

Tipo de documento: Preprint

A Cross-Cultural Study of Everyday Moral Experiences

Autoría ditelliana: Navajas, Joaquín

Otras autorías: Murray, Samuel; Jiménez-Leal; William; Amaya, Santiago; et-al.

¿Cómo citar este documento?

Murray, S., Jimenez-Leal, W., Amaya, S., Silverstein, P., Kekecs, Z., Chartier, C. R., ... Schmidt, K. (2024, January 26). A Cross-Cultural Study of Everyday Moral Experiences.

<https://doi.org/10.31234/osf.io/jktnv>

El presente documento se encuentra alojado en el Repositorio Digital de la Universidad Torcuato Di Tella bajo una licencia Atribución 4.0 Internacional CC BY 4.0 DEED.

<https://repositorio.utdt.edu/handle/20.500.13098/12660>

A Cross-Cultural Study of Everyday Moral Experiences

Samuel Murray, Providence College

William Jiménez-Leal, Universidad de los Andes

Santiago Amaya, Universidad de los Andes

Priya Silverstein, Ashland University, Institute for Globally Distributed Open Research and
Education

Zoltan Kekecs, Eötvös Loránd University, Institute of Psychology

Christopher R. Chartier, Ashland University

Tia Blackall, Ashland University

Cristina E. Salvador, Duke University

Gerit Pfuhl, UiT The Arctic University of Norway

Maria Montefinese, University of Padova

Halil E. Kocalar, Muğla Sıtkı Koçman University

Natalia L. Tome, University of Brasilia, Social and Affective Neuroscience Institute

Anna Polemikou, University of the Aegean

William E. Davis, Wittenberg University

Mariola Paruzel-Czachura, University of Silesia in Katowice, University of Pennsylvania

Monira I. Almohizea, King Saud University

Seda Merve Şahin, Middle East Technical University

Guilherme W Wendt, Western Paraná State University

Bastien Trémolière, Université Toulouse Jean Jaurès

Wang Zheng, Guangxi Normal University

Cleno Couto, University of Brasilia, Social and Affective Neuroscience Institute

Asil Özdoğru, Marmara University

Maja Becker, Université de Toulouse

Yoshihiko Kunisato, Senshu University

David C. Vaidis, Université de Toulouse

Oscar J. Galindo-Caballero, Universidad Manuela Beltrán

Robert M. Ross, Macquarie University

Ziqian Xia, Tongji University

Minjun Zhang, Kyushu University

Gavin B. Sullivan, International Psychoanalytic University Berlin

Thomas R Evans, University of Greenwich

Krystian Barzykowski, Jagiellonian University

Ali H. Al-Hoorie, Royal Commission for Jubail and Yanbu

Daniel Nettle, Institut Jean Nicod

Hendrik Godbersen, FOM University of Applied Sciences

Xiaoming Jiang, Shanghai International Studies University

Yuki Yamada, Kyushu University

Bradley Walker, University of Western Australia

Sergio C. Torres, Leibniz Institut für Wissensmedien

Marta Rokosz, University of Wrocław

Handan Akkas, Ankara Science University

Harry Manley, HELP University Subang 2

Huanxu Liu, Kyushu University

Mahmoud M. Elsherif, University of Leicester

Hongfei Du, Beijing Normal University

Arathy Puthillam, University of California San Diego

Lindsay Hahn, University at Buffalo

Samantha K. Stanley, The Australian National University

Carlota Batres, Franklin and Marshall College

Agnieszka Sorokowska, University of Wrocław

Danielle L. Wagstaff, Federation University

Steve M. J. Janssen, University of Nottingham Malaysia

Candela I. Jantus, Universidad Torcuato Di Tella

Ahmed Khaoudi, Moulay Ismail University

Jackson G. Lu, Massachusetts Institute of Technology

Yang Wu, Huazhong University of Science and Technology

Marta Kowal, University of Wrocław

Andrés Camargo, Fundación Universitaria del Área Andina

Lusine Grigoryan, University of York

Joaquin Navajas, Universidad Torcuato Di Tella, National Scientific and Technical Research

Council (CONICET)

Kevin Vezirian, University of Western Australia

Petros Roussos, National and Kapodistrian University of Athens

Michal Misiak, University of Wroclaw, University of Oxford

Anna Studzinska, Icam School of Engineering

Susana Ruiz-Fernandez, Brandenburg University of Technology

Argiro Vatakis, Panteion University of Social and Political Sciences

Anna Oleszkiewicz, University of Wroclaw, Technische Universität Dresden

Joonha Park, Nagoya University of Commerce and Business

Ruth Wambui Gichanga, St. Paul's University

Erin M. Buchanan, Harrisburg University of Science and Technology

Sumara Masood Ul Hassan, National University of Sciences and Technology

Magdalena Kękuś, SWPS University

Vaitsa Giannouli, Aristotle University of Thessaloniki

Lamya Mouhssine, Moulay Ismail University

Michal Pieniak, University of Wroclaw

Michele Anne, Taylor's University

Fatima Zahra Sahli, Ibn Tofaïl University

Aurélien Graton, Université Paris Cité

Brian Otieno, Art and Flickers Africa, AMREF international University

Tim Kurz, University of Western Australia

Kyoshiro Sasaki, Kansai University

Paulo R. S. Ferreira, Universidade Federal da Grande Dourados

S. Craig Roberts, University of Wroclaw, University of Stirling

Benedict Ntinyari Gituma, St Paul's University

Wakil Ajibola Asekun, University of Lagos

Barnaby J. W. Dixon, The University of the Sunshine Coast, The University of Queensland

Kathleen Schmidt, Ashland University

Author Note

The author list and order as presented here is preliminary, incomplete, and not necessarily reflective of final contribution levels on the project; the order of authors 11-86 is randomized.

Correspondence concerning this article should be addressed to Samuel Murray,

smurray7@providence.edu. *Last updated:* 12 March 2024.

Abstract

Do people in different societies experience morality differently in everyday life? Using experience sampling methods, we investigate everyday moral experiences in a sample from 20 countries across 6 continents, thereby replicating and extending a large-scale study originally conducted in the United States and Canada. We aim to replicate key findings about the kinds of moral experiences people have, the relationship between such experiences and religious or political affiliation, how such experiences affect momentary feelings, and how such experiences influence moral behavior. An international sample enables us to assess the generalizability of these findings across different socio-cultural contexts.

Keywords: morality, cultural variation, Moral Foundations Theory, experience sampling, happiness, purpose in life, moral licensing

Morality is frequently studied experimentally, either in the laboratory or via online surveys, but how do people experience it in everyday life? Hofmann et al. (2014) provided insight into everyday morality with their large-scale experience sampling study conducted in the United States and Canada. They collected nearly 4,000 moral experience reports from 1,250 participants by surveying them 5 times daily over 3 days. Participants were asked to report whether they had committed, witnessed, learned about, or been the target of a moral or immoral event in the past hour, and if so, to briefly describe the event. They also reported their momentary feelings of happiness, sense of purpose, and different emotional experiences over the preceding hour.

Hofmann et al. found that moral or immoral experiences were reported on 28.9% of responses. Moral and immoral events were reported with similar frequency. Participants who identified as religious reported stronger self-conscious emotions—such as pride and guilt—after committing moral and immoral deeds, respectively, compared to non-religious individuals. Moral experiences were associated with greater momentary feelings of happiness and purpose compared to immoral experiences. Evidence for the dynamic effects of moral experiences was also identified. For example, committing a moral act decreased the likelihood of subsequently committing another moral act (moral licensing), while being the target of a moral act increased the likelihood of subsequently committing a moral act (moral contagion). These results were a naturalistic complement to predominantly lab-based studies of morality (Efron et al., 2009; Epley & Dunning, 2000; Phillips et al., 2011; Quiles & Bybee, 1997).

Descriptions of moral events were categorized by experts according to a taxonomy of moral dimensions based on Moral Foundations Theory (MFT; Graham et al., 2013): care/harm, fairness/unfairness, loyalty/disloyalty, authority/subversion, purity/degradation,

honesty/dishonesty, liberty/oppression, and self-discipline/lack of self-discipline. Participants reported significantly more individual concerns related to care, fairness, and honesty compared to more collective concerns related to loyalty, authority, and purity. Liberals reported more experiences of fairness and unfairness than conservatives, while conservatives reported more experiences of purity and degradation than liberals, which supported predictions of MFT that moral thinking is politically divided (Graham et al., 2009; Haidt, 2012).

While informative about everyday moral experiences, these findings have not been replicated nor tested for cross-cultural generalizability. Some evidence suggests that moral experiences vary substantially by culture. The relative importance of moral dimensions varies across cultures (Atari et al., 2023; Bago et al., 2022), and different cultures make different assumptions about the values central to morality (Buchtel et al., 2015; Murray et al., 2023); both of these cultural features might make different kinds of everyday moral events more or less salient. Therefore, we will replicate and extend Hofmann et al.'s study with an international sample to explore further cultural variation in moral experiences.

We will test whether moderators of moral cognition identified in prior research also have effects on everyday moral experiences. The flexibility of interpersonal networks (relational mobility; Thomson et al., 2018) and basic values, or the values that are acknowledged across cultures (Schwartz et al., 2012), have been identified as culturally-bounded moderators of moral judgment (Awad et al., 2020; Zapko-Willmes et al., 2021). Disgust sensitivity (i.e., the extent to which common triggers elicit feelings of disgust; Rozin et al., 2009) positively predicts judgments of purity-based moral violations, but not violations in other moral domains (Wagemans et al., 2018). Moral identity internalization reflects how central being morally good is to one's identity (Aquino & Reed, 2002) and has been identified as a moderator of moral

licensing effects (Clot et al., 2014; Effron et al., 2009; Meijers, 2014). Finally, the second version of the Moral Foundations Questionnaire (MFQ-2) measures the degree to which people endorse different foundations as important to moral reasoning. A recent large-scale cross-cultural study identified differences in the relative importance of the foundations across cultures, which might be associated with different kinds of moral experiences (Atari et al., 2023).

Aims and Hypotheses

This research aims to replicate 12 findings reported in Hofmann et al. (2014) and test their cross-cultural generalizability. We have summarized evidence for the claims reported by Hofmann et al. in our Supplementary Materials 2, including evidence from additional studies. We will also test for various moderators of these effects and examine the robustness of effects across different sample characteristics (described in *Secondary analyses*, below). The findings can be grouped into four categories: (1) *The content of moral experience* (Hypotheses 1 - 4), (2) *Political and religious variability in moral experience* (Hypotheses 5 - 7), (3) *Static effects of moral experience* (Hypotheses 8 - 10), and (4) *Dynamic effects of moral experience* (Hypotheses 11 & 12).

The Content of Moral Experiences

1) For 28.9% of the responses, participants reported a moralized event.

We will test whether the proportion of moral to non-moral reports is statistically equivalent to the original proportion reported by Hofmann et al.

2) Moral and immoral events occurred at similar rates, with moral events being slightly more common at 15.3% compared to 13.6% for immoral events.

We will test whether the proportion of moral to immoral events is statistically equivalent to the proportion reported by Hofmann et al.

- 3) Participants were more likely to report committing or being the target of a moral versus an immoral act, and were more likely to learn about an immoral rather than a moral act.**

We will test whether participants are: (a) more likely to report committing a moral act compared to committing an immoral act, (b) more likely to be the target of a moral act compared to an immoral act, and; (c) more likely to learn about an immoral act rather than about a moral act.

- 4) The majority of morally relevant experiences (50.6%) fell under the Care/Harm Foundation (per Moral Foundations Theory).**

We will test whether moral experiences are categorized into different moral domains of Moral Foundations Theory in proportions that are statistically equivalent to the proportions reported by Hofmann et al.

Political and Religious Variability in Moral Experience

- 5) Political ideology was significantly associated with the content of morally relevant experiences.**

We will test whether, after controlling for religiosity: (a) being liberal is associated with a greater likelihood of reporting a fairness experience compared to being conservative, and; (b) being conservative is associated with a greater likelihood of reporting a purity experience compared to being liberal.

- 6) Religious participants reported fewer immoral experiences overall.**

We will test whether identifying as religious is associated with a lower likelihood of reporting an immoral experience compared to someone who does not identify as religious.

- 7) Religious participants reported more intense self-conscious emotions (guilt, embarrassment, disgust, pride, and gratefulness) compared to non-religious participants in response to committing immoral and moral acts.**

We will test whether people who identify as religious report feeling more inspired or proud after reporting a committed moral act *and* report feeling more angry or guilty after reporting a committed immoral act relative to people who do not identify as religious.

Static Effects of Moral Experience

- 8) Moral acts were associated with higher levels of momentary happiness compared to immoral acts.**

We will test whether experiences of moral events are associated with higher momentary happiness compared to experiences of immoral events.

- 9) The effect of moral experience on happiness was moderated by perspective, where the strongest gain and loss in happiness were observed when participants were the targets of moral and immoral acts, respectively.**

We will test whether being the target of a moral or immoral act has the largest effect on momentary happiness, while learning about a moral or immoral act has the smallest effect on momentary happiness.

- 10) The commission of moral and immoral acts had the strongest effect on sense of purpose compared to all other perspectives.**

We will test whether committing a moral act has the largest positive effect, and committing an immoral act the largest negative effect, on one's momentary sense of purpose relative to other perspectives.

Dynamic Effects of Moral Experience

11) Committing a moral act earlier in the day was associated with a below average likelihood of committing a moral act later in the day.

We will test whether committing a moral act earlier in the day predicts a lower likelihood of committing a moral act later in the day relative to the probability of committing a moral act subsequent to other experience types.

12) Being the target of a moral act was associated with an above average likelihood of committing a moral act later in the day.

We will test whether being the target of a moral act earlier in the day predicts a greater likelihood of committing a moral act later in the day relative to the probability of committing a moral act subsequent to other experience types.

Additionally, Hofmann et al. (2014) found evidence that committing a moral act earlier in the day was associated with an above average likelihood of committing an immoral act later in the day. Given our limited resources, we do not have sufficient power to detect this effect with adequate probability (see *Sample size justification*, below, and Supplementary Materials 3). We will explore this effect but will not draw statistical inferences because of the low power.

Almost none of these effects have been directly replicated (but see Waytz & Hofmann, 2020 for evidence about the frequency of reporting moral vs. immoral experiences). Our study's design enables direct replication of multiple hypotheses while also examining cross-cultural generalizability.

Method

Sample Size and Justification

Diverse samples will be recruited from collaborating laboratories through the Psychological Science Accelerator (PSA) network. Participants will be university students and members of the local community. When laboratory collection efforts cannot achieve the minimum sample size needed for adequate power, polling agencies will be used to supplement data collection (see *Data collection procedure*, below). We plan to collect 250 participants from each of 20 countries ($N = 5000$; see Table 1). Sample countries are subject to change because of collaborator attrition or logistical challenges.

In selecting countries, we prioritized regional representation. We used a measure of cultural distance (Muthukrishna et al., 2020) to quantify the underlying cultural variability of selected countries. Where possible, we aimed to balance countries that were culturally similar to and distant from the United States.

Table 1

List of Target Countries and Regions

Country	Region	Cultural distance index
<i>Argentina</i>	South America	.058
<i>Australia</i>	Oceania	.038
<i>Brazil</i>	South America	.071
<i>China</i>	Asia	.238
<i>Colombia</i>	South America	.102
<i>France</i>	Europe	.079
<i>Greece</i>	Europe	.078
<i>India</i>	Asia	.093
<i>Indonesia</i>	Asia	.178
<i>Japan</i>	Asia	.115
<i>Kenya</i>	Africa	.089
<i>Malaysia</i>	Asia	.125
<i>Morocco</i>	Africa	.149
<i>Nigeria</i>	Africa	.121
<i>Pakistan</i>	Middle East	.178
<i>Poland</i>	Europe	.076
<i>Saudi Arabia</i>	Middle East	N/A

<i>Turkey</i>	Asia/Europe	.12
<i>United Kingdom</i>	Europe	.08
<i>United States</i>	North America	.00

Note. Cultural distance index (CDI) was calculated from an adapted algorithm to reflect values from World Values Survey (WVS) 6 and 7. CDI could not be estimated for Saudi Arabia. The United States was the “center” country used to compute CDI.

A power analysis was conducted for each hypothesis based on analyses of original data provided by Hofmann et al. (2014; see Supplementary Materials 3 for power analysis code). Estimated power for detecting each predicted effect individually is summarized in **Table S1** (see Supplementary Materials). To account for possible inflation of original effect sizes, we subtracted the original standard error from reported fixed effects as a lower-bound estimate of the original effect (this does not apply to Claims 1, 2, or 4, which are based on descriptive statistics). Power analyses were based on 5000 simulations of models fitted on data that were simulated from original data provided by Hofmann et al., with additional random intercepts for countries assuming moderate variance across countries. Our sample size provides 95% power or greater to detect 18 of the predicted 23 effects, including claims related to moral contagion (Claim 11) and moral licensing (Claim 12).

Even though the power of our study is reasonably high to detect each hypothesized effect individually, the chance of encountering some false negatives or positives is worth noting given the number of hypothesis tests. Thus, when interpreting non-significant findings we will also discuss the direction of the observed difference or relationship and the size of the difference or relationship. Furthermore, we will discuss the number of non-supported hypotheses in the study in light of the estimated probability of observing that number of non-significant findings out of the number of hypotheses being tested given the estimated power for each.

Stopping Rules

Recruitment will stop when the sample size target is reached. We will monitor participant enrollment weekly—which does not require analyzing any data—once 80% of the sample is collected within a country ($N = 200$) and daily once 90% of the sample is collected ($N = 225$). Participating labs and polling agencies will be notified when these benchmarks are reached. When 250 participants are collected within each country, we will issue a recruitment stop for all sites within the country. Nevertheless, given the extensive data collection infrastructure, it is possible that participants will be recruited in excess of the necessary number. Data from these participants will be included in the analysis. To prevent analyzing data before collection is finalized, we will wait to analyze data until 7 days after issuing the recruitment stop for the final country.

If target sample size within a country is not achieved within 18 months of starting data collection, a recruitment stop will be issued for all sites in that country. If data collection stops before reaching the target sample size for any reason, then we will explicitly discuss why collection was stopped. Additionally, we will discuss the resulting impact on confirmatory analyses and the implication of premature stopping on our findings and their interpretation.

Data Collection Procedure

Collaborating laboratories recruited through the PSA network will conduct the first wave of data collection. Each lab, or “site”, begins data collection after receiving the appropriate ethical approval. Ethics approval is verified from each site prior to data collection. The second wave of data collection begins after all sites within each country finish data collection. Where necessary, polling agencies will recruit participants within the country to bring the sample to the

minimum size needed to adequately power statistical tests. If the sites within a country can jointly meet the minimum sample size, no polling agencies will be used in that country.

Inclusion Criteria

Eligibility for participation will be based on age (≥ 18 years), being a fluent speaker of the language of the ESM prompts, having a smartphone compatible with the RealLife Exp app, and any additional requirements that each site determines based on IRB feedback.

Exclusion Criteria

We will exclude any participants who voluntarily withdraw from the study before it is concluded and request their data be deleted. Participants will have two months after finishing the study to request their data be deleted. “Nonsense” responses will be excluded from analyses (see *Experience Classification*, below). For any participants recruited during the second wave of data collection, we will exclude anyone who does not pass quality control checks set by the polling companies used to collect supplementary responses. All other data will be included for hypothesis testing.

Missing Data

Participants who have some missing data or skip some responses will be included in analyses that do not require these data. Missing data will be checked using the *mice* package in R to confirm that missingness is random (Schafer & Graham, 2002). Findings about the pattern of missingness will be discussed in the paper and might trigger additional exploratory analyses, but they will not influence the primary confirmatory analyses.

Mobile Application

We will adopt LifeData (lifedatcorp.com) as a platform for data collection. Participants download the *RealLife Exp* application managed by LifeData to their mobile phones. Within the

app, participants download the study, which can then be completed off-line. The app sends notifications to participants during set intervals (see “Momentary assessment”, below).

Participants will receive real-time feedback about compliance through the app. LifeData securely stores information about the time, date, signal identifier, participant ID, and overall progress in the task.

Intake Survey

After enrolling, participants first complete a demographic survey.

Demographic Items

Age: “How old are you?” (*Numeric response*)

Gender: “What best describes your gender identity?” (*Man, Woman, Nonbinary, Prefer not to answer*)

Sample status: “Are you currently enrolled as a full-time student at a College or University?” (*Yes/No*)

Country of residence: “What is your current country or region of residence?” (*Multiple choices listed, full list: <https://www.britannica.com/topic/list-of-countries-1993160>*)

Country of birth: “What is your country or region of birth?” (*Same list as above*)

Political ideology (I): “How would you describe your overall political preferences? If you are middle of the road or neither label applies to you, please select 4” (*1 = Very liberal, 7 = Very conservative*)

Political ideology (II): “Where would you place yourself on the following scale of political orientation? If you are neutral or neither label applies to you, please select 4” (*1 = Very left-wing, 7 = Very right-wing*)

Education: “What is the highest level of education you have attained?” (1 = *Less than a high school degree (or equivalent)*, 2 = *High school degree (or equivalent)*, 3 = *Bachelor’s degree (or equivalent)*, 4 = *Postgraduate degree (MA, PhD, JD, etc.)*).

Socio-economic status: Below is an income scale on which 1 indicates the lowest income group and 10 the highest income group in your country. We would like to know in what group your household is. Please, specify the appropriate number, counting all wages, salaries, pensions and other incomes that come in (1 = *lowest group*, 10 = *highest group*, no midpoint).

Social media usage: “How often do you check your social media accounts in any given week?” (1 = *Once a week or less*; 2 = *2 - 3 times per week*; 3 = *Daily*; 4 = *A few times per day*; 5 = *Hourly*).

Ethnicity: “Would most people in your country consider you to be a member of an ethnic or racial minority?” (0 = *No*, 1 = *Yes*, 2 = *Unsure*).

Ethnicity follow-up: “How would you best describe your ethnicity?” (*Open response*)

Religious affiliation: “What is your present religion, if any?” (*Buddhist, Christian, Hindu, Jewish, Muslim, None, Other (please specify)*)

Religious affiliation follow-up: “How would you describe your religious affiliation?” (*Open response*)

Religiosity: “How religious are you?” (1 = *not at all*, 7 = *very much*).

Personal afterlife: “Do you believe in a personal afterlife? If you are unsure, please select 0” (-3 = *definitely no*, +3 = *definitely yes*).

The Religiosity and Political Ideology (I) measures were taken from Hofmann et al. The Political ideology (II) measure was adapted from Feldman (2013). The use of two political ideology measures is motivated by the meta-analytic results reported by Kivikangas et al. (2021), who showed that, in a global sample, these measures failed to exhibit unidimensionality. Social media usage was taken from the Pew Research Center. Socio-economic status was adapted from WVS-7. All other items were developed internally.

Participants also complete the Basic Values survey (Schwartz et al., 2012; see Supplementary materials §4) during intake.

Momentary Assessment

After completing the intake survey, each participant begins the momentary assessment part of the survey the next day. Each participant is signaled 5 times a day for 4 days. The daily observation window (9:00 to 21:30) is divided into four blocks of 120 minutes (9:00 to 11:00; 11:45 to 13:45; 14:30 to 16:30, and; 17:15 to 19:15) and one block of 90 minutes (the final block is always 20:00 to 21:30). This is different from the sample windows used by Hofmann et al. (2014). Within *RealLife Exp*, signals cannot be sensitive across blocks. Thus, to ensure that signals cannot occur within 45 minutes of each other, the start of blocks 2 - 5 are set to begin 45 minutes after the previous block closes. Within this constraint, 2-hour blocks were needed. The last block was reduced to 90 minutes so that signals would not occur too late in the day.

Within each block, signal time is randomly selected. Each assessment remains open for 45 minutes. Participants are encouraged to respond promptly and a reminder will be sent within 15 minutes if a response is not provided. Only one response per measurement occasion is allowed. While Hofmann et al. sampled participants for three days, we will allow an additional

day for signals (4 days total) to have a better chance to observe the temporal dynamics hypothesized in Claims 11 & 12.

Signal (Daily Survey)

Each signal includes the following questions:

Experience: Within the past hour, did you have a moral or immoral experience? [I committed a moral act; I committed an immoral act; I learned about a moral act; I learned about an immoral act; I was the target of a moral act; I was the target of an immoral act; I witnessed a moral act; I witnessed an immoral act; None of the above]

[In the instructions, participants are told that if they experienced multiple events within the past hour, they should respond based on the most recent event. The order of these questions is fixed across measurement occasions]:

Description: Please describe the event using the box below.

Location: Please tell us where this occurred (Public setting, My home, the home of a family member or friend, Work/School, Other).

Note. We added a 'Work/school' item to help disambiguate 'Public setting'.

[If participants indicate that they learned about an event, they are presented with the following item instead of Location]:

Source: How did you learn about this event? (Personal communication, social media, internet news/magazine, Internet: Other, TV: news, TV: other, Radio: news, Radio: Other, Newspaper/magazine, Book, Podcast, Other).

Note. We added a 'Podcast' item to reflect differences in how people consume information relative to the original study.

Intensity: “How morally right or wrong do you consider this act?” [-3 (*totally wrong/immoral*) to +3 (*totally right/moral*)].

Happiness: “How happy do you feel at the moment?” [-3 (*very unhappy*) to +3 (*very happy*)]

Purpose: “Do you feel that your life has a clear sense of purpose at this moment?” [0 (*not at all*) to 5 (*very much*)]

Affect: Within the past hour, did you feel...?

Inspired (0 = *not at all*, 5 = *very much*)

Proud (0 = *not at all*, 5 = *very much*)

Angry (0 = *not at all*, 5 = *very much*)

Guilty (0 = *not at all*, 5 = *very much*)

Note. We also selected four emotions (two positive and two negative) from the original nine measured by Hofmann et al. so we could test a subset of their emotion findings. We changed *elevated* to **inspired**, which is taken from the short-form Positive and Negative Affect Schedule (PANAS; Watson et al., 1988), a widely-used self-report measure of affect.

As an exploratory addition, if participants indicate that they did not experience something moral or immoral, they are asked if they *thought about* something moral or immoral in the past hour. If they did, they receive the description, intensity, happiness, sense of purpose, and affect questions. If they also didn't think about something moral or immoral, they receive the happiness, sense of purpose, and affect questions.

Exit Survey

When the daily assessment phase is complete, participants complete an exit survey with the following items:

Moral definition: Using the box below, please describe in your own words what it means for something to be moral and for something to be immoral. There are no right or wrong answers, but please respond honestly!

Disgust sensitivity: 27-item scale (Rozin et al., 2009)

Relational mobility: 12-item scale (Thomson et al., 2018)

Moral Foundations Questionnaire-2: 36-item scale (Atari et al., 2023)

Moral identity internalization: 5-item scale (Aquino & Reed, 2002).

These measures can be found in Supplementary Materials.

Translation

All materials will be translated into site-appropriate languages, following the translation process of the PSA (<https://psysciacc.org/translation-process/>). For each language, a language-wide coordinator will assign at least 1 translator and 1 back translator to translate study materials. If only 2 people are available, the language-wide coordinator will serve as the translator. A translation from English to the target language will be created by the translator(s). When this process is complete, the translated material will be translated back into English by the back translator(s). The language-wide coordinator will compare both versions. If any major discrepancies are detected, the language-wide coordinator creates a corrected version of the translation. Translated materials are sent to at least one external reader and to site collaborators to receive feedback on readability and potential cultural adjustments. The language-wide coordinator will make further adjustments based on this feedback.

Experience Classification

After the ESM phase of the study, three individuals from each country sampled during the daily assessment phase will serve as experience coders. Data collection teams will assist with

recruiting these individuals. Each group of coders will undergo a brief training period with the primary research team to learn about MFT. This training will provide theoretical background on MFT, key elements of each foundation, and practice sorting sample reports into foundations to ensure that coders understand the theory sufficiently to complete their task. Each coder will blind-code all responses from their country according to the MFT taxonomy used by Hofmann (see below). This approach represents a departure from the procedure in Hofmann et al., who had one expert coder blind-coding all responses with a second coder blind-coding 50% of responses. We opted for additional coders to mitigate the prevalence of errors in MFT categorization. To assess inter-rater reliability, we will compute simple rater agreement and Gwet's *ACI* coefficient. Gwet's coefficient is a measure of inter-rater agreement designed for more than two raters and is recommended for skewed distributions (e.g., categories that are rare or prevalent; see Heyman et al., 2014; Neuendorf, 2017). If the *ACI* coefficient is $\geq .70$, this is sufficient to establish agreement (Fleiss, 1981; Gwet, 2014). In cases where raters disagree, a 2-rater majority will determine classification. In cases without a majority, the experience will be rated "Too vague".

If the *ACI* coefficient does not meet the .70 threshold, we will compute Cohen's *kappa* for each pair of raters. If one pair of raters achieves *kappa* $> .75$, then the third rater will be discarded and ratings from the pair will be used in content analyses. In cases of disagreement, the experience will be rated as "Too vague".

Following Hofmann et al. (2014), coders will have 10 options for classification: Care/Harm, Fairness/Unfairness, Loyalty/Disloyalty, Authority/Subversion, Purity/Degradation, Liberty/Oppression, Honesty/Dishonesty, Self-Discipline/Lack of Self-Discipline, None of the above, Nonsense. The latter two were included to provide coders with the opportunity to sort reports into potentially new categories not represented within the traditional taxonomy.

Additionally, this would remove the pressure to fit reports into unsuitable categories. If two or more coders rate a response as “Nonsense”, that response and all associated data for that measurement occasion will be removed.

We opted to recruit coders from the same culture as the participants who generated the reports for two reasons. First, this partly replicates the original Hofmann et al. procedure, where they used intra-cultural coders to sort experiences. Second, given the important role of context in understanding the moral dimensions of experiences (Schein, 2020), we believe that coders operating within the same cultural framework are best positioned to classify experience reports. Some reports might contain context clues or implicit meanings that could not be detected by an “outside” observer. Thus, coders within the same culture are best positioned to capture the potential cultural variability underlying the content of moral experiences.

Informed Coding Decisions and Exclusions

We will flag inconsistent responses (e.g., a moral act rated as “very wrong”; see *External validation*, below) but retain them in primary analyses. Robustness analyses will be performed with these potentially errant responses recoded. When participants label the location of a moral experience as ‘Other’, but the report makes clear that the location falls under an existing category, coders will be allowed to change participant responses (e.g., Hofmann et al. report that one participant put ‘Other’ when an experience occurred in a restaurant, which was changed to “Public setting” for location).

If participants do not provide a text entry for an experience report, it will be excluded from content analyses. Responses coded as “None of the above” will not be included in content analyses.

External Validation

Hofmann et al. recruited external raters for two purposes: (a) to assess the extent to which perceptions of moral and immoral deeds are socially shared, and; (b) to obtain an independent measure of the “intensity” of reported moral and immoral deeds. We will validate responses for each country. 150 external raters are needed for each country. Raters will see no more than 60 descriptions randomly selected from responses given during the ESM phase. PSA labs and polling agencies will recruit external raters for this part of the study.

Raters will not be given information about the purpose of the study or the people who provided the original entries. Responses will be delivered via Qualtrics survey so that responses can be presented in random order to each rater. Raters will be asked the following questions:

Valence: “How morally right or wrong do you consider this act or deed?” [-3 = *totally wrong/immoral* to 3 = *totally right/moral*; midpoint labeled *unsure*]

Permissibility: “How morally permissible do you consider this act or deed?” [-3 = *totally impermissible* to 3 = *totally permissible*; midpoint labeled *unsure*]

We will compute Pearson’s correlation coefficient to assess correspondence between external ratings of valence and permissibility. If $r > .7$, responses will be averaged to a single item. We will then compute Pearson’s correlation coefficient to assess correspondence between self-reported intensity and external ratings of intensity.

Data Analysis Plan

For hierarchical models, random intercepts for the participant and country will be included to account for within-person and between-country variation. The same default level for categorical predictors will be used in each test (political ideology = “conservative”, religiosity = “non-religious”, and perspective = “witnessed”). Statistical significance will be determined according to the p-value of the Wald-Z test provided in the `glmer()` model summary in the *lme4*

package in R (Bates et al., 2015), and Wald confidence intervals will be computed for the intercept and the fixed effects from the Standard Error estimates provided in the `glmer()` model summary (Luke, 2017).

Primary Analyses

See Supplementary Materials for protocols on creating variables related to basic values, disgust sensitivity, moral foundations, relational mobility, and moral identity internalization.

Claim 1. We will test whether the proportion of moral to non-moral reports is statistically equivalent to the original proportion reported by Hofmann et al. Based on our analyses of the original dataset, the predicted probability of a moral report compared to a non-moral report was 27.8% (odds ratio = 2.59, 95% CI[2.44, 2.76]). We will fit a generalized linear model with report type (1 = moral, 0 = nonmoral) as a binomial outcome. The model will contain only intercepts (participant and country will be included in the model as random effects). We will conduct an equivalence test using the two one-tailed analysis approach. If the confidence bounds of the estimated probability of moral (as opposed to nonmoral) reports are within 5 percentage points of the original estimate based on our analysis of the original dataset, then we will interpret this as the equivalence hypothesis being supported.

Claim 2. We will test whether the proportion of moral to immoral events is statistically equivalent to the proportion reported by Hofmann et al. Based on our analyses of the original dataset, the predicted probability of reporting an immoral event compared to a moral event was 47.4% (odds ratio = 0.90, 95% CI[0.83, 0.98]). We will fit a generalized linear model with report valence (0 = moral, 1 = immoral) as a binomial outcome. The model will contain only intercepts (participant and country will be included in the model as random effects). We will conduct an equivalence test using the two one-tailed analysis approach. If the confidence bounds of the

estimated probability of moral (as opposed to immoral) reports will be within 5 percentage points of the original estimate reported by Hofmann et al., then we will interpret this as the equivalence hypothesis being supported.

Claim 3. We will test whether participants are more likely to report committing a moral act compared to committing an immoral act (3a), that participants will be more likely to be the target of a moral act compared to an immoral act (3b), and that participants will be more likely to learn about an immoral act rather than about a moral act (3c). We will fit a generalized linear mixed effect regression predicting response valence from commit, target, and learned about perspectives (dummy-coded) with random intercepts of participant and country. Since due to the coding of valence, *moral* is the default level response, significant negative effects of “commit” and “target” perspectives on valence (i.e., more likely to report *moral*) and positive effects of “learn about” perspective on valence (i.e., more likely to report *immoral*) will be interpreted as supporting the hypothesis. Claims 3a, 3b, and 3c will be tested separately.

Claim 4. We will test whether the proportion of moral experiences within each category of Moral Foundations Theory will be statistically equivalent to the proportions reported by Hofmann et al. We will create eight dummy-coded variables corresponding to the eight separate categories used by Hofmann et al. We will fit eight separate hierarchical generalized linear models that contain only intercepts (participant and country will be coded as random effects). We will conduct an equivalence test using the two one-tailed analysis approach. We will assess whether the confidence bounds of the probability of the event of interest (the specific foundation category being observed vs. not) falls within some percentage points around the estimated probability of the same event of interest based on the data from Hofmann et al. Table 2 shows the

probability and odds ratios of each foundation category based on our own analyses of the data shared by Hofmann et al. (see *Table 2* below):

Table 2

Probabilities reported in Hofmann et al. (2014) and smallest effect size of interest (SESOI; see Lakens, 2017) ranges for Claim 4 (note: the probability refers to the probability that a report falls in the category rather than any other category)

Foundation	Original Probability	SESOI Range (log odds scale)	SESOI Range (probability)
<i>Care</i>	50.6%	[-0.20, 0.20]	44.9% - 54.9%
<i>Fairness</i>	13.9%	[-2.5, -1.54]	7.6% - 17.6%
<i>Loyalty</i>	4.8%	-2.74*	6.1%
<i>Authority</i>	5.6%	-2.51*	7.5%
<i>Purity</i>	5.2%	-2.93*	5.1%
<i>Liberty</i>	3.3%	-2.93*	5.0%
<i>Honesty</i>	12.8%	[-2.80, -1.68]	5.7% - 15.7%
<i>Discipline</i>	3.8%	-2.77*	5.9%

Note. * = Denotes upper-bound for smallest effect of interest because frequency is predicted to be low.

For the Care, Fairness, and Honesty dimensions, we will test against both the lower and upper bounds of the 5 percentage point difference smallest effect of interest range. For the other categories, we will test against the upper bound of +5 percentage points difference as our smallest effect of interest range. We will conduct one-sided tests for these other categories because we predict that they will occur with relatively low frequency. We are including two

additional categories in our coding scheme (None of the above and Nonsense), but we predict that only a small number of reports will fall in these categories. Thus, we are not including these categories in our hypothesis tests. The hypotheses related to the eight foundation categories will be tested separately. Significant equivalence will be interpreted as supporting the hypothesis.

Claim 5. We will test whether, after controlling for religiosity, being liberal is associated with a greater likelihood of reporting an experience of Fairness compared to being conservative (5a), while being conservative will raise the likelihood of reporting an experience of Purity compared to being liberal (5b). We will create dummy-coded variables corresponding to the Moral Foundations categories fairness (vs. any other foundation category) and purity (vs. any other foundation category). We will fit two hierarchical generalized linear models to predict the likelihood of reporting an experience of fairness or purity. Political ideology (I) and religiosity will be categorical predictors. Each will be coded based on Hofmann et al.'s original coding scheme (0 = liberal, 1 = conservative; -1 = non-religious, 1 = religious). Participants who respond above the midpoint on the political ideology (I) question will be coded as conservative, while those who respond below the midpoint will be coded as liberal (participants who respond at the midpoint are excluded from this analysis). Likewise, participants who respond above the midpoint on the religiosity question will be coded as religious, while those who respond below the midpoint will be coded as non-religious (midpoint responses will also be excluded). Each model will include random intercepts for participants and countries. If there is evidence of overdispersion (e.g., zero inflation), we will fit a negative binomial model that corrects for this. Hypotheses 5a and b will be tested separately. A main effect of political ideology will be interpreted as supporting the hypothesis.

Claim 6. We will test whether religious individuals are less likely to report an immoral experience compared to non-religious individuals. We will fit a generalized linear mixed effect model to predict report valence (0 = moral, 1 = immoral) from religiosity (contrast-coded, -1 = non-religious, 1 = religious) with random intercepts for participants and countries. An effect of religiosity will be interpreted as supporting the hypothesis.

Claim 7. We will test whether religious people report feeling more inspired or proud after reporting a committed moral act (7a) and report feeling more angry or guilty after reporting a committed immoral act (7b) compared to non-religious people. We will analyze the subset of data where people reported committing either moral or immoral acts, respectively. If the Pearson correlation of these items exceeds .70, we will average scores for use in analyses. Otherwise, responses will be analyzed separately. We will fit separate hierarchical linear models predicting affect from religiosity (contrast-coded as described above). We will include random intercepts for participants and countries. Claims 7a and b will be tested separately. A main effect of religiosity will be interpreted as supporting the hypothesis.

Claim 8. We will test whether people experience greater happiness after reporting moral experiences compared to reporting immoral experiences. We will fit a hierarchical linear model to predict momentary happiness from report valence (moral vs. immoral), perspective, and their interaction (see 9 for why perspective is involved in this model). The model will include random intercepts for participants and countries. Since *moral* is the default level in the analysis, a significant negative main effect of valence will be interpreted as supporting the hypothesis.

Claim 9. We will test whether being the target of a moral or immoral act will have the largest effect on momentary happiness (9a), while learning about a moral or immoral act will have the smallest effect on momentary happiness (9b). We will use the same model as Claim 8.

We will extract effect sizes using the *emmeans* package in R. An interaction between perspective and valence will be interpreted as evidence of moderation. If the moral vs. immoral difference has the largest Cohen's *d* effect size within the "target of" perspective, we will conclude that event valence had the strongest effect when participants were targets of moralized acts, supporting Claim 9a. If the moral vs. immoral difference has the smallest Cohen's *d* effect size within the "learned about" perspective, we will conclude that event valence had the weakest effect when participants learned about a moralized act supporting Claim 9b.

Claim 10. We will test whether committing a moral or immoral act has the largest effect on momentary sense of purpose relative to other perspectives. We will follow an analysis plan similar to testing Claim 9, with the exception that the outcome variable for the hierarchical linear regression will be sense of purpose. An interaction between valence and perspective will be taken as supporting the moderation. If committing a moral act is associated with the highest average sense of purpose, this will be taken as supporting the hypothesis.

Claims 11 - 12 and exploratory analysis. We will test whether committing a moral act earlier in the day predicts a lower likelihood of committing a moral act later in the day (Claim 11). We predict that being the target of a moral act predicts a greater likelihood of committing a moral act later in the day (Claim 12) relative to a baseline probability of committing a moral act subsequent to nonmoral experiences. Additionally, we will explore whether committing a moral act earlier in the day predicts a greater likelihood of committing an immoral act later in the day (Claim 12) relative to a baseline probability of committing an immoral act subsequent to nonmoral experiences.

Following the analysis plan of Hofmann et al., we will create two new variables called "precedes_committed_moral_act" and "precedes_committed_immoral_act". For each signal for

which a response was recorded, the variable “precedes_committed_moral_act” indicates whether at least one moral act was committed subsequent to that given signal response on the same day (1) or not (0). The variable “precedes_committed_immoral_act” works the same way but for committing immoral acts (we assume, with Hofmann et al., that *days* are the fundamental units within which people trade-off moral behavior, therefore signal 5 cannot serve as a preceding signal). Non-moral events will form the reference category for the perspective variable, so that each level of the factor tests the difference in predicted likelihoods from the sample average likelihood of occurrence.

To test Claims 11 and 12 and conduct our exploratory analysis, we will fit a hierarchical generalized linear model to predict the likelihood of committing a moral act (11 and 12) or immoral act (exploratory) subsequent to committing a moral act (11 and exploratory) or being the target of a moral act (12) from perspective with random intercepts for participant and country. For Claim 11 we will consider a significant negative effect of the “committing a moral act” perspective as supporting the hypothesis. For the exploratory analysis, we will consider a significant positive effect of the “committing an immoral act” perspective as supporting the hypothesis. For Claim 12 we will consider a significant positive effect of the “target of a moral act” perspective as supporting the hypothesis.

Secondary Analyses

We plan to collect data on 17 demographic variables. Some of these are necessary for conducting our primary analyses (political ideology and religiosity). Others will be part of secondary exploratory analyses to assess how effects are moderated across different sample characteristics (see *Intake Survey*, above). Additionally, we will examine the impact of historical pathogen prevalence (Murray & Schaller, 2010), gross domestic product, and cultural tightness

(Gelfand et al., 2011) for each country sampled where data is available. We will also examine the moderating effect of individual difference variables where there is evidence of an association with moral judgment or experience (see *Individual difference measures*, above).

For each model, we will follow a maximal-to-minimal modeling process (Barr et al., 2013) so that if a model fails to converge, we will eliminate the random intercepts closer to zero (Barr et al., 2013; Brauer & Curtin, 2018; Meteyard & Davies, 2020).

Besides the main results, we will explore whether there is any indication that different sample characteristics (student/community *or* compensated/non-compensated) moderate effects. Where appropriate, we will discuss these potential effects in the discussion section. We will explore how the exclusion of participants for low response rates alters our findings and discuss these changes, if needed.

Specific Exploratory Analyses

Claim 1. We will alter the model structure slightly from the primary analysis. We will fit a hierarchical generalized linear model to predict changes in the likelihood of experiencing a moral (compared to non-moral) event from different moderators, particularly disgust sensitivity.

Claim 3. We will explore the extent to which participants are more likely to witness a moral event compared to an immoral event. We will also explore the extent to which social media use moderates the effect of different perspectives on moral experience. Because people use social media to express moral outrage and social media usage is more widespread than 10 years ago, we will assess whether people are more likely to witness or learn about immoral events compared to moral events.

We will explore whether participants from non-Western countries are more likely to report experiences of moral compared to immoral events. To do this, we will include country as a fixed effect in the model and include a term for the interaction between country and perspective.

Claim 4. We will explore the extent to which different demographic and individual difference variables moderate the likelihood of experiencing an event with content in a particular moral foundation. There is evidence that several factors are associated with stronger endorsement of different norm categories within MFT as measured by the MFQ (Atari et al., 2023). Greater historical pathogen prevalence (van Leeuwen et al., 2012), lower relational mobility (Graham et al., 2016), and tighter cultural norms (Gelfand et al., 2011) are associated with stronger endorsements of Binding norms (Authority, Purity, and Loyalty). As there is evidence that GDP is significantly correlated with endorsements of different norms within MFT (van Leeuwen et al., 2012), we will control for GDP in all regressions.

Claim 5. We will explore the extent to which effects of political ideology on moral experience change depending on the measure of political ideology. Primary analyses use the first measure (“How would you describe your overall political preferences?”). As reported in Kivikangas et al. (2021), outside the US, effects of political ideology on moral judgment changed when people were asked to place themselves on a left-wing/right-wing spectrum. We will refit each model testing for effects of political ideology with the second measure (“Where would you place yourself on the following scale of political orientation?”). We will also explore how the importance of different moral foundations (as measured by the MFQ-2) predict the content of moral experience. We will fit separate generalized linear mixed models to predict the likelihood of reporting an experience with care-, loyalty-, authority-, or purity-related content from the self-reported importance of these categories in moral judgment, respectively. Outcome variables

will be dummy-coded. Additionally, we will explore how different fairness-related constructs (equality and proportionality) predict the likelihood of reporting an experience with fairness-related content to see whether people are more likely to experience issues of equality or proportionality. Each model will include random intercepts for participants and countries.

Claim 6. We will explore the extent to which being a member of a particular religious denomination modulates the likelihood of having a positive or negative moral experience. We will explore whether disgust sensitivity and social media use moderate the effect of religiosity on the likelihood of having an immoral experience. A recent meta-analysis (Yu et al., 2022) found a significant positive association between religiosity and disgust sensitivity. Thus, we will explore whether the effect of religiosity on the likelihood of having an immoral experience is mediated by disgust sensitivity. Furthermore, given the use of social media to express moral outrage (Brady et al., 2021), we will examine whether increased social media use among religious individuals increases the likelihood of learning about immoral acts.

Claim 7. We will explore whether being a member of a particular religious denomination moderates the effect of religiosity on the effect following moral experiences. Additionally, we will explore how results change when religiosity is treated as a continuous predictor and interacts with religious affiliation. We will also include an interaction with belief in a personal afterlife to see whether this interacts with denomination and religiosity.

Claims 8 and 9. We will explore whether identifying with a particular religious denomination or believing in a personal afterlife modulates the effect of moral experiences on happiness.

Claim 10. A recent study found that conservatives tend to report greater meaning-in-life compared to liberals (Newman et al., 2019). Hence, we will explore how model coefficients change when controlling for ideology.

Claims 11 and 12. Many of the proposed factors that explain the effects of moral licensing cannot be studied within an ESM design. For example, we cannot experimentally manipulate ethical mindset or ease of rationalizing target behavior (see Mullen & Monin, 2016). However, we will explore the degree to which the internalization of moral identity modulates the likelihood of licensing, since the importance of moral identity is a proposed moderator of moral licensing (Clot et al., 2014; Effron et al., 2009; Meijers, 2014). Monin and Miller (2001), building on Moral Accounting Theory (Thaler, 1985), suggested that behavioral licensing is likely to increase the possibility of behavior within the same domain. We will explore whether the domain of the initial experience raises the likelihood of licensed behavior within the same domain. We will also explore alternative coding schemes to measure the temporal dynamics of licensing and contagion effects. Hofmann et al. assume that days are the units within which people trade-off moral behavior. We will explore the effects of two different modeling strategies: (1) extending the range of subsequent signals to the first signal of the next day (thus, e.g., signal 1 of day 1 could impact signal 1 of day 2), and (2) extending the range of subsequent signals to cover two days (thus, e.g., signal 1 of day 1 could impact signal 5 of day 2). This changes the assumption about when moral credentials “expire”.

References

- Aquino, K., & Reed II, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology*, 83(6), 1423–1440. <https://doi.org/10.1037/0022-3514.83.6.1423>
- Atari, M., Haidt, J., Graham, J., Koleva, S., Stevens, S. T., & Dehghani, M. (2023). Morality beyond the WEIRD: How the nomological network of morality varies across cultures. *Journal of Personality and Social Psychology*. Advance online publication. <https://doi.org/10.1037/pspp0000470>
- Awad, E., Dsouza, S., Shariff, A., Rahwan, I., & Bonnefon, J.-F. (2020). Universals and variations in moral decisions made in 42 countries by 70,000 participants. *PNAS Proceedings of the National Academy of Sciences of the United States of America*, 117(5), 2332–2337. <https://doi.org/10.1073/pnas.1911517117>
- Bago, B., Kovacs, M., Protzko, J., Nagy, T., Kececs, Z., Palfi, B., Adamkovic, M., Adamus, S., Albaloooshi, S., Albayrak-Aydemir, N., Alfian, I. N., Alper, S., Alvarez-Solas, S., Alves, S. G., Amaya, S., Andresen, P. K., Anjum, G., Ansari, D., Arriaga, P., Aruta, J. J. B. R., ... Aczel, B. (2022). Situational factors shape moral judgements in the trolley dilemma in Eastern, Southern and Western countries in a culturally diverse sample. *Nature Human Behaviour*, 6(6), 880–895. <https://doi.org/10.1038/s41562-022-01319-5>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). "Fitting linear mixed-effects models using lme4." *Journal of Statistical Software*, 67(1), 1–48.

<https://doi.org/10.18637/jss.v067.i01>

Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances*, 7(33),

<https://doi.org/10.1126/sciadv.abe5641>

Brauer, M., & Curtin, J. J. (2018). Linear mixed-effects models and the analysis of nonindependent data: A unified framework to analyze categorical and continuous independent variables that vary within-subjects and/or within-items. *Psychological Methods*, 23(3), 389–411. <https://doi.org/10.1037/met0000159>

Buchtel, E. E., Guan, Y., Peng, Q., Su, Y., Sang, B., Chen, S. X., & Bond, M. H. (2015). Immorality east and west: Are immoral behaviors especially harmful, or especially uncivilized? *Personality and Social Psychology Bulletin*, 41(10), 1382-1394.

<https://doi.org/10.1177/0146167215595606>

Clot, S., Grolleau, G., & Ibanez, L. (2014). Smug alert! Exploring self-licensing behavior in a cheating game. *Economics Letters*, 123(2), 191–194.

<https://doi.org/10.1016/j.econlet.2014.01.039>

Effron, D. A., Cameron, J. S., & Monin, B. (2009). Endorsing Obama licenses favoring whites. *Journal of Experimental Social Psychology*, 45(3), 590–593.

<https://doi.org/10.1016/j.jesp.2009.02.001>

Epley, N., & Dunning, D. (2000). Feeling "holier than thou": are self-serving assessments produced by errors in self-or social prediction? *Journal of Personality and Social Psychology*, 79(6), 861. <https://doi.org/10.1037/0022-3514.79.6.861>

- Feldman, S. (2013). Political ideology. In L. Huddy, D. O. Sears, & J. S. Levy (Eds.), *The Oxford handbook of political psychology* (pp. 591–626). Oxford University Press.
- Fleiss, J. L., Levin, B., and Paik, M.C. (2003). The measurement of interrater agreement. In Fleiss, J.L., Levin, B., and Paik, M.C. (Eds.), *Statistical methods for rates and proportions*, 3rd edition (pp. 598 - 626). John Wiley & Sons.
- Gelfand, M. J., Raver, J. L., Nishii, L., Leslie, L. M., Lun, J., Lim, B. C., Duan, L., Almaliaich, A., Ang, S., Arnadottir, J., Aycan, Z., Boehnke, K., Boski, P., Cabecinhas, R., Chan, D., Chhokar, J., D'Amato, A., Ferrer, M., Fischlmayr, I. C., Fischer, R. ... Yamaguchi, S. (2011). Differences between tight and loose cultures: a 33-nation study. *Science*, 332(6033), 1100–1104. <https://doi.org/10.1126/science.1197754>
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029–1046. <https://doi.org/10.1037/a0015141>
- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. In *Advances in experimental social psychology* (Vol. 47, pp. 55-130). Academic Press. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>
- Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology*, 8, 125–130. <https://doi.org/10.1016/j.copsyc.2015.09.007>
- Gwet, K. L. (2014). *Handbook of inter-rater reliability: The definitive guide to measuring the extent of agreement among raters*. Advanced Analytics, LLC.

- Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.
- Heyman, R., Lorber, M. F., Eddy, J.M., & Schellati, T. (2014). Behavioral observation and coding. In H.T. Reis, & C.M. Judd (Eds.), *Handbook of research methods in social and personality psychology* (2 ed., pp. 345–372). Cambridge University Press.
- Hofmann, W., Wisneski, D. C., Brandt, M. J., & Skitka, L. J. (2014). Morality in everyday life. *Science*, 345(6202), 1340–1343. <https://doi.org/10.1126/science.1251560>.
- Kivikangas, J. M., Fernández-Castilla, B., Järvelä, S., Ravaja, N., & Lönnqvist, J.-E. (2021). Moral foundations and political orientation: Systematic review and meta-analysis. *Psychological Bulletin*, 147(1), 55–94. <https://doi.org/10.1037/bul0000308>
- Lakens, D. (2017). Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science* 8(4), 355–362. <https://doi.org/10.1177/1948550617697177>
- Luke, S. G. (2017). Evaluating significance in linear mixed-effects models in R. *Behavior Research Methods*, 49(4), 1494–1502. <https://doi.org/10.3758/s13428-016-0809-y>
- Meijers, M. H. C. (2014). *On justifying eco-unfriendly behaviors* (Doctoral dissertation, Universiteit van Amsterdam [Host]).
- Meteyard, L., & Davies, R. A. (2020). Best practice guidance for linear mixed-effects models in psychological science. *Journal of Memory and Language*, 112, 104092. <https://doi.org/10.1016/j.jml.2020.104092>
- Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. *Journal of Personality and Social Psychology*, 81(1), 33-43. <https://doi.org/10.1037/0022-3514.81.1.33>

Mullen, E., & Monin, B. (2016). Consistency versus licensing effects of past moral behavior.

Annual Review of Psychology, 67, 363-385.

<https://doi.org/10.1146/annurev-psych-010213-115120>

Murray, S., Carmona, G., Vega, L., Jiménez-Leal, W., & Amaya, S. (2023). Loyalty from a personal point of view: A cross-cultural prototype study of loyalty. *Journal of Experimental Psychology: General*.

Journal of Experimental Psychology: General.

Murray, D. R., & Schaller, M. (2010). Historical prevalence of infectious diseases within 230 geopolitical regions: A tool for investigating origins of culture. *Journal of Cross-Cultural Psychology*, 41(1), 99–108.

<https://doi.org/10.1177/0022022109349510>

Muthukrishna, M., Bell, A. V., Henrich, J., Curtin, C. M., Gedranovich, A., McInerney, J., &

Thue, B. (2020). Beyond Western, Educated, Industrial, Rich, and Democratic (WEIRD) psychology: Measuring and mapping scales of cultural and psychological distance.

Psychological Science, 31(6), 678–701. <https://doi.org/10.1177/0956797620916782>

Neuendorf, K. A. (2017). *The content analysis guidebook*. Sage Publications Inc.

<https://doi.org/10.4135/9781071802878>

Newman, D. B., Schwarz, N., Graham, J., & Stone, A. A. (2019). Conservatives report greater meaning in life than liberals. *Social Psychological and Personality Science*, 10(4),

494–503. <https://doi.org/10.1177/1948550618768241>

Phillips, J., Misenheimer, L., & Knobe, J. (2011). The ordinary concept of happiness (and others like it). *Emotion Review*, 3(3), 320–322. <https://doi.org/10.1177/1754073911402385>

Quiles, Z. N., & Bybee, J. (1997). Chronic and predispositional guilt: Relations to mental health, prosocial behavior, and religiosity. *Journal of Personality Assessment*, 69(1), 104–126.

https://doi.org/10.1207/s15327752jpa6901_6

- Rozin, P., Haidt, J., & McCauley, C. (2009). *Disgust: The body and soul emotion in the 21st century*. American Psychological Association. <https://doi.org/10.1037/11856-001>
- Schafer, J. L., & Graham, J. W. (2002). Missing data: our view of the state of the art. *Psychological methods*, 7(2), 147–177.
- Schein, C. (2020). The importance of context in moral judgments. *Perspectives on Psychological Science*, 15(2), 207–215. <https://doi.org/10.1177/1745691620904083>
- Schwartz, S. H., Cieciuch, J., Vecchione, M., Davidov, E., Fischer, R., Beierlein, C., Ramos, A., Verkasalo, M., Lönnqvist, J.-E., Demirutku, K., Dirilen-Gumus, O., & Konty, M. (2012). Refining the theory of basic individual values. *Journal of Personality and Social Psychology*, 103(4), 663–688. <https://doi.org/10.1037/a0029393>
- Thaler, R. (1985). Mental accounting and consumer choice. *Marketing science*, 4(3), 199–214.
- Thomson, R., Yuki, M., Talhelm, T., Schug, J., Kito, M., Ayanian, A. H., Becker, J. C., Becker, M., Chiu, C. Y., Choi, H. S., Ferreira, C. M., Fülöp, M., Gul, P., Houghton-Illera, A. M., Joasoo, M., Jong, J., Kavanagh, C. M., Khutkyy, D., Manzi, C., Marcinkowska, U. M. ... Visserman, M. L. (2018). Relational mobility predicts social behaviors in 39 countries and is tied to historical farming and threat. *Proceedings of the National Academy of Sciences of the United States of America*, 115(29), 7521–7526. <https://doi.org/10.1073/pnas.1713191115>
- van Leeuwen, F., Park, J. H., Koenig, B. L., & Graham, J. (2012). Regional variation in pathogen prevalence predicts endorsement of group-focused moral concerns. *Evolution and Human Behavior*, 33(5), 429–437. <https://doi.org/10.1016/j.evolhumbehav.2011.12.005>

- Wagemans, F., Brandt, M. J., & Zeelenberg, M. (2018). Disgust sensitivity is primarily associated with purity-based moral judgments. *Emotion, 18*(2), 277–289.
<https://doi.org/10.1037/emo0000359>
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology, 54*(6), 1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>
- Waytz, A., & Hofmann, W. (2020). Nudging the better angels of our nature: A field experiment on morality and well-being. *Emotion, 20*(5), 904–909.
<https://doi.org/10.1037/emo0000588>
- Yu, Z., Bali, P., Tsikandilakis, M., & Tong, E. M. W. (2022). ‘Look not at what is contrary to propriety’: A meta-analytic exploration of the association between religiosity and sensitivity to disgust. *British Journal of Social Psychology, 61*, 276–299.
<https://doi.org/10.1111/bjso.12479>
- Zapko-Willmes, A., Schwartz, S. H., Richter, J., & Kandler, C. (2021). Basic value orientations and moral foundations: Convergent or discriminant constructs? *Journal of Research in Personality, 92*, 104099. <https://doi.org/10.1016/j.jrp.2021.104099>