances: How self-disclosure on Facebook affects perceived respect and likability in the professional context. *Computers in Human Behavior*, 74, 265–276. <https://doi.org/10.1016/j.chb.2017.04.033>
- Baumeister, R. F., & Leary, M. R. (1995). The need to belong: Desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin*, 117(3), 497–529. <https://doi.org/10.1037/0033-2909.117.3.497>
- Brooks, A. W., Gino, F., & Schweitzer, M. E. (2015). Smart people ask for (my) advice: Seeking advice boosts perceptions of competence. *Management Science*, 61(6), 1421–1435. <https://doi.org/10.1287/mnsc.2014.2054>
- Brunswik, E. (1956). *Perception and the representative design of psychological experiments*. University of California Press.
- Buffardi, L. E., & Campbell, W. K. (2008). Narcissism and social networking web sites. *Personality and Social Psychology Bulletin*, 34(10), 1303–1314. <https://doi.org/10.1177/0146167208320061>
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's mechanical Turk. *Perspectives on Psychological Science*, 6(1), 3–5. <https://doi.org/10.1177/1745691610393980>
- Chancellor, J., & Lyubomirsky, S. (2013). Humble beginnings: Current trends, state perspectives, and hallmarks of humility. *Social and Personality Psychology Compass*, 7(11), 819–833. <https://doi.org/10.1111/spc3.12069>
- Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology*, 40, 61–149. [https://doi.org/10.1016/S0065-2601\(07\)00002-0](https://doi.org/10.1016/S0065-2601(07)00002-0)
- Duggan, M., & Smith, A. (2016). *The political environment on social media*. Pew Internet & American Life Project. <http://www.pewinternet.org/2016/10/25/the-political-environment-on-social-media>
- Fetterman, A. K., Curtis, S., Carre, J., & Sassenberg, K. (2019). On the willingness to admit wrongness: Validation of a new measure and an exploration of its correlates. *Personality and Individual Differences*, 138, 193–202. <https://doi.org/10.1016/j.paid.2018.10.002>
- Fetterman, A.K. (2021). *Data for "When you are wrong on Facebook, just admit it wrongness admission leads to better interpersonal impressions on social media"*. <https://osf.io/a4tgp/>
- Fetterman, A. K., & Sassenberg, K. (2015). The reputational consequences of failed replications and wrongness admission among scientists. *PLoS One*, 10(12), e0143723. <https://doi.org/10.1371/journal.pone.0143723>
- Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: Warmth and competence. *Trends in Cognitive Sciences*, 11(2), 77–83. <https://doi.org/10.1016/j.tics.2006.11.005>
- Funder, D. C. (2012). Accurate personality judgment. *Current Directions in Psychological Science*, 21(3), 177–182. <https://doi.org/10.1177/0963721412445309>
- Gable, S. L., & Reis, H. T. (2001). Appetitive and aversive social interaction. In J. Harvey & A. Wenzel (Eds.), *Close romantic relationships: Maintenance and enhancement* (pp. 169–194). Erlbaum.
- Gigerenzer, G., & Kurz, E. M. (2001). Vicarious functioning reconsidered: A fast and frugal lens model. In K. R. Hammond & T. R. Stewart (Eds.), *The essential Brunswik: Beginnings, explications, applications* (pp. 342–348). Oxford University Press.
- Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass*, 10(10), 535–549. <https://doi.org/10.1111/spc3.12267>
- Hagá, S., & Olson, K. R. (2017). 'If I only had a little humility, I would be perfect': Children's and adults' perceptions of intellectually arrogant, humble, and diffident people. *The Journal of Positive Psychology*, 12(1), 87–98. <https://doi.org/10.1080/17439760.2016.1167943>
- Harris, K., & Vazire, S. (2016). On friendship development and the Big Five personality traits. *Social and Personality Psychology Compass*, 10(11), 647–667. <https://doi.org/10.1111/spc3.12287>
- John, L. K., Jeong, M., Gino, F., & Huang, L. (2019). The self-presentational consequences of upholding one's stance in spite of the evidence. *Organizational Behavior and Human Decision Processes*, 154, 1–14. <https://doi.org/10.1016/j.obhdp.2019.07.001>
- Kaye, L. K., Wall, H. J., & Hird, A. T. (2020). Less is more when rating extraversion: Behavioral cues and interpersonal perceptions on the platform of facebook. *Psychology of Popular Media*, 9(4), 465–474. <https://doi.org/10.1037/ppm0000263>
- Krämer, N. C., & Winter, S. (2008). Impression management 2.0: The relationship of self-esteem, extraversion, self-efficacy, and self-presentation within social networking sites. *Journal of Media*

- Psychology*, 20(3), 106–116. <https://doi.org/10.1027/1864-1105.20.3.106>
- Leary, M. R., Diebels, K. J., Davisson, E. K., Jongman-Sereno, K. P., Isherwood, J. C., Raimi, K. T., Deffler, S. A., & Hoyle, R. H. (2017). Cognitive and interpersonal features of intellectual humility. *Personality and Social Psychology Bulletin*, 43(6), 793–813. <https://doi.org/10.1177/0146167217697695>
- Lee, S. Y., & Jang, K. (2019). Antecedents of impression management motivations on social network sites and their link to social anxiety. *Media Psychology*, 22(6), 890–904. <https://doi.org/10.1080/15213269.2019.1580588>
- Levordashka, A., & Utz, S. (2017). Spontaneous trait inferences on social media. *Social Psychological and Personality Science*, 8(1), 93–101. <https://doi.org/10.1177/1948550616663803>
- Maxwell, S. E., & Cole, D. A. (2007). Bias in cross-sectional analyses of longitudinal mediation. *Psychological Methods*, 12(1), 23–44. <https://doi.org/10.1037/1082-989X.12.1.23>
- Meagher, B. R., Leman, J. C., Heidenga, C. A., Ringquist, M. R., & Rowatt, W. C. (2021). Intellectual humility in conversation: Distinct behavioral indicators of self and peer ratings. *The Journal of Positive Psychology*, 16(3), 417–429. <https://doi.org/10.1080/17439760.2020.1738536>
- Munro, K., Hartt, C. M., & Pohlkamp, G. (2015). Social media discourse and genetically modified organisms. *The Journal of Social Media in Society*, 4(1), 38–65.
- Neubaum, G., Cargnino, M., Winter, S., & Dvir-Gvirsman, S. (2021). “You’re still worth it”: The moral and relational context of politically motivated unfriending decisions in online networks. *PLoS One*, 16(1), e0243049. <https://doi.org/10.1371/journal.pone.0243049>
- Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. *Political Behavior*, 32(2), 303–330. <https://doi.org/10.1007/s11109-010-9112-2>
- Orben, A. C., & Dunbar, R. I. M. (2017). Social media and relationship development: The effect of valence and intimacy of posts. *Computers in Human Behavior*, 73, 489–498. <https://doi.org/10.1016/j.chb.2017.04.006>
- Porter, T., Baldwin, C. R., Warren, M. T., Murray, E. D., Cotton Bronk, K., Forgeard, M. J. C., Snow, N. E., & Jayawickreme, E. (2021). Clarifying the content of intellectual humility: A systematic review and integrative framework. *Journal of Personality Assessment*, 1–13. <https://doi.org/10.1080/00223891.2021.1975725>
- Qin, Y., Cho, H., Li, P., & Zhang, L. (2021). First impression formation based on valenced self-disclosure in social media profiles. *Frontiers in Psychology*, 12. <https://doi.org/10.3389/fpsyg.2021.656365>
- Quadflieg, S., & Penton-Voak, I. S. (2017). The emerging science of people-watching: Forming impressions from third-party encounters. *Current Directions in Psychological Science*, 26(4), 383–389. <https://doi.org/10.1177/0963721417694353>
- Ranzini, G., & Hoek, E. (2017). To you who (I think) are listening: Imaginary audience and impression management on Facebook. *Computers in Human Behavior*, 75, 228–235. <https://doi.org/10.1016/j.chb.2017.04.047>
- Roberts, R. C., & Wood, W. J. (2007). *Intellectual virtues: An essay in regulative epistemology*. Oxford University Press.
- Rodriguez, D., Hook, J. N., Farrell, J. E., Mosher, D. K., Zhang, H., Van Tongeren, D. R., Davis, D. E., Aten, J. D., & Hill, P. C. (2019). Religious intellectual humility, attitude change, and closeness following religious disagreement. *The Journal of Positive Psychology*, 14(2), 133–140. <https://doi.org/10.1080/17439760.2017.1388429>
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27–42. <https://doi.org/10.1177/2515245917745629>
- Schulz, K. (2010). *Being wrong: Adventures in the meaning of error*. ECCO.
- Seidman, G. (2013). Self-presentation and belonging on Facebook: How personality influences social media use and motivations. *Personality and Individual Differences*, 54(3), 402–407. <https://doi.org/10.1016/j.paid.2012.10.009>
- Stopfer, J. M., Egloff, B., Nestler, S., & Back, M. D. (2013). Being popular in online social networks: How agentic, communal, and creativity traits relate to judgments of status and liking. *Journal of Research in Personality*, 47(5), 592–598. <https://doi.org/10.1016/j.jrp.2013.05.005>
- Stopfer, J. M., Egloff, B., Nestler, S., & Back, M. D. (2014). Personality expression and impression formation in online social networks: An integrative approach to understanding the processes of accuracy, impression management and meta-accuracy. *European Journal of Personality*, 28(1), 73–94. <https://doi.org/10.1002/per.1935>
- Tennie, C., Frith, U., & Frith, C. D. (2010). Reputation management in the age of the world-wide web. *Trends in Cognitive Sciences*, 14(11), 482–488. <https://doi.org/10.1016/j.tics.2010.07.003>
- Utz, S. (2010). Show me your friends and I will tell you what type of person you are: How one’s profile, number of friends, and type of friends influence impression formation on social network sites. *Journal of Computer-Mediated Communication*, 15(2), 314–335. <https://doi.org/10.1111/j.1083-6101.2010.01522.x>
- Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: An identity-based model of political belief. *Trends in Cognitive Sciences*, 22(3), 213–224. <https://doi.org/10.1016/j.tics.2018.01.004>
- Waggoner, A. S., Smith, E. R., & Collins, E. C. (2009). Person perception by active versus passive perceivers. *Journal of Experimental Social Psychology*, 45(4), 1028–1031. <https://doi.org/10.1016/j.jesp.2009.04.026>
- Walther, J. B., Van Der Heide, B., Hamel, L. M., & Shulman, H. C. (2009). Self-generated versus other-generated statements and impressions in computer-mediated communication: A test of warranting theory using facebook. *Communication Research*, 36(2), 229–253. <https://doi.org/10.1177/0093650208330251>
- Walther, J. B., & Whitty, M. T. (2021). Language, psychology, and new new media: The hyperpersonal model of mediated communication at twenty-five years. *Journal of Language and Social Psychology*, 40(1), 120–135. <https://doi.org/10.1177/0261927X20967703>
- Wang, M. Y., & Silva, D. E. (2018). A slap or a jab: An experiment on viewing uncivil political discussions on Facebook. *Computers in Human Behavior*, 81, 73–83. <https://doi.org/10.1016/j.chb.2017.11.041>

History

Received June 11, 2021

Revision received January 3, 2022

Accepted February 3, 2022

Published online March 18, 2022

Authorship

Adam K. Fetterman and Nicole L. Muscanell conceived of and designed the studies. A.K. Fetterman managed the data collection. A.K. Fetterman and Dongjie Wu conducted the analyses. Adam K. Fetterman led the writing of the manuscript. All authors edited the manuscript.

Open Data

We report how we determined our sample sizes, all data exclusions, all manipulations, and all measures in the studies. Data, code, and materials can be accessed through the Open Science Framework (OSF): <https://osf.io/a4tpg/>. We did not preregister our hypotheses.

ORCID

Adam K. Fetterman

 <https://orcid.org/0000-0001-5957-7670>

Adam K. Fetterman

Department of Psychology

University of Houston

3695 Cullen Boulevard

Houston, TX 77204-5022

USA

akfetterman@uh.edu

https://econtent.hogrefe.com/contentReq.requestUri} - Sunday, May 05, 2024 2:26:50 AM - IP Address: 18.118.226.105



Not Getting What You Want

Aggression, Prosocial Behaviors, and Popularity

Martin H. Jones¹, Toby J. Cooke¹, and Jennifer Symonds²

¹Department of Individual, Family and Community Education, University of New Mexico, Albuquerque, NM, USA

²Department of Educational Psychology, School of Education, University College Dublin, Ireland

Abstract. As adolescents desire the benefits of having greater social status, some teenagers cannot acquire their desired level of popularity. The current study uses a single high school to examine how the discrepancy between popularity goals and actual popularity aligns with aggression and prosocial behaviors. The current study suggests that the discrepancy between popularity and popularity goals aligns with having more aggression and prosocial behaviors when adolescents are in less popular peer groups. Within one's own peer group, adolescents with greater discrepancy between popularity and popularity goals utilize more aggression. The results provide additional insight toward understanding how pursuing popularity might align with using aggression and prosocial behaviors toward peers in school.

Keywords: popularity, popularity goals, social goals, aggression, prosocial behaviors

Popular adolescents hold greater social capital, are more admired by their peers, and appear more attractive (Cillessen et al., 2011; Dawes & Xie, 2014, 2017). Given these potential benefits, adolescents often desire popularity more than having friendships and academic achievement (LaFontana & Cillessen, 2010). Indeed, the desire for acquiring social status peaks during adolescence (Brendgen et al., 2018; Laursen, 2018) but, comes with some interpersonal consequences. Desiring social status, or one's popularity goals, can lead to both greater aggression and greater prosocial behaviors (Dawes, 2017; Jarvinen & Nicholls, 1996; Kiefer & Wang, 2016). Popularity goals may not equate with having popularity (Dawes, 2017; Dawes & Xie, 2014), which creates a discrepancy between adolescents' social goals and actual social status. This discrepancy between social status and popularity goals is rarely examined, especially regarding how such a discrepancy might relate with aggression and prosocial behaviors.

Theoretical support for the discrepancy between popularity goals, social status, aggression, and prosocial behaviors comes from Social Information Processing (SIP) theory (Crick & Dodge, 1994). In part, SIP suggests individuals pursue, and then adjust, their behaviors in pursuit of social goals (e.g., popularity goals). Thus, when discrepancies emerge between popularity goals and social status, adolescents may adjust other behaviors (e.g., aggression and prosocial behaviors) to pursue desired social outcomes. Theoretically, students might therefore employ more aggression or prosocial behaviors to overcome discrepant popularity goals and actual popularity in hopes of obtaining greater social status.

Social Status

Social status is often defined as someone or some peer group having perceived popularity or social preference (Cillessen & Marks, 2011). Perceived popularity refers to how well-known an individual or peer group is within their social network, whereas social preference is how well-liked an individual is by their peers. Perceived popularity closely aligns with popularity goals as many adolescents want social status, regardless of being well-liked (Dawes & Xie, 2014; Estell, 2020). The current study exclusively focuses on perceived popularity.

Perceived popularity can subdivide into social network popularity and within-group popularity (Jones & Estell, 2010; Cairns & Cairns, 1994). Social network popularity is the perceived popularity of one's peer group in comparison to other peer groups. Alternatively, within-group popularity is the perceived popularity of someone inside their peer group.

Popularity Goals, Aggression, and Prosocial Behaviors

Popularity goals are the psychological desires for acquiring greater social standing and prominence (Dawes, 2017; Dawes & Xie, 2014). Popularity goals can help a student gain status by using aggression against peers (e.g., Kiefer & Wang, 2016). Popularity goals can also help students acquire social status through being prosocial (Dawes, 2017; Dawes & Xie, 2014; Jarvinen & Nicholls, 1996). For some

students, entitled bistrategic controllers, aggression and prosocial behaviors are deliberately employed with peers to gain social status or other desired outcomes (Hawley, 2003). Knowing when to be aggressive and/or prosocial can help students gain social status (Hawley & Bower, 2018), but less literature examines if aggression and prosocial behaviors increase when students have not acquired their desired social status.

Discrepancy Between Popularity and Popularity Goals

Popularity goals can align with having greater popularity, but this alignment does not occur for all students (e.g., Dawes & Xie, 2014). A discrepancy can form between what a student wants (i.e., popularity goals) and their actual popularity. This discrepancy is particularly important to examine as students with high popularity goals could exert even more aggression and prosocial behaviors to compensate for their lower than desired social status. The current study explores this possibility. In addition, the study examines whether discrepancies between popularity goals and social status differ by actual social status levels.

Method

Participants and Procedures

Ninth through 12th graders participated from one small rural high school in the southwestern United States (70% participation rate). Self-reported genders were 156 girls, 112 boys, 7 genderqueer/nonbinary, and 1 nonreporting. Self-reported ethnicities were as follows: Hispanic/Latinx – 130, White – 69, Indigenous – 35, multicultural – 27, Asian-American – 9, Black/African-American – 5, and other ethnicity – 1. In regard to socioeconomic status, all students were eligible for free/reduced-price lunch. All participants had parental consent, knew their answers were confidential, and received no financial compensation. Survey response time was 30–35 minutes. Participants answered surveys approximately 1 month into the spring semester.

Instruments

Popularity

Popularity assessments came from a social network analysis procedure entitled social cognitive mapping (SCM; Cairns et al. 1988; Cairns & Cairns, 1994). SCM asks all students in

a social network, “Are there some kids in your classroom who hang around together a lot? If yes, who are they?” (Cairns et al., 1995). SCM has students report on each other’s peer group membership and the group’s members. SCM collates these nominations across all student respondents. In this way, students report and SCM procedures objectively reciprocate peer group membership. Students receiving the most nominations have the greatest perceived popularity. Similarly, peer groups with the most nominations have the greatest perceived popularity. Four levels of perceived popularity occur for both social network and within-group popularity: social isolate (no peer group membership), peripheral member (least popular), secondary member (somewhat popular), and nuclear member (most popular). Past research suggests that SCM findings parallel observed peer affiliations in school and findings have strong reliability for both social network popularity and within-group popularity (Chen et al., 2008; Gest et al., 2003).

Popularity Goals

Popularity goals were measured using six questions in regard to gaining perceived popularity (Li & Wright, 2014). Example items are “I want to be popular among my peers” and “I want to be socially central among my peers.” Participants responded on a 5-point Likert scale ranging from 1 (= *never*) to 5 (= *all the time*), with higher scores indicating greater desire for popularity ($\alpha = .86$). The scale was previously validated and deemed reliable (Li & Wright, 2014).

Aggression

Aggression measurement included eight questions assessing both overt and relational aggression (Li & Wright, 2014). Example items included “How often do you start fights with others?” and “How often do you tell a peer that they cannot be in the group?” Responses were on a 5-point Likert scale ranging from 1 (= *never*) to 5 (= *all the time*). Higher scores represented greater self-reported aggression ($\alpha = .71$). The scale is both valid and reliable (Li & Wright, 2014).

Prosocial Behaviors

Prosocial behaviors were measured with four questions ($\alpha = .77$; Li & Wright, 2014), including “How often do you help, cooperate or share with others?” and “How often do you tell another peer you care about them?” Responses were on a 5-point Likert scale ranging from 1 (= *never*) to 5 (= *all the time*). Higher scores indicated greater self-reported prosocial behaviors. The scale is valid and reliable (Li & Wright, 2014).

Measuring Discrepancy

Discrepancy scores were the difference between popularity goals and popularity (popularity coded as 1 = *social isolate*, 2 = *peripheral*, 3 = *secondary*, and 4 = *nuclear*). This approach results in the range of difference scores being limited by one's popularity level. For example, nuclear peer group members' raw differences could range from 1 to -3, whereas secondary group members' raw difference scores could range from 2 to -2. Converting raw differences scores to z-scores eliminates this problem. Thus, standardized discrepancy scores were generated for both social network popularity and within-group popularity. Positive discrepancy scores represent popularity goals that are greater than actual popularity level, whereas negative discrepancy scores suggest the opposite.

Results

Preliminary Analyses

Given past research (Rose & Smith, 2018; Salmivalli & Peets, 2018), preliminary analyses examined potential gender differences in aggression and prosocial behaviors. A multiple ANOVA using gender as the grouping variable and aggression and prosocial behaviors as outcome variables suggested significant differences, $F(4, 532) = 3.09$, $p = .02$. Post hoc comparisons suggested significant differences for prosocial behaviors, $F(2, 267) = 5.88$, $p = .003$, but not aggression. Post hoc comparisons with Bonferroni adjustment suggested that females had higher prosocial behaviors than males ($p = .01$, $d = 0.37$, $M_{\text{females}} = 3.82$, $SD_{\text{females}} = 0.79$; $M_{\text{males}} = 3.51$, $SD_{\text{males}} = 0.89$), which is considered a small to moderate effect size.

Social Network Popularity and Popularity Goals Discrepancies

A path analysis tested whether the discrepancy between popularity goals and social network popularity aligned with aggression and prosocial behaviors. The model fit was good, $\chi^2(3, N = 276) = 2.05$, $p = .56$, RMSEA = 0.00 (0.00, 0.09), CFI = 1.00. As seen in Figure 1, the discrepancy between popularity goals and social network popularity related to aggression and prosocial behaviors, although the relationship was smaller for prosocial behaviors.

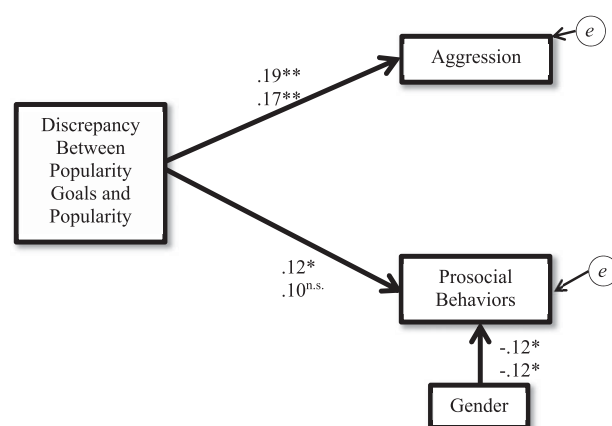


Figure 1. Discrepancies, prosocial behaviors, and aggression. Coefficients above are for social network popularity discrepancies, whereas numbers below are for within-group popularity discrepancies. ** $p < .01$, * $p < .05$, n.s. = not significant.

Discrepancy Between Within-Group Popularity and Popularity Goals

A second path analysis tested whether the discrepancy between popularity goals and within-group popularity aligned with aggression and prosocial behaviors. The model fit was good, $\chi^2(3, N = 276) = 2.29$, $p = .52$, RMSEA = 0.00 (0.00, 0.09), CFI = 1.00. Discrepancy scores significantly aligned with aggression, but not with prosocial behaviors (see Figure 1).

Discrepancies Across Social Statuses

Social Network Popularity

The final analyses tested whether discrepancies varied by social status level. For social network popularity, an ANOVA suggested significant differences, $F(3, 272) = 95.79$, $p < .001$. All social network popularity levels significantly differed from each other using pairwise comparisons with Bonferroni adjustments at $p < .001$ (see Table 1), having moderate to large effect sizes (d 's = 0.47–1.00). Social isolates had a positive discrepancy, whereas nuclear members had a negative discrepancy. Although seemingly counterintuitive for nuclear group members, the negative differences score is likely a ceiling effect as there is no higher social network popularity than nuclear peer group membership.

Within-Group Popularity

An additional ANOVA tested whether discrepancies varied by within-group popularity levels. The ANOVA was significant, $F(3, 272) = 133.14$, $p < .001$, with differences across all social statuses at $p < .001$ (see Table 1), with all having large effect sizes (d 's < 0.90). Social isolates re-

Table 1. Descriptive statistics

	Social network popularity									Within-group popularity								
	Popularity goals		Aggression		Prosocial behaviors		Discrepancy*			Popularity goals		Aggression		Prosocial behaviors		Discrepancy**		
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>N</i>	<i>M</i>	<i>SD</i>
Social isolate	2.18	0.92	3.50	1.64	3.63	0.78	61	1.12	0.70	2.18	0.92	3.50	1.64	3.63	0.78	61	1.20	0.63
Peripheral	2.14	0.85	4.13	1.69	3.76	0.88	60	0.32	0.65	2.32	0.89	2.88	1.61	3.66	0.87	35	0.62	0.61
Secondary	2.26	0.93	2.88	1.63	3.55	0.91	86	−0.34	0.70	2.20	0.87	3.13	1.74	3.61	0.97	56	−0.15	0.59
Nuclear	2.61	0.98	3.13	1.72	3.86	0.79	69	−0.84	0.75	2.40	0.98	4.13	1.67	3.77	0.82	124	−0.70	0.67

Note. *Discrepancy score between popularity goals and social network popularity. **Discrepancy between popularity goals and within-group popularity.

ported having popularity goals greater than their popularity level, whereas within-group nuclear members showed negative discrepancy scores. Again, this likely indicates ceiling and floor effects.

Discussion

The current study examined how discrepancies between adolescents' popularity goals and popularity aligned with aggression and prosocial behaviors. The results suggest the discrepancy between social network popularity and popularity goals aligned with greater aggression and prosocial behaviors, whereas discrepancies between within-group popularity and popularity goals related to aggression. These findings are supported by SIP theory, which suggests that desiring to be in more popular peer groups (without actually having it) aligns with adjusting one's behaviors in pursuit of social goals (Crick & Dodge, 1994). Indeed, the current study expands this notion by suggesting that students with greater discrepancies between desired and actual social status may be more aggressive and more prosocial. The implication of this finding is that aggression and prosocial behaviors are not solely used by bistrategic controllers (Hawley, 2003), but also by those who have not achieved their desire for social status. SIP theory suggests that students with greater discrepancies could have less social status than desired because they use aggression and prosocial behaviors ineffectively in pursuit of their social goals.

In contrast, students holding nondiscrepant popularity goals may have no reason to be concerned about their popularity status and therefore do not use aggression and prosocial tactics to pursue their popularity goals. The current study expands past work by suggesting that the discrepancy (and not just the goal) between popularity goals and actual popularity may bolster increased aggression and prosocial behaviors (Dawes, 2017; Jarvinen & Nicholls, 1996; Kiefer & Wang, 2016).

Discrepancies Across Social Statuses

Holding popularity goals does not always result in achieving popularity (Dawes, 2017; Dawes & Xie, 2014). The current study expands this finding in demonstrating that discrepancies appear for the social status of one's peer group as well as membership within one's peer group. These discrepancies occur for social isolates and also for less socially prominent group members (e.g., peripheral group members). In contrast, more socially central group members (secondary and nuclear members) do not hold such discrepancies as they likely achieved their popularity goal or experience ceiling popularity effects.

Limitations and Future Directions

This study's methodology creates inherent limitations. Foremost, a single high school limits the generalizability of findings. In addition, the cross-sectional data do not reflect the dynamic nature of social relationships in how discrepancies form across time and relate with aggression and prosocial behaviors. Finally, utilizing self-report questionnaires introduces the potential for social desirability bias in reporting aggression and prosocial behaviors. In a similar way, different social network analyses might produce different results, such as having students nominate the most popular students at school or via other objective social network analysis procedures.

The current study does not explore the many subtypes of aggression and prosocial behaviors. Additional research may better clarify how adolescents with greater discrepancy between popularity goals and social status align with overt and relational aggression uniquely or in tandem. Similarly, additional work may clarify whether proactive versus reactive prosocial behaviors related to discrepancies between social status and popularity goals (Boxer et al., 2004; Findley-Van Nostrand & Ojanen, 2018).

References

- Boxer, P., Tisak, M. S., & Goldstein, S. E. (2004). Is it bad to be good? An exploration of aggressive and prosocial behavior subtypes in adolescence. *Journal of Youth and Adolescence*, 33, 91–100. <https://doi.org/10.1023/b:joyo.0000013421.02015.ef>
- Brendgen, M., Ouellet-Morin, I., & Boivin, M. (2018). Peer relations and psychosocial development: Perspectives from genetic approaches. In W. M. Bukowski, B. Laursen, & K. H. Rubin (Eds.), *Handbook of peer interactions, relationships, and groups* (2nd ed., pp. 123–140). The Guilford Press.
- Cairns, R. B., & Cairns, B. D. (1994). *Lifelines and risks: Pathways of youth in our time*. Harvester Wheatsheaf.
- Cairns, R. B., Leung, M., Buchannan, L., & Cairns, B. D. (1995). Friendships and social networks in childhood and adolescence: Fluidity, reliability, and interrelations. *Child Development*, 66, 1330–1345. <https://psycnet.apa.org/doi/10.2307/1131650>
- Cairns, R. B., Cairns, B. D., Neckerman, H. J., Gest, S., & GariCpy, J.-L. (1988). Social networks and aggressive behavior: Peer support or peer rejection? *Developmental Psychology*, 24, 815–823.
- Chen, X., Chang, L., Liu, H., & He, Y. (2008). Effects of the peer group on the development of social functioning and academic achievement: A longitudinal study in Chinese children. *Child Development*, 79(2), 235–251. <https://doi.org/10.1111/j.1467-8624.2007.01123.x>
- Cillessen, A. H. N., & Marks, P. E. L. (2011). Conceptualizing and measuring popularity. In A. H. N. Cillessen, D. Schawartz, & L. Mayeux (Eds.), *Popularity in the peer system* (pp. 25–56). The Guilford Press.
- Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information-processing mechanisms in children's social adjustment. *Psychological Bulletin*, 115, 74–101. <https://doi.org/10.1037/0033-2909.115.1.74>
- Dawes, M. (2017). Early adolescents' social goals and school adjustment. *Social Psychology of Education*, 20, 299–328. <https://doi.org/10.1007/s11218-017-9380-3>
- Dawes, M., & Xie, H. (2014). The role of popularity goal in early adolescents' behaviors and popularity status. *Developmental Psychology*, 50(2), 489–497. <https://doi.org/10.1037/a0032999>
- Dawes, M., & Xie, H. (2017). The trajectory of popularity goal during the transition to middle school. *Journal of Early Adolescence*, 37(6), 852–883. <https://doi.org/10.1177/0272431615626301>
- Estell, D. B. (2020). Popularity/antisocial goals. In M. H. Jones (Ed.), *Social goals in the classroom* (pp. 154–169). Routledge.
- Findley-Van Nostrand, D., & Ojanen, T. (2018). Forms of prosocial behaviors are differentially linked to social goals and peer status in adolescents. *The Journal of Genetic Psychology*, 179(6), 329–342. <https://doi.org/10.1080/00221325.2018.1518894>
- Gest, S. D., Farmer, T. W., Cairns, B. D., & Xie, H. (2003). Identifying children's peer social networks in school classrooms: Links between peer reports and observed interactions. *Social Development*, 12(4), 513–529. <https://doi.org/10.1111/1467-9507.00246>
- Hawley, P. H. (2003). Prosocial and coercive configurations of resource control in early adolescence: A case for the well-adapted Machiavellian. *Merrill-Palmer Quarterly*, 49(3), 279–309. <https://doi.org/10.1353/mpq.2003.0013>
- Hawley, P. H., & Bower, A. R. (2018). Evolution and peer relations: Considering the functional roles of aggression and prosociality. In W. M. Bukowski, B. Laursen, & K. H. Rubin (Eds.), *Handbook of peer interactions, relationships, and groups* (2nd ed., pp. 106–122). The Guilford Publications.
- Jarvinen, D. W., & Nicholls, J. G. (1996). Adolescents' social goals, beliefs about the causes of social success, and satisfaction in peer relations. *Developmental Psychology*, 32(3), 435–441. <https://doi.org/10.1037/0012-1649.32.3.435>
- Jones, M. (2022). Data for “Not getting what you want: Aggression, prosocial goals, and popularity”. <https://osf.io/7Z2MC/>
- Jones, M. H., & Estell, D. B. (2010). When elementary students change peer groups: Intra-group centrality, inter-group centrality, and self-perceptions of popularity. *Merrill-Palmer Quarterly*, 56(2), 164–188.
- Kiefer, S. M., & Wang, J. H. (2016). Associations of coolness and social goals with aggression and engagement during adolescence. *Journal of Applied Developmental Psychology*, 44, 52–62. <https://doi.org/10.1016/j.appdev.2016.02.007>
- LaFontana, K. M., & Cillessen, A. H. N. (2010). Developmental changes in the priority of perceived status in childhood and adolescence. *Social Development*, 19(1), 130–147. <https://doi.org/10.1111/j.1467-9507.2008.00522.x>
- Laursen, B. (2018). Peer influence. In W. M. Bukowski, B. Laursen, & K. H. Rubin (Eds.), *Handbook of peer interactions, relationships, and groups* (2nd ed., pp. 447–469). The Guilford Press.
- Li, Y., & Wright, M. F. (2014). Adolescents' social status goals: Relationships to social status insecurity, aggression, and prosocial behavior. *Journal of Youth and Adolescence*, 43, 146–160. <https://doi.org/10.1007/s10964-013-9939-z>
- Rose, A. J., & Smith, R. L. (2018). Gender and peer relationships. In W. M. Bukowski, B. Laursen, & K. H. Rubin (Eds.), *Handbook of peer interactions, relationships, and groups* (2nd ed., pp. 571–589). The Guilford Press.
- Salmivalli, C., & Peets, K. (2018). Bullying and victimization. In W. M. Bukowski, B. Laursen, & K. H. Rubin (Eds.), *Handbook of peer interactions, relationships, and groups* (2nd ed., pp. 302–321). The Guilford Press.

History

Received August 28, 2021

Revision received January 14, 2022

Accepted January 17, 2022

Published online March 18, 2022

Acknowledgments

Thank you to the SMAART research group at the University of New Mexico for data collection and entry.

Authorship

Study idea conception occurred during a walk to the bus stop at University College Dublin between Martin Jones and Jennifer Symonds. Toby J. Cooke executed this idea as part of his Master's coursework at the University of New Mexico. Martin Jones then wrote a majority of the current paper, with Toby Cooke and Jennifer Symonds providing subsequent amounts of writing.

Open Data

Data for this study are accessible at <https://osf.io/7Z2MC/> (Jones et al., 2022).

Martin H. Jones

Department of Individual
Family and Community Education
MSC 05 3040
1 University of New Mexico
Albuquerque, NM 87131
USA
martinjones@unm.edu

Instructions to Authors

Social Psychology is a publication dedicated to international research in social psychology as well as a forum for scientific discussion and debate.

Social Psychology publishes innovative and methodologically sound research and serves as an international forum for scientific discussion and debate in the field of social psychology. Topics include all basic social psychological research themes, methodological advances in social psychology, as well as research in applied fields of social psychology. The journal focuses on original empirical contributions to social psychological research, but is open to theoretical articles, critical reviews, and replications of published research.

The journal welcomes original empirical and theoretical contributions to basic research in social psychology, to social psychological methods, as well as contributions covering research in applied fields of social psychology, such as economics, marketing, politics, law, sports, the environment, the community, or health. Preference will be given to original empirical and experimental manuscripts, but theoretical contributions, critical reviews, and replications of published research are welcome as well.

Social Psychology aims to increase transparency and openness of the research process and encourages authors to share their data and materials and if possible, pre-register their studies.

Social Psychology publishes the following types of article: Original Articles, Research Reports, Replications.

Manuscript submission: All manuscripts should in the first instance be submitted electronically at <http://www.editorialmanager.com/sopsy>. Detailed instructions to authors are provided at <http://www.hgf.io/sopsy>

Copyright Agreement: By submitting an article, the author confirms and guarantees on behalf of him-/herself and any coauthors that the manuscript has not been submitted or published elsewhere, and that he or she holds all copyright in and titles to the submitted contribution, including any figures, photographs, line drawings, plans, maps, sketches, tables, and electronic supplementary material, and that the article and its contents do not infringe in any way on the rights of third parties. ESM will be published online as received from the author(s) without any conversion, testing, or reformatting. They will not be checked for typographical errors or functionality. The author indemnifies and holds harmless the publisher from any third-party claims.

The author agrees, upon acceptance of the article for publication, to transfer to the publisher the exclusive right to reproduce and distribute the article and its contents, both physically and in nonphysical, electronic, or other form, in the journal to which it has been submitted and in other independent publications, with no limitations on the number of copies or on the form or the extent of distribution. These rights are transferred for the duration of copyright as defined by international law. Furthermore, the author transfers to the publisher the following exclusive rights to the article and its contents:

1. The rights to produce advance copies, reprints, or off-prints of the article, in full or in part, to undertake or allow translations into other languages, to distribute other forms or modified versions of the article, and to produce and distribute summaries or abstracts.
2. The rights to microfilm and microfiche editions or similar, to the use of the article and its contents in videotext, teletext, and similar systems, to recordings or reproduction using other media, digital or analog, including electronic, magnetic, and optical media, and in multimedia form, as well as for public broadcasting in radio, television, or other forms of broadcast.
3. The rights to store the article and its content in machine-readable or electronic form on all media (such as computer disks, compact disks, magnetic tape), to store the article and its contents in online databases belonging to the publisher or third parties for viewing or downloading by third parties, and to present or reproduce the article or its contents on visual display screens, monitors, and similar devices, either directly or via data transmission.
4. The rights to reproduce and distribute the article and its contents by all other means, including photomechanical and similar processes (such as photocopying or facsimile), and as part of so-called document delivery services.
5. The right to transfer any or all rights mentioned in this agreement, as well as rights retained by the relevant copyright clearing centers, including royalty rights to third parties.

Online Rights for Journal Articles: Guidelines on authors' rights to archive electronic versions of their manuscripts online are given in the "Guidelines on sharing and use of articles in Hogrefe journals" on the journal's web page at <http://www.hgf.io/sopsy>

December 2021

Hogrefe OpenMind

Open Access Publishing? It's Your Choice!

Your Road to Open Access

Authors of papers accepted for publication in any Hogrefe journal can choose to have their paper published as an open access article as part of the Hogrefe OpenMind program. This means that anyone, anywhere in the world will – without charge – be able to read, search, link, send, and use the article, in accordance with the internationally recognized Creative Commons licensing standards.

The Choice Is Yours

1. Open Access Publication:

The final “version of record” of the article is published online with full open access. It is freely available online to anyone in electronic form.

2. Traditional Publishing Model:

Your article is published in the traditional manner, available worldwide to journal subscribers online and in print and to anyone by “pay per view.”

Whichever you choose, your article will be peer-reviewed, professionally produced, and published both in print and in electronic versions of the journal. Every article will be given a DOI and registered with CrossRef.

How Does Hogrefe's Open Access Program Work?

After submission to the journal, your article will undergo exactly the same steps, no matter which publishing option you choose: peer-review, copy-editing, typesetting, data preparation, online reference linking, printing, hosting, and archiving. In the traditional publishing model, the publication process (including all the services that ensure the scientific and formal quality of your paper) is financed via subscriptions to the journal. Open access publication, by contrast, is financed by means of a one-time article fee (€ 2,500 or US \$3,000) payable by you the author, or by your research institute or funding body.

Once the article has been accepted for publication, it's your choice – open access publication or the traditional model. We call it OpenMind!

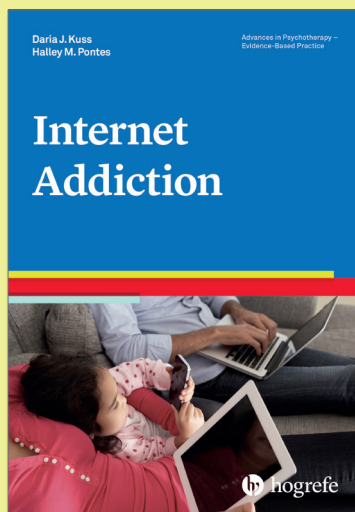
Learn more about open access and Hogrefe OpenMind: hgf.io/openmind-us

For authors from Germany – open access publication is possible under a publish-and-read agreement with 100+ institutions.

A compact guide to the assessment and treatment of Internet addiction

“This excellent book is a pleasure to read. At a time when clinicians are scrambling to learn what they can about the rapidly developing problem of Internet addiction, this book offers them an excellent place to start.”

Hilarie Cash, PhD, Chief Clinical Officer and Co-Founder of reSTART Life, PLLC, Fall City, WA – the first residential treatment program for Internet addiction in the US



Daria J. Kuss / Halley M. Pontes

Internet Addiction

Advances in Psychotherapy –
Evidence-Based Practice, vol. 41
2019, iv / 86 pp.
US \$29.80 / € 24.95
ISBN 978-0-88937-501-7
Also available as eBook

This book examines how you can identify, assess, and treat Internet addiction in the most effective manner. Internet use has become an integral part of our daily lives, but at what point does it become problematic? What are the different kinds of Internet addiction? And how can professionals best help clients? This compact, evidence-based guide written by leading experts from the field helps disentangle the debates and controversies around Internet addiction, including social media addiction and Internet gaming disorder,

and outlines the current assessment and treatment methods. The book presents a 12–15 session treatment plan for Internet and gaming addiction using the method and setting with the best evidence: group CBT. Printable tools in the appendix help clinicians implement therapy. This accessible book is essential reading for clinical psychologists, psychiatrists, psychotherapists, counselors, social workers, teachers, researchers, as well as students and parents.



17th European Congress of Psychology

Psychology as the Hub Science: Opportunities & Responsibility

5-8 July 2022, Ljubljana, Slovenia

Important Dates

08 MAR 2021	Submission of abstracts & registration opens
10 NOV 2021	Abstract Submission closes
10 FEB 2022	Acceptance notification
10 MAR 2022	Early Registration & Active participants deadline

Venue & Congress secretariat Information

**CANKARJEV DOM,
Cultural and Congress
Centre**

Gregor Rogac
Prešernova 10,
1000 Ljubljana,
Slovenia
+386 1 2417145
gregor.rogac@cd-cc.si

**Scientific
Programme
Info**

Prof. Dr.
Mojca Juriševič
ecp2022@dps.si

MAIN DOMAINS /TRACKS

General Psychology
Psychology Of Development Across Lifespan
Psychology In Education
Psychology And Neuroscience
Clinical Psychology And Psychopathology
Psychology In Public Health
Social Psychology

Work And Organizational Psychology
Psychology And Data Science
Cyberpsychology
Psychology Of Art
Psychology And Research Methodology
Psychological Assessment
Psychology And Personality

Biological Psychology
Environmental Psychology
Sport Psychology
Traffic And Transportation Psychology
Military And Space Psychology
Links To Psychology /Interdisciplinary



Stay updated, subscribe:
www.ecp2022.eu