



HAL
open science

Social Norm Deviation as a Process Underlying the Induced-Hypocrisy Effect: Findings from a Five-Study Experimental Program

Mauduy Maxime, Daniel Priolo, Nicolas Margas, Cécile Sénémeaud

► To cite this version:

Mauduy Maxime, Daniel Priolo, Nicolas Margas, Cécile Sénémeaud. Social Norm Deviation as a Process Underlying the Induced-Hypocrisy Effect: Findings from a Five-Study Experimental Program. 2025. <hal-05061392>

HAL Id: hal-05061392

<https://hal.science/hal-05061392v1>

Preprint submitted on 9 May 2025

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

**Social Norm Deviation as a Process Underlying the Induced-Hypocrisy Effect:
Findings from a Five-Study Experimental Program**

Maxime Mauduy¹, Daniel Priolo², Nicolas Margas³, Cécile Sénémeaud⁴

¹ Université Paris Cité, Laboratoire de Psychologie Sociale : contextes et régulation (LPS, UR 4471, Boulogne-Billancourt, France.

² Université Paul Valéry Montpellier 3, EPSYLON, 34199 Montpellier, France.

³ Institute of Sport Sciences, University of Lausanne, Switzerland.

⁴ Université de Caen-Normandie, LPCN, 14000 Caen, France.

Author notes

Maxime Mauduy, ORCID iD: <https://orcid.org/0000-0002-4604-5314>

Daniel Priolo, ORCID iD: <https://orcid.org/0000-0003-3430-3459>

Nicolas Margas, ORCID iD: <https://orcid.org/0000-0001-6388-4032>

Cécile Sénémeaud, ORCID iD: <https://orcid.org/0000-0003-2195-1457>

All data, analysis code, and materials have been made publicly available on OSF and can be accessed in the Supplementary materials at

https://osf.io/k53mq/?view_only=d4752d9b66e240f987390e51e6dc6fcb.

Studies 1, 2, 3 were not preregistered. Design, hypotheses, and analysis plan of Studies 4a and 4b were preregistered on OSF (Study 4a: <https://osf.io/h8fr7>, Study 4b: <https://osf.io/tkxgs>)

Correspondence concerning this article should be addressed to Maxime Mauduy, Université Paris Cité, Laboratoire de Psychologie Sociale (LPS, UR 4471), Institut de Psychologie, 71 avenue Édouard Vaillant, 92774 Boulogne-Billancourt, France. Email: maxime.mauduy@u-paris.fr. Phone Number: +33659576492.

Word count (From abstract to References, excluding title page, figures and tables): 14200

Abstract

Induced-hypocrisy, a cognitive dissonance paradigm, effectively promotes prosocial behaviors through a two-step procedure. First, a person advocates for a prosocial behavior and then recalls their past failures to comply with it, which prompts the person to adopt the behavior. Initially, the explanation for such hypocrisy effect focused on restoring a positive self-image threatened by past transgressions. A recent theoretical explanation, suggest that behavioral change results from the perceived discrepancy between social norms and transgressive behavior. Our five-study experimental program (*Total N = 774*) was designed to test whether the process of social norm deviation underlies the hypocrisy effect. Results showed the influence of social norms strength on the hypocrisy effect (Studies 1, 2, 3) and the mediating effect of perceived social norm deviation in increasing prosocial behaviors consecutive to the hypocrisy procedure (Studies 4a and 4b). An integrative analysis (*N = 602*) of Studies 1, 2 and 4a, as well as two additional datasets, confirmed the influence of social norms in producing the hypocrisy effect. Taken together, these results strongly support social norm deviation as a key process underlying the hypocrisy effect, in that participants with stronger social normative beliefs exhibit a more pronounced hypocrisy effect due to a greater perceived deviation from social norms. These experimental findings constitute a significant contribution to understanding and improving the hypocrisy effect. They are discussed from both theoretical and practical perspectives, questioning the implication of the self and the affiliation motivation in this social norm deviation process and emphasizing the relevance of reinforcing social norms in contexts where transgressions are less salient, such as in discrimination prevention.

Keywords

Induced-hypocrisy paradigm, social norm, deviation, behavior change, discrimination

Social Norm Deviation as a Process Underlying the Induced-Hypocrisy Effect: Findings from a Five-Study Experimental Program

Everyday situations often reveal a gap between socially approved issues and personal actions related to these issues. These situations are classic examples of cognitive dissonance (Festinger, 1957) and have led to an effective strategy for promoting prosocial behaviors, namely induced-hypocrisy paradigm (IHP, Aronson et al., 1991). This paradigm involves two steps (Stone & Fernandez, 2008): participants express support for a prosocial behavior (normative advocacy step), then recall instances where they failed to act in accordance with it (transgression salience step). This inconsistency generates the hypocrisy effect, prompting people to adopt future behaviors consistent with the prosocial behavior. Three decades of research have shown induced hypocrisy to be effective in various domains, such as environment (e.g., water-saving, Dickerson et al., 1992) and health (e.g., binge drinking, Sénémeaud et al., 2014). However, although induced hypocrisy relates to socially normative domains (Liégeois et al., 2017), social norms have been underexplored in explaining hypocrisy effects. A recent meta-analysis (Mauduy, Priolo, et al., 2023) has indirectly demonstrated the influence of the normative salience in producing the hypocrisy effect, and this leads naturally to the need for an experimental demonstration of this effect. This article presents five experiments and one integrative analysis designed to experimentally test and understand better the involvement of social norms in producing the hypocrisy effect.

Self-Standard Model of dissonance: a dual pathway to hypocrisy effects

The hypocrisy effect was originally explained by the motivation to restore a threatened positive self-image during the transgression salience step, as proposed by self-consistency theory (Aronson, 1999). Consequently, researchers focused heavily on this step, often

neglecting the first one (Mauduy, Priolo, et al., 2023). It wasn't until Stone and Fernandez (2008) that a dual pathway explanation for the hypocrisy effect was introduced, based on the Self-Standard Model of dissonance (SSM, Stone & Cooper, 2001). According to the SSM, individuals evaluate their past behavior against standards of judgment related to prosocial behavior, which can be either personal or normative. Personal standards reflect personal norms (Thøgersen, 2006), while normative standards represent social norms (Cialdini & Trost, 1998). When personal norms are salient, behavior is assessed against them, leading to a threat to self-concept if the behavior contradicts these norms (personal norm deviation pathway). However, when social norms are salient, behavior is compared to others' expectations, not necessarily involving self-threat. The hypocrisy effect thus arises from the inconsistency between behavioral standards (either normative or personal) emphasized in the first step and the transgressions of these standards highlighted in the second step (Stone & Fernandez, 2008).

Because Aronson et al. (1991) initially developed the IHP to support his self-consistency view of dissonance (Aronson, 1968, 1999), hypocrisy effects have long remained explained solely in terms of self-consistency, corresponding to the personal norm deviation pathway of the SSM.

A well-documented personal norm deviation pathway

Self-consistency theory (Aronson, 1968, 1999) posits that dissonance arousal involves the self-concept, particularly threatening it. The self-concept represents personal norms and people self-integrity (i.e., positive self-image). It is threatened when individuals engage in behaviors contrary to their personal norms. Building on this theoretical framework, researchers have therefore explored and demonstrated the role of self-threat in the hypocrisy effect. They found that self-esteem moderates the hypocrisy effect (Peterson et al., 2008) and self-integrity manipulations (threatening vs. supporting) also influence this effect (Fointiat, 2004). In

addition, researchers investigated the optimal conditions for performing the transgression salience step (Stone & Fernandez, 2008). Effective transgressions recall must be personal, anonymous, infrequent, recent, and described without justification to generate the hypocrisy effect. Moreover, numerous transgressions or significant (vs. trivial) elaborate transgressions are essential (Stone & Fernandez, 2011).

However, this emphasis on the self and the transgressions has overlooked the behavioral standard salience step and, more generally, the social norm deviation pathway.

A neglected social norm deviation pathway

Since behaviors targeted in the IHP are always socially desirable, Priolo et al., (2016) propose that the perception of inconsistency between social norm and transgressive behavior underlies the hypocrisy effect. According to the perceived social norm deviation approach, the more people perceive a gap between their behaviors and their beliefs about social norms, the stronger the hypocrisy effect and the greater the behavioral change. While characteristics of recalled transgressions could contribute to widening this gap (Liégeois et al., 2017), another avenue lies in the first IHP step, associated to social norms.

Priolo et al. (2016) showed that the hypocrisy effect could arise from inconsistency with both descriptive (i.e., individual beliefs about majority behavior) and injunctive norms (i.e., individual beliefs about majority approval). Additionally, Mauduy et al. (2022) found a greater hypocrisy effect when both types of social norms are made salient. Mauduy & Mange (2023) also found that making a situational injunctive norm salient (i.e., this school's adults disapprove of school bullying), rather than a general injunctive norm (i.e., school bullying is socially frowned upon), enhances the hypocrisy effect. Finally, Mauduy et al.'s meta-analysis (2023) showed that the hypocrisy effect was larger in studies categorized as strongly (vs. weakly) emphasizing the social norm associated with the behavior in the first IHP step. These results

are first indirect evidence of the influence of social norms strength in hypocrisy effects, but direct and experimental tests are needed. Thus, the general aim of this research is to experimentally test the involvement of the social norm deviation process in generating the hypocrisy effect. Testing this process in the context of discrimination prevention seems particularly relevant.

Investigating social norm deviation process in the context of discrimination prevention

Discrimination, defined as “the differential treatment of people on the basis of their membership in a given group” (Bodenhausen & Richeson, 2010, p. 343), remains a major societal issue that requires sustained efforts for its reduction (DARES, 2021; Kaas & Manger, 2012; Sasaki et al., 2017). Surprisingly, few studies have applied the IHP to the reduction of discriminatory behaviors (Mauduy et al., 2022; Son Hing et al., 2002). While promising, these studies face a critical limitation: the difficulty of eliciting the recall of one’s own transgressions. In the domain of discrimination, individuals often suppress, minimize, or fail to recognize their own prejudiced or discriminatory actions, especially in public or research settings (Crandall et al., 2002). As such, strategies that rely on increasing the salience of past transgressions - commonly used to strengthen the hypocrisy effect (e.g., Sénémeaud et al., 2014) - may not be suitable when addressing discrimination.

From a social norm deviation perspective, a more suitable approach may be to amplify individuals’ perceived deviation by reinforcing the salience of anti-discriminatory norms, rather than by focusing on transgressions. Importantly, several well-established procedures exist to strengthen perceived social norms, such as exposing participants to (ostensibly) scientific evidence of widespread egalitarian values (e.g., Gabarrot et al., 2009). The context of discrimination thus offers a particularly relevant setting for testing our theoretical hypothesis

that strengthening social norms activation within the IHP procedure can increase participants' perceived norm deviation and, in turn, promote behavior change.

This approach is further supported by previous research in the discrimination context showing that perceived norm deviation - typically assessed via the "Should-Would Discrepancy" questionnaire (Devine et al., 1991) - elicits psychological discomfort, including guilt and self-directed threat, which in turn can influence behavioral intentions and actions (Amodio et al., 2007; Fehr & Sassenberg, 2010; Monteith et al., 2002; Plant & Devine, 1998). However, this questionnaire has rarely been used with intervention goals, and no studies have attempted to increase its effectiveness by actively manipulating perceived deviation. Given the conceptual proximity between the IHP and this "Should-Would Discrepancy" questionnaire - both of which rely on highlighting discrepancies between norms and behavior - our work could provide a promising opportunity to extend and strengthen the use of social norm deviation processes in discrimination reduction.

Research overview

The present research aims to experimentally test the impact of perceived social norm deviation on the generation of the hypocrisy effect in the context of discrimination prevention. As a strategy for manipulating social norm deviation would be to enhance, rather than reduce, the transgression salience, the person's social normative beliefs, we manipulated, drawing on normative influence research, the strength of social norms within the IHP. Reinforcing them should heighten the perception of social norm deviation after transgressions of them have been recalled, thereby amplifying the hypocrisy effect itself. Thus, we hypothesized that, in an induced-hypocrisy situation, (1) stronger social normative beliefs would yield a greater hypocrisy effect (*General Hypothesis 1, GH1*) because (2) participants would perceive their past transgressive behaviors as deviating more from their social normative beliefs (*GH2*).

To test these two general hypotheses, we set up a program of five experimental studies. The first three studies tested the influence of the strength of social norms on the hypocrisy effect (*GH1*). More specifically, we tested the influence of social norm belief reinforcement on the hypocrisy effect (Study 1). Then, we replicated this effect of social norm reinforcement while also testing the effect of weakening it (Study 2). Following this, we checked that our experimental manipulations of social norms (i.e., strengthening and weakening) were effective in modifying beliefs relating to social norms without simultaneously affecting those relating to personal norms (Study 3). The last two studies tested the mediating effect of social norm deviation perception in increasing prosocial behaviors consecutive to the hypocrisy effect (*GH2*). Given that experimental-chain-designs are now recommended to provide stronger evidence for mediation effects (Pirlott & MacKinnon, 2016), we chose to test our mediation hypothesis regarding the social norm deviation process using an experimental-chain design. We first measured the perceived social norm deviation following the hypocrisy procedure (Study 4a, pre-registered, <https://osf.io/h8fr7>) and then manipulated this social norm deviation - independently of a hypocrisy procedure - to observe whether its effects on engagement in prosocial behavior were similar to those obtained in a hypocrisy procedure (Study 4b, pre-registered, <https://osf.io/tkxgs>). Finally, we conducted an integrative analysis including five studies (Studies 1, 2, 4a, and two additional datasets) selected because they tested the influence of social norm strength on the reinforcement of the hypocrisy effect.

Ethics, Transparency and Openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study, and we follow JARS (Appelbaum et al., 2018). All data, analysis code, and research materials are available at https://osf.io/k53mq/?view_only=d4752d9b66e240f987390e51e6dc6fcb. Data were analyzed using Jamovi, version 2.4.14, and R, version 4.1.1 (R Core Team, 2021). Furthermore, the study

was conducted in full agreement with the Declaration of Helsinki (World Medical Association, 2008) and the ethical standards set by a Psychology Department, which follows the American Psychological Association Ethical Principles of Psychologists and the Code of Conduct (American Psychological Association, 2017) for the ethical treatment of human participants. With an information letter, participants were informed about the study task, its duration, and that their participation was voluntary, confidential and anonymous. Participants could refuse to participate and could withdraw from the study at any time without any consequences. Informed, individual consent to participate was requested before and after the study completion to authorize the retention of their data for research purposes. This transparency, openness and ethics were the same for all the studies in this research.

Study 1

Study 1 aims to experimentally test for the first time the influence of social norms strength on the hypocrisy effect. If deviation from social norm underlies the hypocrisy effect, then reinforcing social norms should increase the hypocrisy effect.

Method

Participants

We recruited 180 Caen university students to participate in this laboratory study, without any counterpart. Since 20 participants did not fully complete the experimental protocol, the final sample consisted of 160 students from the University of Caen ($M = 20.0$, $SD = 2.4$; 78.12% female). This sample size, with an expected effect size of $OR = 2.47$ for the contrast of interest, allowed us to achieve a satisfactory statistical power above 90% (see Supplementary Materials 1, Section 1.1).

Design and Procedure

Participants were individually invited to take part in the study. They were informed about the study via an information letter and then asked to sign a consent form for their participation (see Supplement 1, Section 1.2). Subsequently, participants were randomly assigned to one of four experimental conditions: control, norm, standard hypocrisy, and reinforced hypocrisy¹.

In the *reinforced hypocrisy condition*, participants were exposed to a strong anti-discrimination social norm through the results of a purported study surveying students at Caen University about their attitudes and behaviors towards French people of foreign origin (FPFO). Subsequently, they underwent both induced-hypocrisy steps. They were asked to write a short statement against discrimination towards FPFO, followed by the recall of their past discriminatory behaviors towards FPFO. Finally, measures of behavioral intention were introduced as dependent variables. Participants in the *standard hypocrisy condition* followed a similar protocol to those in the reinforced hypocrisy condition, with the exception that social norms were not reinforced beforehand. However, they engaged in the same exercise, focusing on a purported "neutral" study (i.e., extracurricular activities of students belonging to their university). Cognitive processing and activation of the "Caen university students" group were thus comparable to the reinforced hypocrisy condition. Participants in the *norm condition* underwent the same protocol as those in the standard hypocrisy condition, excluding the transgression-salience step. Participants in the *control condition* directly responded to the dependent measures.

Upon completion of the study, all participants received debriefing, including clarification of the study's true purpose and correction of misinformation about the purported study results and the prevention association. Any identifying information from participants who

¹ The same procedure was followed for each of the experiments presented in this article. We will no longer mention it in the description of the following experiments.

agreed to join the association was not retained and was returned to them. Participants were then invited to complete a post-consent declaration (see Supplement 1, Section 1.3) to authorize the retention of their data for research purposes and were sincerely thanked for their participation in the study².

Materials

Strengthening social norms.

To reinforce social norms, we followed the method of Gabarrot et al. (2009) by providing participants with normative information against discrimination (see Supplement 1, Section 1.4. for an illustration). Participants were informed about the results of a purported study carried out with a representative sample of students from Caen University. Results informed participants that 82.25% of students consider as illegitimate favoring French people over French People of Foreign Origin, and that 89.26% of students do not discriminate FPFO. This normative information was supposed to activate strong injunctive and descriptive anti-discrimination norms (Cialdini et al., 1990; Gabarrot et al., 2009). Participants then answered a comprehension question to confirm their understanding of the study results, as recommended by Smith & Louis (2008).

Normative Advocacy.

The most commonly used procedure to conduct the first induced-hypocrisy step consists of performing a normative advocacy toward a prosocial cause or behavior (Mauduy, Priolo, et al., 2023; Stone & Fernandez, 2008). Therefore, we asked participants to write three arguments demonstrating why it is important to treat people equally regardless of their origins (see Supplement 1, Section 1.5 for an illustration). As recommended, we emphasized the public

² This post-experimental debriefing was conducted for all experiments presented in this article. We will no longer mention it in the description of the following experiments.

nature of this advocacy by informing participants that their arguments would be used to build prevention programs for young people.

Recall of past transgressive behaviors.

Participants filled out a questionnaire concerning their past behaviors in discrimination situations (Fried, 1998). Given that the number of recalled transgressions influences the magnitude of the hypocrisy effect (e.g., Sénémeaud et al., 2014), we neutralized this effect by asking participants to recall the same number of transgressions. A two-phase pre-test ($N = 60$; $M = 22.3$, $SD = 1.7$; 78.3% female, see Supplement 1, Section 1.6 for more details) identified four discriminatory situations towards FPFO most frequently carried out by university students that thus constituted the transgression recall questionnaire (i.e., negative thoughts, negative remarks, keeping distance, and stereotyping French people of African origin). Consistent with recommendations (Fried, 1998; Stone & Fernandez, 2008), participants were asked for specifics about each situation they encountered by answering three questions (i.e., When was the last time?; Where were you?; Who was involved?).

Dependent measures.

Prosocial behavioral intention is the most classic mean used for measuring the hypocrisy effect (Priolo et al., 2019). Therefore, participants' antidiscrimination behavioral intention was measured by asking them whether or not they would agree to distribute flyers on behalf of a fictitious association fighting against FPFO discrimination. Participants who agreed to join were asked to indicate the amount of time they wished to allocate (from 0: half an hour to 8: eight half-hours) as well as their name, surname, and email address for future contact (see Supplement 1, Section 1.7 for an illustration). This procedure allowed us to measure (1) the *engagement to volunteer* (acceptance rate of the target request as binary variable) and (2) the *amount of volunteering* (i.e., the time allocated to the association). The higher the proportion of

participants agreeing to the target request and the more time promised to the association, the higher the hypocrisy effect.

Hypotheses and Statistical Analyses Plan

We expected participants' engagement to volunteering (H₁) and amount of volunteering (H₂) to show a positive trend from the control and norm conditions, through the standard hypocrisy condition, and up to the reinforced hypocrisy condition.

To test our specific hypotheses, we used planned comparisons as they are recommended to be more efficiently than omnibus tests (see Abdi & Williams, 2010; Judd et al., 2017). Three orthogonal contrasts, including our contrast of interest (C1: Control = -2, Norm = -2, Standard Hypocrisy = +1, Reinforced Hypocrisy = +3) and its two residuals (C2: Control = -1, Norm = +1, Standard Hypocrisy = 0, Reinforced Hypocrisy = 0; C3: Control = +1, Norm = +1, Standard Hypocrisy = -5, Reinforced Hypocrisy = +3) were used to test our hypotheses. C1 should be positive and significant, while C2 and C3, as well as their combined residual effect, should not be significant to support our hypotheses (Abdi & Williams, 2010). H₁ was tested using a binomial logistic regression model on the engagement to volunteer, with the three contrasts as predictors. H₂ was tested using a linear regression model on the amount of volunteering, with the three contrasts as predictors.

Results

Preliminary analyses (see Supplement 2, Section 1.1) ensured the random distribution of participants across experimental conditions based on socio-demographic variables (gender and age). Additionally, we confirmed that the number of transgressions recalled did not significantly differ between our two hypocrisy conditions, considering its influence on the hypocrisy effect (e.g., Sénémeaud et al., 2014).

Descriptive data are displayed in **Table 1**.

Table 1.

Engagement to volunteer and amount of volunteering among the four experimental conditions (Study 1).

	Engagement to volunteer (target-request acceptance)	Amount of volunteering (in minutes)
Control	22.50 (9/40)	14.3 (29.6)
Norm	15.0 (6/40)	10.5 (32.3)
Standard hypocrisy	32.50 (13/40)	28.5 (46.6)
Reinforced hypocrisy	42.50 (17/40)	31.5 (51.7)

Note. $N = 160$. For engagement to volunteer, data show percentages (ratio); for amount of volunteering data show means (standard deviations).

Effect on engagement to volunteer

The logistic regression model on engagement to volunteer is globally significant, $\chi^2(3) = 8.66$, $p = .03$., $R^2_{Nagelkerke} = .076$, More specifically, our contrast of interest C1 is significant ($Estimate = 0.23$, $\chi^2_{Wald}(1) = 2.78$, $p = .005$, $OR = 1.27$, 95% CI [1.07, 1.50]) while the residuals C2 ($Estimate = -0.25$, $\chi^2_{Wald}(1) = -0.85$, $p = .39$, $OR = 0.78$, 95% CI [0.44, 1.38]) and C3 ($Estimate = -0.01$, $\chi^2_{Wald}(1) = -0.11$, $p = .91$, $OR = 0.99$, 95% CI [0.89, 1.11]) and their combined effect ($\chi^2(2) = 0.751$, $p = .69$) are not. These results support our hypothesis H₁ and indicate that the proportion of participants who agreed to take part in the anti-discrimination actions follows an upward trend from the control and norm conditions, through the standard hypocrisy condition, and up to the reinforced hypocrisy condition.

Effects on the amount of volunteering

The linear regression model on the amount of volunteering is not globally significant at the α threefold of .05, $F(3,156) = 2.54$, $p = .058$, $R^2 = .047$. However, our contrast of interest C1 is significant ($Estimate = 0.14$, $\beta = 0.21$, 95% CI [0.05, 0.36], $p = .008$) while the two residual contrasts C2 ($Estimate = -0.06$, $\beta = -0.03$, 95% CI [-0.19, 0.12], $p = .68$) and C3 ($Estimate = -0.02$, $\beta = -0.05$, 95% CI [-0.20, 0.11], $p = .55$) and their combined residual effect

($\Delta R^2 = .003$, $F(2,156) = 0.261$, $p = .77$) are not. As expected (H_2), participants' amount of volunteering to the anti-discrimination association shows an upward trend from the control and advocacy-only conditions, through the standard hypocrisy condition, and up to the reinforced hypocrisy condition.

Discussion

Results of Study 1 support the hypocrisy effect in the context of discriminatory behavior prevention (Mauduy et al., 2022; Son Hing et al., 2002). More importantly, as expected (GHI), they showed that strengthening social norms enhanced the hypocrisy effect. In the next study, we aimed to replicate this first finding while deepening the influence of social norm strength. Indeed, if the social norm deviation process drives the hypocrisy effect, then both strengthening and weakening social norms should impact the hypocrisy effect, by increasing or reducing it. Furthermore, we also aimed to check that the pro-social cause against discrimination was also a personal cause for our study population, a detail which we did not check in our first study. This is important because the anti-discrimination normative advocacy carried out as part of the first step of the IHP could in itself be a source of dissonance if the participants are not personally against discrimination.

Study 2

Study 2 aimed to (1) replicate the higher hypocrisy effect in the strengthening social norms condition, and (2) test the lower hypocrisy effect in the weakening social norms condition. Therefore, we provided some participants with normative information aimed at reducing the strength of social norms.

Method

Participants

We recruited 180 Caen university students to take part in this second laboratory experiment. Twenty participants were withdrawn because they had not fully completed the induced hypocrisy procedure (i.e., no transgression recalled). Our final sample consisted of 160 students ($M = 19.6$, $SD = 1.54$; 78.1% female). This sample size, with an expected effect size of $OR = 2.47$ for the contrast of interest, allows to achieve a satisfactory statistical power above 80% (see Supplement 1, Section 2.1).

Design and Procedure

Participants in the *control condition* were asked to directly respond to the dependent variable questions, namely the behavioral intention measures (i.e., engagement and amount of volunteering in discrimination prevention association), a discriminatory behavior measure, based on the Hierarchical Organizational Task (Anier et al., 2018), and a scale measuring personal anti-discrimination beliefs. Participants in the three hypocrisy conditions were asked to complete an exercise in understanding the results of a purported study. The results were intended either to reinforce anti-discriminatory social norms (*reinforced hypocrisy condition*), to weaken these norms (*weakened hypocrisy condition*), or to activate an irrelevant norm (i.e., related to students' extracurricular activities; *standard hypocrisy condition*). Next, they were asked to carry out the classic induced-hypocrisy two steps: the normative advocacy (i.e., writing three arguments against origin-related discrimination) and the recall of past transgressions (i.e., recall of five situations of discrimination against FPFO). Finally, they filled in the measures for the dependent variables.

Materials

Strengthening and weakening social norms.

To reinforce social norms, we followed the same procedure as in Study 1. To weaken anti-discriminatory social norms, we also followed the same procedure, but told participants that a minority of Caen students did not discriminate against FPFO (i.e., 29.26%) and disapproved of discrimination against them (i.e., 22.25%, see Supplement 1, Section 2.2 for more details).

Normative Advocacy.

The procedure and material were the same as Study 1.

Recall of past transgressive behaviors.

The procedure was the same as the Study 1 but the material was not. The pre-test of Study 1 aimed at identifying discriminatory behaviors had not been carried out on our study population, students at Caen university. To improve the relevance of the identification of the most frequent discriminatory behaviors in our study population, we carried out a second pre-test ($N = 150$, $M = 19.57$, $SD = 2.53$; 70% female; see Supplement 1, Section 2.3 for more details). Thus, the transgression recall (see Supplement 1, Section 2.4) included five discrimination situations (i.e., keeping one's distance, criticizing, excluding, staring at and ignoring a FPFO).

Dependent measures.

The first two dependent variables were the same as those used in Study 1. First, participants were asked whether they would like to volunteer for a discrimination prevention association (engagement to volunteer), and how much time they would like to give to the association (amount of volunteering). The higher the proportion of participants accepting the target request and the time allocated to the association, the higher the hypocrisy effect.

As a third dependent variables, we also used the Organizational Hierarchy Task (Michinov et al., 2005) to obtain an indirect measure of a discriminatory behavior. The procedure and materials were taken from Anier et al. (2018: study 2); see Supplement 1, Section 2.5 for an illustration). Participants were informed that the university was setting up an association designed to facilitate the integration of FPFO students. The members of the association had already been recruited. The university needed the participants' advice to organize the association board. In concrete terms, participants were presented with an organizational hierarchy and the pictures of six individuals (including their first name). There were two North-African men (stereotypically named "Abdel" and "Rachid"), two Caucasian men (named "Alexandre" and "Pascal") and two Caucasian women (named "Laetitia" and "Sandrine"). Photos of real individuals have been used from Langner et al.'s database (2010). After presenting this context, participants were asked to assign a position – among eight along four hierarchy levels – in this organizational hierarchy to each of these individuals as well as themselves. Indirect discriminatory behaviors were measured by looking at the position assigned to the two North-African individuals. The more North African targets are assigned to the bottom, the more the participant discriminates against them³. Following Anier et al. (2018), we assigned a score to each of the four levels of the organizational hierarchy (from 1 for the highest position to 4 for the lowest), then averaged the score of the two North-African targets (from 1.5, meaning the two highest positions were assigned to the two North African targets, to 4, meaning the two North African targets were assigned the lowest positions) and the average score of the two Caucasian men. The average position assigned to Caucasian targets ("Alexandre" and "Pascal") was subtracted from that assigned to North-African targets ("Abdel" and "Rachid") to operationalize the measure of indirect discriminatory behavior. The

³ This measure was created within the framework of the social dominance theory (Sidanius & Pratto, 2001). In this context, the more North African target assigns at the bottom, the more the participant reinforces existing hierarchical structures between minority and majority group by discriminating.

more negative the difference, the less participants discriminated against North-African individuals.

Finally, to measure anti-discriminatory personal beliefs, we used 6 items on a 7-point Likert scale (from 1 - *Strongly Disagree* to 7 - *Strongly Agree*). Based on the work of Thøgersen (2006), we measured personal normative beliefs using three items relating to the introjected norm (e.g., “*I would have a bad conscience if I treated people of foreign origin differently from people of French origin*”) and three items relating to the integrated norm (e.g., “*Because of my own values, I would feel bad about discriminating against people of foreign origin*”). To complement these 6 items, 15 filler items were added to the questionnaire to limit social desirability (see Supplement 1, Section 3.2 for all items used), as recommended by Vautier et al. (2003). These 15 items measure participants' normative beliefs on themes other than discrimination (i.e., blood donation, organic food and cannabis use).

Hypotheses and Data Analysis Plan

We expected participants' engagement in volunteering (H₁) and amount of volunteering (H₂) to show a positive trend from the control condition, through the weakened hypocrisy condition, through the standard hypocrisy condition, and up to the reinforced hypocrisy condition. We also expected participants' discriminatory behaviors (H₃) to show a negative trend from the control condition, through the weakened hypocrisy condition, through the standard hypocrisy condition, and up to the reinforced hypocrisy condition.

As for Study 1, data were analyzed by using contrast analyses (Abdi & Williams, 2010), and more specifically by means of polynomial contrasts (see Supplement 1, Section 2.6). The linear contrast represented our hypotheses (i.e., control= -3; weakened hypocrisy= -1; standard hypocrisy= +1; reinforced hypocrisy= +3). So, we expected that it was significant while the quadratic and cubic contrasts and their combined residual effect were not. H₁ was tested using a binomial logistic regression model on engagement to volunteer, with the three contrasts as

predictors. H₂ and H₃ were tested using linear regression models on the amount of volunteering and the discriminatory behavior score, with the three contrasts as predictors.

Results

Preliminary analyses (see Supplement 2, Section 2.1) ensured the random distribution of participants across experimental conditions based on socio-demographic variables (gender and age) and that the number of transgressions recalled did not significantly differ between the three hypocrisy conditions. Finally, none of the four conditions differed significantly on the measure of personal beliefs (6 items, McDonald's $\omega = .70$), further indicating high levels of personal anti-discrimination beliefs for all 4 experimental conditions (control: $M = 6.71$, $SD = 0.55$; weakened-hypocrisy: $M = 6.64$, $SD = 0.51$; standard-hypocrisy: $M = 6.63$, $SD = 0.55$; reinforced-hypocrisy: $M = 6.68$, $SD = 0.60$). This indicates that the antidiscrimination prosocial cause is also a personal cause in our study population.

Descriptive data are displayed in **Table 2**.

Table 2.

Engagement to volunteer, amount of volunteering and score of discriminatory behaviors (SD) among the four experimental conditions (Study 2)

	Engagement to volunteer (target-request acceptance)	Amount of volunteering (in minutes)	Discrimination score
Control	17.5 (7/40)	14.3 (34.0)	-0.66 (1.16)
Weakened Hypocrisy	27.5 (11/40)	15.0 (31.0)	-0.85 (0.97)
Standard hypocrisy	37.5 (15/40)	22.5 (35.1)	-1.02 (0.91)
Reinforced hypocrisy	50.0 (20/40)	34.5 (45.3)	-1.14 (0.64)

Note. $N = 160$. For engagement to volunteer, data show percentages (ratio); for amount of volunteering and discrimination score (The more negative the difference, the less participants discriminated against North-African persons), data show means (standard deviations).

Effect on engagement to volunteer

The logistic regression model on engagement to volunteer is globally significant, $\chi^2(3) = 10.69$, $p = .014$, $R^2_{Nagelkerke} = .09$. More specifically, results indicate that the linear contrast is significant ($Estimate = 0.26$, $\chi^2_{Wald}(1) = 3.12$, $p = .002$, $OR = 1.29$, 95% CI [1.10, 1.52]) while the residual quadratic ($Estimate = -0.02$, $\chi^2_{Wald}(1) = -0.10$, $p = .92$, $OR = 0.98$, 95% CI [0.69, 1.39]) and cubic contrasts ($Estimate = 0.01$, $\chi^2_{Wald}(1) = 0.11$, $p = .91$, $OR = 1.01$, 95% CI [0.87, 1.17]) and their combined effect ($\chi^2_{(2)} = 0.02$, $p = .99$) are not. These results support our hypothesis H₁ and indicate that the proportion of participants who agreed to take part in the anti-discrimination association actions follows an upward trend from the control condition, through the weakened hypocrisy condition, through the standard hypocrisy condition, and up to the reinforced hypocrisy condition.

Effect on the amount of volunteering

The linear regression model on the amount of volunteering is not globally significant at the α threefold of .05, $F(3,156) = 2.54$, $p = .054$, $R^2 = .048$. However, the linear contrast is significant ($Estimate = 0.11$, $\beta = 0.21$, 95% CI [0.05, 0.36], $p = .01$) while the residual quadratic ($Estimate = 0.09$, $\beta = 0.08$, 95% CI [-0.08, 0.23], $p = .34$) and cubic contrasts ($Estimate = -0.004$, $\beta = -0.006$, 95% CI [-0.16, 0.15], $p = .93$) and their combined residual effect ($\Delta R^2 = .006$, $F(2,156) = 0.471$, $p = .63$) are not. As expected (H₂), participants' amount of volunteering to the anti-discrimination association shows an upward trend from the control condition, through the weakened hypocrisy condition, through the standard hypocrisy condition, and up to the reinforced hypocrisy condition.

Effect on discriminatory behavior in the hierarchical organizational task

The linear regression model on the discriminatory behavior score is not globally significant at the α threefold of .05, $F(3,156) = 1.97$, $p = .12$, $R^2 = .037$. However, the linear contrast is

significant ($Estimate = -0.08$, $\beta = -0.19$, 95% CI [-0.35, -0.04], $p = .017$), while the quadratic ($Estimate = 0.02$, $\beta = 0.02$, 95% CI [-0.14, 0.18], $p = .81$) and cubic contrasts ($Estimate = 0.03$, $\beta = 0.01$, 95% CI [-0.15, 0.16], $p = .93$) and their combined residual effect ($\Delta R^2 = .001$, $F(2,155) = 0.03$, $p = .97$) are not. These results support our hypothesis H₃ and indicate that participants' discriminatory behavior score follows an upward trend from the control condition, through the weakened hypocrisy condition, through the standard hypocrisy condition, and up to the reinforced hypocrisy condition.

Discussion

Consistent with our hypothesis (*GHI*), Study 2 replicated and extended Study 1 by showing that the stronger the social norms strength in hypocrisy procedure, the greater the hypocrisy effect. Nevertheless, at this stage, these results testing our hypothesis of social norms deviation have one limitation. Personal norms could be derived from social norms (Thøgersen, 2006) and the impact of social norms on behavior may be partly mediated by personal norms (e.g., Doran & Larsen, 2016; Thøgersen, 2014). As such, it is possible that the observed effects could be attributed to the reinforcement/weakening of personal norms, rather than social norms, not completely evacuating the explanation in terms of self-consistency (Aronson, 1999) and, more generally, in terms of personal norm deviation pathways (Stone & Cooper, 2001). In order to test whether the social norm deviation strength underlies the results of the first two studies, we conducted a third experiment to examine whether our operationalizations of the normative salience step of IHP (i.e., reinforcing/weakening social norms and/or normative advocacy) impact the strength of social and personal normative beliefs.

Study 3

Study 3 aimed to check that our experimental manipulations of social norms (i.e., strengthening and weakening) were effective in modifying beliefs relating to social norms,

without simultaneously affecting those relating to personal norms. Using the protocols from Studies 1 and 2, we tested four versions of the normative salience step: reinforcing social norms, weakening social norms, implementing only normative advocacy, and activating an irrelevant norm before normative advocacy.

Method

Participants

We recruited 140 Caen university students to take part in this third study. After excluding 16 participants who did not fully complete the experimental protocol our final sample consisted of 124 students ($M = 19.4$, $SD = 1.28$; 81.4% female). This sample size, with an expected effect size of $f = .35$ for the contrast of interest, enabled us to achieve a satisfactory statistical power of 94% (see Supplement 1, Section 3.1).

Design and Procedure

Participants in the *normative advocacy only* condition ($n = 29$) were asked to perform the normative advocacy and then respond to the dependent variable questions, namely a scale measuring social and personal normative anti-discrimination beliefs. Before completing the normative advocacy measure, participants in the other three conditions were asked to complete the exercise (i.e., reading the results of a so-called survey) that aimed at reinforcing (*reinforcing social norms* condition, $n = 28$), weakening (*weakening social norms* condition, $n = 35$) anti-discriminatory social norms or activating a neutral norm (*irrelevant norm* condition, $n = 32$). The questionnaire measuring dependent variables was then completed.

Materials

Strengthening and weakening social norms.

The procedure and material were the same as Studies 1 and 2.

Normative Advocacy.

The procedure and material were the same as Studies 1 and 2.

Dependent measures.

To measure anti-discriminatory social and personal normative beliefs, the procedure and material were the same as Study 2, with the addition of a measure of social normative beliefs. Based on the work of Cialdini et al. (1991) and Cialdini & Trost (1998), social norm beliefs were measured using three descriptive norm items (e.g., “*I think that most students at the Caen University do not discriminate against FPFO*”) and three injunctive norm items (e.g., “*I think that most students at the Caen University disapprove of discrimination against FPFO*”). Thus, we used 12 items on a 7-point Likert scale (from 1 - *Strongly Disagree* to 7 - *Strongly Agree*), supplemented by 15 filler items.

Hypotheses and Data Analysis Plan

We posed two (alternative) hypotheses depending on the type of deviation pathway that participants in Studies 1 and 2 would have used. First, if the social norm deviation pathway underlies the hypocrisy effects in Studies 1 and 2, then our manipulations of the normative salience step should have impacted participants' social normative beliefs, such that the stronger (vs. weaker) social norms are, the higher (vs. lower) participants' social normative beliefs will be (H₁). In other words, participants' social normative beliefs will show a positive trend from the *weakening social norms* condition, through the *irrelevant norm* and *normative advocacy only* conditions, and up to the *reinforcing social norms* condition. However, if the personal norm deviation pathway underlies the hypocrisy effects in Studies 1 and 2, then our manipulation of the normative salience step should impact participants' personal normative beliefs (H₂). Concretely, personal normative beliefs will show a positive trend from the *weakening social norms* condition, through the *irrelevant norm* and *normative advocacy only* conditions, and up to the *reinforcing social norms* condition.

As for previous studies, we tested our hypothesis using contrast analyses. The contrast codes are detailed in the Supplement 1 (Section 3.3). C1 compares the *weakening social norms* condition to the *reinforcing social norms* condition. C2 compares the *irrelevant norm* condition to the *normative advocacy only* condition. C3 compares the *weakening social norms* and *reinforcing social norms* conditions to the *irrelevant norm* and *normative advocacy only* conditions. To support H₁ or H₂, C1 should be positive and significant, while the two other residual contrasts C2 and C3 and their combined residual effect should not be significant. H₁ and H₂ were tested using linear regression models on social normative beliefs and personal normative beliefs, respectively, with the three contrasts as predictors.

Results

Preliminary analyses (see Supplement 2, Section 3.1) revealed that participants' age significantly differs according to the experimental conditions. In the model testing our hypothesis, we have therefore included participants' age as a covariate. In addition, a confirmatory factorial analysis was performed on the 12 items measuring normative beliefs and confirmed the bi-dimensionality of our measurement scale, namely social normative beliefs (6 items McDonald's $\omega = .87$) and personal normative beliefs (6 items, McDonald's $\omega = .82$).

Descriptive data are displayed in **Table 3**.

Table 3.

Social and personal normative belief scores according to the four experimental conditions (Study 3).

Condition	Social normative beliefs	Personal normative beliefs
Weakening social norms	4.91 (1.01)	6.47 (1.01)
Irrelevant norm	5.35 (0.81)	6.47 (0.85)
Normative advocacy only	5.14 (0.86)	6.72 (0.43)
Reinforcing social norms	5.64 (0.91)	6.59 (0.67)

Note. $N = 124$. Data indicate the mean (standard deviation) of items on a 7-point Likert scale.

Effect on social normative beliefs

The linear regression model on social normative beliefs is globally significant, $F(4,119) = 2.90$, $p = .025$, $R^2 = .089$. More specifically, C1 is significant (*Estimate* = 0.35, $\beta = 0.27$, 95% CI [0.09, 0.44], $p = .003$) while C2 (*Estimate* = -0.12, $\beta = -0.09$, 95% CI [-0.26, 0.9], $p = .33$) and C3 (*Estimate* = 0.01, $\beta = 0.01$, 95% CI [-0.16, 0.19], $p = .87$) and their combined residual effect ($\Delta R^2 = .007$, $F(2,119) = 0.491$, $p = .61$) are not. Furthermore, participants' age is not significantly associated with social normative beliefs (*Estimate* = 0.03, $\beta = 0.08$, 95% CI [-0.10, 0.25], $p = .39$). These results support H₁, indicating that participants' social normative beliefs follow an upward trend from the *weakening social norms* condition, through the *irrelevant norm* and *normative advocacy only* conditions, and up to the *reinforcing social norms* condition (see **Table 3**).

Effect on personal normative beliefs

The linear regression model on personal normative beliefs is not globally significant at the α threefold of .05, $F(4,119) = 0.570$, $p = .66$, $R^2 = .018$. More specifically, neither C1 (*Estimate* = 0.01, $\beta = 0.01$, 95% CI [-0.18, 0.19], $p = .95$), C2 (*Estimate* = 0.07, $\beta = 0.06$, 95% CI [-0.12, 0.24], $p = .42$) nor C3 (*Estimate* = 0.02, $\beta = 0.02$, 95% CI [-0.16, 0.20], $p = .82$) are significant.

Furthermore, participants' age is not significantly associated with personal normative beliefs ($Estimate = 0.04$, $\beta = 0.11$, 95% CI [-0.07, 0.30], $p = .22$). These results do not support H₂.

Discussion

The results of Study 3 supported H₁, showing that our manipulations of normative salience in Studies 1 and 2 effectively influenced participants' social normative beliefs rather than personal normative beliefs. Collectively, the findings from the first three studies support *GHI*, indicating that social norm strength contributes to the induced-hypocrisy effect and providing evidence for the social norm deviation approach.

However, according to normative focus theory (Cialdini et al., 1991; Kallgren et al., 2000), the increased hypocrisy effect observed in our reinforced hypocrisy conditions could be driven solely from the activation of additional normative information. If this is the case, the heightened behavioral engagement observed in these conditions may not necessarily reflect a stronger hypocrisy effect per se, but rather a combination of hypocrisy and normative salience effects. To disentangle these possibilities, we need to determine whether the enhanced behavioral engagement in reinforced hypocrisy conditions is genuinely due to a stronger induced-hypocrisy experience.

A critical test of this hypothesis involves examining two key predictions: (i) whether participants in the hypocrisy condition perceive their past transgressions as deviating from social norms - and whether this perception is amplified when the norm is reinforced, and (ii) whether this perceived deviation from social norms mediates the effect of induced-hypocrisy on prosocial behavioral engagement. This is precisely what the social norm deviation approach predicts: the induced-hypocrisy effect arises from individuals' perceived gap between their past behavior and prevailing social norms (Priolo et al., 2016).

Therefore, to test this mediating effect of perceived social norm deviation (*GH2*), we conducted two additional studies using an experimental-chain design recommended for testing mediation hypotheses (Pirlott & MacKinnon, 2016). Study 4a (pre-registered, <https://osf.io/h8fr7>) examined whether hypocrisy induction increases perceived social norm deviation ($X \rightarrow M$). Study 4b (pre-registered, <https://osf.io/tkxgs>) then directly manipulated perceived social norm deviation - independently of hypocrisy induction - to assess whether it produces the same effect on prosocial engagement as hypocrisy induction ($M \rightarrow Y$).

Study 4a

Study 4a (<https://osf.io/h8fr7>) aimed to test whether IHP affects participants' perceived social norm deviation.

Method

Participants

We recruited 193 Caen university students to take part in this fourth online experiment. Twenty-eight participants were withdrawn because they did not fully complete the experimental protocol (9 did not fill in the socio-demographic data and 19 did not recall any transgressions in the hypocrisy conditions). Our final sample consisted of 165 students ($M = 21.5$, $SD = 4.58$; 73.9% female). This sample size, with an expected effect size of $OR = 1.72$ for the contrast of interest, allowed us to obtain a satisfactory statistical power above 80% (see Supplement 1, Section 4.1).

Design and Procedure

Due to the COVID-19 pandemic, we were compelled to conduct this experiment online rather than face-to-face. Students at the Caen University were asked to participate in a survey via their institutional e-mail address. The university's Information Systems Department (DSI)

developed a server-to-server security system guaranteeing complete anonymity for respondents. The study was carried out on the LimeSurvey® platform.

Participants in the *control condition* directly completed the dependent variables, namely measures of perceived social norm deviation and engagement to volunteer in a real discrimination prevention association (i.e., AFEV – “Association de la fondation étudiante pour la ville”). Participants in the *reinforced hypocrisy condition* completed an exercise that aimed at reinforcing the anti-discriminatory social norms, followed by normative advocacy and recall of past discriminatory behaviors. Participants in the *standard hypocrisy condition* performed only the normative advocacy and the transgressions-recall.

Materials

Strengthening social norms.

The procedure and material were the same as previous studies.

Normative Advocacy.

The procedure and material were the same as previous studies.

Recall of past transgressive behaviors.

The procedure and material were the same as Study 2.

Dependent measures.

First, we measured participants' perceptions of deviating from the anti-discriminatory social norm using a four-item scale (see Supplement 1, Section 4.2). This scale is composed of two items (e.g., “To what extent can your past behaviors toward FPFO be considered a violation of the values and expectations of Caen University students?”) measuring perceived norm violation (Crane & Platow, 2010), one item measuring group norm adequacy (“To what extent do your past behaviors towards FPFO match the values and expectations of Caen University students?”, reversed item), and one item measuring group prototypicality (i.e., “To what extent

do your past behaviors toward FPFO make you representative of Caen University students?”, reversed item, (Steffens et al., 2015). Response modalities were a 7-point scale ranging from 1: *Not at all* to 7: *Absolutely*. High scores on the items indicated a greater perception of deviating from the social norm.

Second, given the paucity of studies that have carried out the induced-hypocrisy procedure online (Dossett, 2009; Gamma et al., 2018; Odou et al., 2019: study 3; Thorman et al., 2020) and the concerns of some researchers about its effectiveness (McGrath, 2018), we also measured the hypocrisy effect in the conventional way via the acceptance of a prosocial request (i.e., becoming a volunteer in a prevention association). Participants' engagement to the City Student Foundation Association (Association de la Fondation Etudiante pour la Ville - AFEV) was measured (see Supplement 1, Section 4.3). Participants were asked whether they wished to “join the association and access the membership form” or “refuse to join the association”. The association was presented as fighting discrimination and inequality, and participants were informed that their engagement to the association consisted in giving two hours a week of their time to a disadvantaged young person for one school year.

Hypotheses and Data Analysis Plan

Our main aim relates to the effect of our experimental conditions on perceived social norm deviation. We expected participants' perceived deviation from the social norm to show a positive trend from the control condition, through the standard hypocrisy condition, and up to the reinforced hypocrisy condition (H_1). A secondary aim was to verify that our online hypocrisy procedure causes a hypocrisy effect. We expected participants' engagement to volunteering to show a positive trend from the control condition, through the standard hypocrisy condition, and up to the reinforced hypocrisy condition (H_2).

As for previous studies, we tested our hypotheses using polynomial contrasts (see Supplement 1, Section 4.4). The linear contrast (Control < Standard hypocrisy < Reinforced hypocrisy) was our contrast of interest, and the quadratic contrast was the residual. With the two contrasts as predictors, H₁ was tested using a linear regression model on perceived deviation from social norm and H₂ was tested using a binomial logistic regression model on engagement to volunteering.

Results

Preliminary analyses (see Supplement 2, Section 4.1) ensured the random distribution of participants across experimental conditions based on socio-demographic variables (gender and age), the absence of a significant difference between the two hypocrisy conditions on the number of transgressions recalled and the internal validity of the measure of perceived deviation from the social norm. As the factorial analyses indicated two dimensions, we calculated our perceived deviation from social norm score using the first dimension including the two items measuring perceived norm violation ($r = .87, p < .001$). The results of the analysis on the second-dimension composite score, including two items measuring group norm adequacy and group prototypicality) are available in supplement 2 (Section 4.2).

Descriptive data are displayed in **Table 4**.

Table 4.

Score of perceived deviation from social norm and engagement to volunteer among the three experimental conditions.

	Perceived deviation from the social norm	Engagement to volunteer (target-request acceptance)
Control	1.21 (0.393)	1.82 (1/55)
Standard hypocrisy	1.70 (1.32)	7.27 (4/55)
Reinforced hypocrisy	2.37 (1.56)	14.5 (8/55)

Note. $N = 165$. For perceived deviation from group norms, data indicate the average score for 2 items on a 7-point scale (*SD*). For engagement to volunteer, data indicate the percentage of engagement in a discrimination prevention association (number of acceptances/number of participants).

Effect on perceived deviation from social norm

The linear regression model on the perceived deviation from social norm score is globally significant, $F(2,162) = 13.1$, $p < .001$, $R^2 = .139$. More specifically, the linear contrast is significant (*Estimate* = 0.58, $\beta = 0.37$, 95% CI [0.23, 0.52], $p < .001$) while the quadratic contrast (*Estimate* = 0.03, $\beta = 0.04$, 95% CI [-0.11, 0.18], $p = .61$) is not. These results support our hypothesis H_1 and indicate that participants' perceived deviation from social norm follows an upward trend from the control condition, through the standard hypocrisy condition, and up to the reinforced hypocrisy condition.

Effect on engagement to volunteer

The very high asymmetry in the distribution of our binary dependent variable (7.9% acceptance, 92.1% refusal) is known to cause a bias in classical logistic regression. To reduce this bias (Heinze et al., 2023; Heinze & Schemper, 2002), we used the intercept-corrected Firth logistic regression (*flic* from the *logistf* R package, see Supplement 1, Section 4.7 for the R script). This intercept correction provides more accurate probability predictions, but does not alter the effects or their statistical significance compared to the original Firth model (Puhr et

al., 2017). The results obtained without using this intercept-corrected are available in the supplement 2 (Section 4.3).

The logistic regression model on engagement to volunteer is globally not significant at the α threefold of .05, $\chi^2(2) = 5.93$, $p = .052$, $R^2_L = .07$. However, results indicate that the linear contrast is significant (*Estimate* = 0.94, $\chi^2_{wald}(1) = 5.88$, $p = .015$, *OR* = 2.56, 95% CI [1.18, 7.92]) while the quadratic contrast (*Estimate* = -0.07, $\chi^2_{wald}(1) = -0.11$, $p = .74$, *OR* = 0.92, 95% CI [0.58, 1.46]) is not. These results support our hypothesis H₂ and indicate that the proportion of participants who accepted to take part in the anti-discrimination association actions follows an upward trend from the control condition, through the standard hypocrisy condition, and up to the reinforced hypocrisy condition.

Discussion

The results of Study 4a indicate that individuals perceive a gap with social norms in hypocrisy situations, more so when their normative beliefs are stronger. This finding suggests that the stronger effects observed on prosocial engagement in this and previous studies stem from a reinforced hypocrisy situation rather than the combination of two distinct effects - namely, induced hypocrisy and normative salience. It supports our general hypothesis that the hypocrisy effect is driven by the process of social norm deviation.

Study 4a thus provides the first step in testing the mediation hypothesis: induced hypocrisy leads to the perception of social norm deviation. Study 4b extends this investigation by examining the second step of the mediation process. If perceived social norm deviation mediates the hypocrisy effect, then manipulating this perception (low vs. high deviation) through a different approach than induced hypocrisy should produce similar behavioral outcomes, such as greater acceptance of prosocial behavior.

Study 4b

Following Study 4a results showing that induced hypocrisy affects participants' perceived social norm deviation, Study 4b (<https://osf.io/tkxgs>) aimed to test whether social norm deviation increases the acceptance of prosocial behavior, similar to an induced-hypocrisy procedure. In addition, research on norm deviance suggests that deviant individuals diminish their opportunities to form and maintain gratifying relationships with ingroup members (Jetten & Hornsey, 2014). Consequently, we also investigated the impact of social norm deviation on the perceived group affiliation from which the social norm originates.

Method

Participants

A total of 173 participants took part in this online study. After attrition ($n = 8$) due to missing socio-demographic variables, our final sample consisted of 165 students from the University of Caen ($M = 20.2$, $SD = 3.20$; 72.7% female). This sample size, with an expected effect size of $OR = 1.72$ for the contrast of interest, enables us to achieve a satisfactory a priori statistical power of 80% (see Supplement 1, Section 5.1).

Design and Procedure

Participants were randomly assigned to one of three experimental conditions: control, low deviation-from-norm, high deviation-from-norm. Participants in the *control condition* were first asked to give their opinion on possible changes in university policy by answering 15 questions. Finally, they were asked to fill in the dependent variables, namely the measures of perceived in-group affiliation and of engagement to volunteering for a discrimination prevention association. Participants in the *low* and *high deviation-from-norm conditions* followed the same procedure as those in the control condition, with one exception. After answering the 15

academic policy questions, we introduced our group norm deviation manipulation in which we indicated that their answers differed either weakly or strongly from those of a representative sample of other students.

Materials

Manipulation of deviation from in-group norm

We drew on Heerdink et al.' procedure (2015) to manipulate participants' deviation from an in-group norm. Participants were asked to give their opinion on some possible changes in their university policies by answering 15 yes/no questions (e.g., "The Caen University should also provide classes on Saturdays"). Once the 15 questions had been answered (see Supplement 1, Section 5.2), participants were told that 4,500 Caen University students had already completed the questionnaire. To activate a strong deviation from in-group norm, we informed participants that a majority of Caen University students disagreed with them. To activate a weak deviation from in-group norm, we informed participants that a minority of Caen University students disagreed with them.

Dependent measures.

On the one hand, two dependent variables were measured. First, we measured participants' perceptions of affiliation with the Caen University students group using five items (e.g., "I feel out of place among Caen University students"; adapted from the Psychological Need States Scale, (Bhavsar et al., 2020) on a 7-point scale ranging from 1 - *Not at all* to 7 - *Definitely* (see Supplement 1, Section 5.3). The higher the score, the more participants perceive themselves to be affiliated with the group. Second, as for Study 4, we measured participants' engagement to the discrimination prevention association (i.e., AFEV). The higher the participants' acceptance rate, the greater the hypocrisy effect.

On the other hand, we checked that our manipulation of the deviation from in-group norm was effective by asking participants to answer one item (i.e., “Having read this result, would you say that the majority of Caen University students share your opinions?”) on a 7-point scale (from 1: *Not at all* to 7: *Absolutely*).

Hypotheses and Data Analysis Plan

Our main aim relates to the effect of our experimental conditions on behavioral engagement in prosocial way. We expected participants' engagement to volunteer in prevention association to show a positive trend from the control condition, through the low deviation-from-norm condition, and up to the high deviation-from-norm condition (H₁). A secondary aim was to test whether our experimental conditions impact group affiliation also. We expected participants' perceived group affiliation to show a negative trend from the control condition, through the low deviation-from-norm condition, and up to the high deviation-from-norm condition (H₂).

As for previous studies, we tested our hypotheses using polynomial contrasts (see Supplement 1, Section 5.4). The linear contrast (Control < Low deviation-from-norm < High deviation-from-norm) was our contrast of interest, and the quadratic contrast was the residual. With the two contrasts as predictors, H₁ was tested using a binomial logistic regression model on engagement to volunteer and H₂ was tested using a linear regression model on perceived group affiliation.

Results

Preliminary analyses (see Supplement 2, Section 5.1) ensured the random distribution of participants across experimental conditions based on socio-demographic variables (gender and age), the effectiveness of our experimental manipulation on participants' perceived deviation from in-group norm and the one-dimensionality of our measurement of perceived group affiliation (5 items, Cronbach's $\alpha = .85$).

Descriptive data are displayed in **Table 5**.

Table 5.

Perceived affiliation to in-group and engagement to volunteer in prevention association among the three experimental conditions.

	Engagement to volunteer (Acceptance of prosocial behavior)	Perceived group affiliation
Control	1.82 (1/55)	5.24 (1.34)
Low deviation-from-norm	7.27 (4/55)	5.13 (1.26)
High deviation-from-norm	12.7 (7/55)	4.63 (1.50)

Note. $N = 165$. For the perceived group affiliation, data indicate the mean score for 5 items on a 7-point Likert scale (*SD*). For engagement to volunteer, data indicate percentage of engagement in a discrimination prevention association (amount of acceptances/number of participants).

Effect on engagement to volunteer

As for Study 4a, we used the intercept-corrected Firth logistic regression (flic from the *logistf* R package, see Supplement 1, Section 5.7 for the R script) to analyze our binary dependent variable, due to the high asymmetry in its distribution (7.3% acceptance, 92.7% refusal). The results obtained without using this intercept-corrected are available in the Supplement 2 (Section 5.2).

The logistic regression model on engagement to volunteer is globally not significant at the α threefold of .05, $\chi^2(2) = 4.72, p = .09, R^2_L = .06$. However, the linear contrast is significant (*Estimate* = 0.86, $\chi^2_{wald}(1) = 4.72, p = .03, OR = 2.37, 95\% CI [1.08, 7.39]$) while the quadratic contrast (*Estimate* = -0.10, $\chi^2_{wald}(1) = -0.18, p = .66, OR = 0.91, 95\% CI [0.57, 1.43]$) is not. These results support our hypothesis H_1 and indicate that the proportion of participants who accepted to take part in the anti-discrimination association follows an upward trend from the control condition, through the low deviation-from-norm condition, and up to the high deviation-from-norm condition.

Effect on perceived group affiliation

The linear regression model on the perceived group affiliation is close to being significant, $F(2,162) = 3.06$, $p = .05$, $R^2 = .036$. More specifically, the linear contrast is significant ($Estimate = -0.32$, $\beta = -0.18$, 95% CI [-0.33, -0.03], $p = .02$) while the quadratic contrast ($Estimate = -0.07$, $\beta = -0.07$, 95% CI [-0.22, 0.08], $p = .38$) is not. These results support our hypothesis H₂ and indicate that participants' perceived group affiliation follows a downward trend from the control condition, through the low deviation-from-norm condition to the high deviation-from-norm condition.

Discussion

Having established in Study 4a that induced hypocrisy increases the perception of social norm deviation (the first step in testing mediation hypotheses), Study 4b investigated whether deviation from social norms (manipulated independently of induced hypocrisy) produces the same behavioral outcomes as the hypocrisy procedure. The findings demonstrate that manipulating perceived deviation from the in-group norm, through normative feedback, impacts individuals' intentions to act in prosocial way. Thus, findings of Studies 4a and 4b support our GH₂, suggesting that people are more likely to engage in prosocial behavior in hypocrisy situations because they perceive their past transgressions as deviating from social norms. These results, combined with those of the first three studies, align with our overall hypothesis that a process of deviation from social norms could underlie the hypocrisy effect.

These research findings provide strong support for our theoretical framework. However, in the present research, our primary analyses relied on polynomial contrasts to test our hypotheses, capturing the hypothesized linear effects of social norm strength on hypocrisy effects. While informative and relevant (Judd et al., 2017), this approach does not directly compare standard and reinforced hypocrisy conditions, the crux of our theoretical proposition. To address this

limitation and further strengthen our conclusions, we conducted an integrative analysis that combined data from studies specifically designed to test the key contrasts of interest: comparisons between the control and standard hypocrisy conditions, and between the standard and reinforced hypocrisy conditions.

A common concern with small-scale internal meta-analyses or integrative analyses is the potential influence of unreported or unpublished studies (Vosgerau et al., 2019). To mitigate this risk, we included in our integrative analysis all relevant studies conducted as part of our broader investigation on this research question. This comprehensive inclusion reduces concerns about potential file-drawer effects and ensures that our conclusions are not biased by selective reporting. Specifically, we combined data from the studies presented in this paper (Studies 1, 2, and 4a) with two additional datasets following the same protocol, which included control, standard hypocrisy, and reinforced hypocrisy conditions. One of these two additional studies has been published (see, Mauduy et al., 2022) while the other remains unpublished. This combined dataset ($N > 600$) enables us to test the critical contrasts of interest with increased and satisfactory statistical power. As such, this integrative analysis provides a robust and unbiased test of our central hypothesis regarding the role of social norm strength in modulating the hypocrisy effect.

Integrative analysis of the effect of social norms strength

Method

To provide a more comprehensive test of our hypothesis regarding the role of social norm strength in the hypocrisy effect, we conducted an integrative analysis combining data from five studies: Study 1, Study 2, Study 4a, and two additional studies (Sup 1 and Sup 2). This dataset includes a total of 602 participants distributed across three experimental conditions: a control condition ($n = 213$), a standard hypocrisy condition ($n = 206$), and a reinforced

hypocrisy condition ($n = 183$). Participants in each study followed the same procedure, in the context of discrimination prevention, as for the control, standard hypocrisy and reinforced hypocrisy conditions described in this research. The key dependent variable was participants' acceptance of a prosocial behavior.

To analyze these data, we conducted a generalized linear mixed model, treating *study* as a random intercept to account for between-study variation and modeling *condition* as a fixed effect with three levels (control, standard hypocrisy, reinforced hypocrisy). Additionally, we included random slopes for the condition, allowing the effect of condition to vary across studies. This approach enables us to test whether the effect of hypocrisy conditions is consistent or differs from study to study. The dependent variable was the binary measure of prosocial behavior acceptance (0 = no acceptance, 1 = acceptance). We used planned comparisons to compare control vs. standard hypocrisy conditions and standard hypocrisy vs. reinforced hypocrisy conditions.

Results and Discussion

The omnibus test for the effect of condition was significant, $\chi^2(2) = 26.4, p < .001, R^2 = .09$. The fixed-effect estimates (see **Table 6**) revealed that participants in the standard hypocrisy condition ($M = 23.1\%, SE = 0.072$) were significantly more likely to accept the prosocial behavior compared to the control condition ($M = 10.6\%, SE = 0.047$). Similarly, the reinforced hypocrisy condition ($M = 36.4\%, SE = 0.087$) led to a significantly higher probability of prosocial behavior acceptance compared to the standard hypocrisy condition. The random effects component indicated that there was some variation between studies (random intercept, $ICC = 0.176$), but the variability in condition effects (random slopes) was small and non-significant ($LRT = 4.50, p = .48$). This suggests that while some studies may exhibit slightly

stronger or weaker effects of hypocrisy conditions, the overall pattern remains stable across studies.

Table 6.

Results of the generalized linear mixed effects model on prosocial behavior acceptance (N = 602).

Effect	<i>Estimate</i>	<i>SE</i>	<i>OR</i>	95% CI [Lower, Upper]	<i>p</i>
Fixed Effects					
Intercept	-1.30	0.393	0.273	[0.127, 0.590]	<.001
Condition 1 (Standard Hypocrisy vs. Control)	0.935	0.315	2.547	[1.374, 4.724]	.003
Condition 2 (Reinforced vs. Standard Hypocrisy)	0.642	0.246	1.900	[1.173, 3.079]	.009
Random Effects (5 studies)					
Intercept (Between-study variation)	0.701				
Condition 1 - slopes variance	0.002				
Condition 2 - slopes variance	0.020				
ICC	0.176				

By integrating data from multiple studies using a mixed-effects model, this analysis supports our hypotheses by showing that both the standard and reinforced hypocrisy conditions lead to greater acceptance of prosocial behavior compared to the control condition. Moreover, the reinforced hypocrisy condition amplifies this effect, supporting the idea that the strength of perceived norm deviation modulates the hypocrisy effect.

General discussion

Our five-study experimental program ($N_{Total} = 774$) investigated the social norm deviation pathway as an explanatory mechanism for the induced-hypocrisy effect, yielding two key findings. First, the strength of social norms positively influences hypocrisy effects. To

further support this, we conducted an integrative analysis of Studies 1, 2, and 4a, along with two additional datasets ($N = 602$). Consistent with Priolo et al.'s meta-analysis (2019), our experimental studies replicated standard hypocrisy effects relative to a control condition ($OR = 2.55$ equivalent to $r = .25$, medium effect). Furthermore, aligning with Mauduy et al.'s meta-analysis (2023), reinforcing social norms in hypocrisy conditions (i.e., reinforced hypocrisy conditions) significantly increased behavioral engagement compared to standard hypocrisy conditions ($OR = 1.9$ equivalent to $r = .17$, small effect). Second, our experimental-chain-design studies (Studies 4a and 4b) demonstrated that perceived social norm deviation mediates the hypocrisy effect. These findings support the idea that social norm deviation is a key process underlying the hypocrisy effect and offering a significant theoretical contribution to the induced-hypocrisy paradigm. This calls for further theoretical discussions on the mechanisms underlying hypocrisy and their broader implications for cognitive dissonance theory.

While our findings support the social norm deviation approach (Liégeois et al., 2017), which provides a compelling explanation for this social norm deviation process by suggesting that it arises from the perception of an inconsistency between social norms and one's own transgressive behavior, alternative theoretical explanations within cognitive dissonance theory may also be relevant in understanding the mechanisms at play.

Notably, the social norm deviation approach assumes that the hypocrisy effect is not primarily driven by self-threat but it does not exclude the possibility that perceiving oneself as deviating from social norms may still impact the self. This aligns with self-consistency theory (Aronson, 1999), which suggests that individuals strive to maintain a self-image congruent with their values and social roles. It is possible that perceiving oneself as a norm-deviant person could induce discomfort related to self-consistency concerns, thereby reinforcing behavioral change. Furthermore, the perceived deviation from social norms might heighten awareness of the potential negative consequences of one's actions, aligning with the New Look Model

(Cooper & Fazio, 1984), which emphasizes the role of aversive consequences in cognitive dissonance processes. Similarly, the deviation from the norm approach is compatible with the Action-Based Model (Harmon-Jones et al., 2015). The latter posits that dissonance is resolved by enhancing perceptions of effective and goal-consistent action. Indeed, perceiving deviation from the norm may lead a person to adopt normative behavior, creating the impression of behaving appropriately and thus being more effective.

A crucial step in disentangling these theoretical explanations - and, by extension, better understanding the hypocrisy effect - is identifying the psychological process underlying the social norm deviation pathway. According to the New Look Model, this process would involve aversive consequences; in the Action-Based Model, it would be a motivation for accuracy, correct action, and effectiveness. However, neither the social norm deviation approach nor the Self-Standards Model (SSM) explicitly defines the process through which norm deviation leads to behavioral change. To address this, it may be useful to extend our theoretical framework beyond cognitive dissonance theory and explore insights from research on norm deviance.

Research on optimal distinctiveness theory (Brewer, 2012) and norm deviance (Jetten & Hornsey, 2014) suggests that deviant individuals risk diminishing their opportunities to form and maintain gratifying relationships with ingroup members, prompting them to seek ways to restore their affiliation. Consistently, our Study 4b results indicated that social norm deviation (manipulated through normative feedback) impacts perceived affiliation with the ingroup, leading to increased intentions to engage in prosocial behavior. Affiliation is a fundamental psychological need, and individuals are motivated to restore it when threatened (Baumeister & Leary, 1995). Thus, theoretical models and findings suggest that affiliation motivation could be the underlying mechanism driving behavior change in hypocrisy situations where participants perceive themselves as deviating from social norms. If affiliation concerns underlie the

hypocrisy effect, this has significant implications for both the understanding of induced hypocrisy and cognitive dissonance theory more broadly.

First, this perspective provides insight into how the social norm deviation pathway in induced hypocrisy situations might threaten the self, an aspect not explicitly considered by the social norm deviation approach (Liégeois et al., 2017) or the SSM (Stone & Fernandez, 2008). Aronson's self-consistency theory characterizes the self as independent (Markus & Kitayama, 1991), emphasizing autonomy, uniqueness, and resistance to conformity (Aronson, 1999). However, a person may also perceive themselves as interdependent, driven to seek and maintain social bonds through behavior aligned with social norms. Yet, whereas maintaining a positive self-image primarily motivates adherence to personal norms (Cialdini & Trost, 1998; Thøgersen, 2006), affiliation motives drive adherence to social norms (Baumeister & Leary, 1995; Cialdini & Goldstein, 2004). Thus, induced hypocrisy may primarily target the interdependent self, which is affiliated and socially connected, rather than the independent self. Supporting this idea, Yoo and Jin (2020) found that people with an interdependent self-construal are more susceptible to the hypocrisy effect. If hypocrisy operates as a signal that one's standing within the group is at risk, then behavioral change may be driven by a motivation to restore social belonging rather than simply resolving cognitive inconsistency.

Second, if induced hypocrisy elicits concerns related to norm deviance (e.g., fear of social rejection or disapproval), then the role of aversive consequences in hypocrisy-induced behavioral change warrants further investigation. However, while the New Look Model highlights aversive consequences, it will remain limited in its explanatory scope, as it does not account for the role of the self in the cognitive dissonance process (Stone & Cooper, 2001), and aversive consequences are not a necessary condition for generating cognitive dissonance effects (Harmon-Jones et al., 1996).

Third, if affiliation motivation drives dissonance regulation in induced hypocrisy situations, the Action-Based Model - which emphasizes motivation for accuracy, correct action, and effectiveness - may not be the most suitable explanatory framework for the hypocrisy effect. Furthermore, this could challenge the exclusivity of the Action-Based Model's assumption that all dissonance-related behavior change is driven by a motivation for effective action, calling for a reassessment of its scope in explaining all forms of cognitive dissonance.

Ultimately, it may be worth critically examining whether the induced-hypocrisy paradigm genuinely reflects a cognitive dissonance process. If its effects are primarily driven by perceived social norm deviation, alternative theoretical frameworks that do not rely on cognitive dissonance – such as optimal distinctiveness theory (Brewer, 2012) or deviance regulation theory (Blanton & Christie, 2003) – may provide equally valid explanations. This suggests that integrating induced hypocrisy into a broader theoretical landscape could enhance our understanding of its mechanisms. More generally, theoretical models derived from cognitive dissonance research, and dissonance processes themselves, could benefit from being informed by and examined through the lens of other theoretical frameworks beyond the field of cognitive dissonance. For instance, theories related to fundamental psychological needs (Vansteenkiste et al., 2020), as suggested by other researchers (Mauduy et al., 2024), could offer valuable insights into the underlying motivations at play. By integrating perspectives from these broader theoretical domains, future research could refine and expand our comprehension of both the hypocrisy effect and cognitive dissonance processes as a whole.

Limitations and Conclusion

This research has three main limitations. First, we used a single method to reinforce social norms by providing normative information. Testing our hypotheses with alternative approaches to influence normative perceptions would enhance the robustness of our findings. Second, we examined the social norm deviation process solely by manipulating the strength of

social norms. While this was our primary focus due to limited research on social norms in the induced-hypocrisy paradigm, future studies should investigate whether this process also occurs when manipulating transgression salience while keeping normative salience constant. Third, we examined social norm deviation in the context of discrimination. One could argue that transgressions involving discriminatory behavior inherently threaten affiliation needs, which we hypothesize as the key driver of the social norm deviation process. However, even if transgressions do elicit affiliation concerns, this alone would not fully account for our findings, as transgression recall remained constant across all hypocrisy conditions in our studies. To further explore the role of affiliation threat in the hypocrisy effect, future research should replicate these findings in domains where transgressive behaviors do not directly challenge affiliation needs.

Despite these limitations, this research provides strong evidence for the role of social norm deviation in driving the hypocrisy effect. It demonstrates that reinforcing social norms enhances behavioral change following a hypocrisy procedure and suggests that affiliation motivation may be a key mechanism in this process. More broadly, induced hypocrisy has traditionally been used to promote prosocial behaviors (e.g., sunscreen and condom use) rather than to prevent problematic behaviors like discrimination. It has also been applied in contexts where recalling transgressions is relatively straightforward (e.g., recycling, sunscreen use), whereas transgression recall is more challenging in cases like school bullying (Mauduy, Bagneux, et al., 2023) and discrimination (Son Hing et al., 2002). Thus, this research also expands the applicability of induced hypocrisy to anti-social behaviors, demonstrating that effective hypocrisy interventions can prevent behaviors that are difficult to recall. In this context, emphasizing social norms is crucial to overcoming the limitations of transgression recall. More broadly, this research refines recommendations for implementing the induced-

hypocrisy paradigm (Mauduy, Priolo, et al., 2023), demonstrating that it is most effective when social normative beliefs are reinforced.

References

- Abdi, H., & Williams, L. J. (2010). Contrast Analysis; How to express a research hypothesis as a contrast. *Encyclopedia of Research Design*, 1–14.
- Amodio, D. M., Devine, P. G., & Harmon-Jones, E. (2007). A Dynamic Model of Guilt. *Psychological Science*, 18(6), 524–530. <https://doi.org/10.1111/j.1467-9280.2007.01933.x>
- Anier, N., Badea, C., Berthon, M., & Guimond, S. (2018). Perceived acculturation preferences of minority groups and intergroup discrimination: When culture-specific intergroup norms matter. *Journal of Applied Social Psychology*, 48(9), 506–518. <https://doi.org/10.1111/jasp.12530>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3–25. <https://doi.org/10.1037/amp0000191>
- Aronson, E. (1968). Dissonance theory: Progress and problems. In R. P. Abelson, W. J. McGuire, T. M. Newcomb, M. J. Rosenberg, & P. H. Tannenbaum (Eds.), *Theories of cognitive consistency: A sourcebook* (pp. 5–27). Rand McNally.
- Aronson, E. (1999). Dissonance, hypocrisy, and the self-concept. In E. Harmon-Jones & J. Mills (Eds.), *Cognitive dissonance: Progress on a pivotal theory in social psychology* (pp. 103–126). American Psychological Association.
- Aronson, Elliot., Fried, Carrie., & Stone, Jeff. (1991). Overcoming denial and increasing the intention to use condoms through the induction of hypocrisy. *American Journal of Public Health*, 81(12), 1636–1638. <https://doi.org/10.2105/AJPH.81.12.1636>

- 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>
- Bhavsar, N., Bartholomew, K. J., Quedsted, E., Gucciardi, D. F., Thøgersen-Ntoumani, C., Reeve, J., Sarrazin, P., & Ntoumanis, N. (2020). Measuring psychological need states in sport: Theoretical considerations and a new measure. *Psychology of Sport and Exercise*, *47*, 101617.
<https://doi.org/10.1016/j.psychsport.2019.101617>
- Blanton, H., & Christie, C. (2003). Deviance Regulation: A Theory of Action and Identity. *Review of General Psychology*, *7*(2), 115–149. <https://doi.org/10.1037/1089-2680.7.2.115>
- Bodenhausen, G. V., & Richeson, J. A. (2010). Prejudice, stereotyping, and discrimination. In R. F. Baumeister & E. J. Finkel (Eds.), *Advanced social psychology: The state of the science* (pp. 341–383). Oxford University Press.
- Brewer, M. B. (2012). Optimal distinctiveness theory: Its history and development. *Handbook of Theories of Social Psychology*, *2*, 81–98.
- Cialdini, R. B., & Goldstein, N. J. (2004). Social Influence: Compliance and Conformity. *Annual Review of Psychology*, *55*(1), 591–621.
<https://doi.org/10.1146/annurev.psych.55.090902.142015>
- Cialdini, R. B., Kallgren, C. A., & Reno, R. R. (1991). A Focus Theory of Normative Conduct: A Theoretical Refinement and Reevaluation of the Role of Norms in Human Behavior. In *Advances in Experimental Social Psychology* (Vol. 24, pp. 201–234). Elsevier.
[https://doi.org/10.1016/S0065-2601\(08\)60330-5](https://doi.org/10.1016/S0065-2601(08)60330-5)
- Cialdini, R. B., Reno, R. R., & Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, *58*(6), 1015–1026. <https://doi.org/10.1037/0022-3514.58.6.1015>
- Cialdini, R. B., & Trost, M. R. (1998). Social influence: Social norms, conformity and compliance. In D. T. Gilbert, S. T. Fiske, & G. Lindzey, *The handbook of social psychology* (Eds, pp. 151–192). McGraw-Hill.

- Cooper, J., & Fazio, R. H. (1984). A New Look at Dissonance Theory. In *Advances in experimental social psychology* (Vol. 17, pp. 229–266). [https://doi.org/10.1016/S0065-2601\(08\)60121-5](https://doi.org/10.1016/S0065-2601(08)60121-5)
- Crandall, C. S., Eshleman, A., & O'Brien, L. (2002). Social norms and the expression and suppression of prejudice: The struggle for internalization. *Journal of Personality and Social Psychology*, 82(3), 359–377. <https://doi.org/10.1037//0022-3514.82.3.359>
- Crane, M. F., & Platow, M. J. (2010). Deviance as adherence to injunctive group norms: The overlooked role of social identification in deviance. *British Journal of Social Psychology*, 49(4), 827–847. <https://doi.org/10.1348/014466609X481416>
- DARES. (2021). *Discrimination à l'embauche des personnes d'origine supposée maghrébine: Quels enseignements d'une grande étude par testing ?* <https://dares.travail-emploi.gouv.fr/publication/discrimination-lembauche-des-personnes-dorigine-supposee-maghrébine-quels-enseignements>
- Devine, P. G., Monteith, M. J., Zuwerink, J. R., & Elliot, A. J. (1991). Prejudice With and Without Compunction. *Journal of Personality and Social Psychology*, 60(6), 817–830. <https://doi.org/10.1037/0022-3514.60.6.817>
- Dickerson, C. A., Thibodeau, R., Aronson, E., & Miller, D. (1992). Using Cognitive Dissonance to Encourage Water Conservation. *Journal of Applied Social Psychology*, 22(11), 841–854. <https://doi.org/10.1111/j.1559-1816.1992.tb00928.x>
- Doran, R., & Larsen, S. (2016). The Relative Importance of Social and Personal Norms in Explaining Intentions to Choose Eco-Friendly Travel Options. *International Journal of Tourism Research*, 18(2), 159–166. <https://doi.org/10.1002/jtr.2042>
- Dossett, A. P. (2009). *Inducing Hypocrisy as a means of mass persuasion* [[Master's thesis, Ohio State University]]. http://rave.ohiolink.edu/etdc/view?acc_num=osu1249582594
- Fehr, J., & Sassenberg, K. (2010). Willing and able: How internal motivation and failure help to overcome prejudice. *Group Processes & Intergroup Relations*, 13(2), 167–181. <https://doi.org/10.1177/1368430209343116>
- Festinger, L. (1957). *A theory of cognitive dissonance*. Stanford University Press.

- Fointiat, V. (2004). "I know what I have to do, but ... " When hypocrisy leads to behavioral change. *Social Behavior and Personality*, 32(8), 741–746. <https://doi.org/10.2224/sbp.2004.32.8.741>
- Fried, C. B. (1998). Hypocrisy and identification with transgressions: A case of undetected dissonance. *Basic and Applied Social Psychology*, 20(2), 145–154. <https://doi.org/10.1207/15324839851036769>
- Gabarrot, F., Falomir-Pichastor, J. M., & Mugny, G. (2009). Being similar versus being equal: Intergroup similarity moderates the influence of in-group norms on discrimination and prejudice. *British Journal of Social Psychology*, 48(2), 253–273. <https://doi.org/10.1348/014466608X342943>
- Gamma, K., Mai, R., & Loock, M. (2018). The Double-Edged Sword of Ethical Nudges: Does Inducing Hypocrisy Help or Hinder the Adoption of Pro-environmental Behaviors? *Journal of Business Ethics*, 161(2), 351–373. <https://doi.org/10.1007/s10551-018-3930-2>
- Harmon-Jones, E., Brehm, J. W., Greenberg, J., Simon, L., & Nelson, D. E. (1996). Evidence That the Production of Aversive Consequences Is Not Necessary to Create Cognitive Dissonance. *Journal of Personality and Social Psychology*, 70(1), 5–16. <https://doi.org/10.1037/0022-3514.70.1.5>
- Harmon-Jones, E., Harmon-Jones, C., & Levy, N. (2015). An Action-Based Model of Cognitive-Dissonance Processes. *Current Directions in Psychological Science*, 24(3), 184–189. <https://doi.org/10.1177/0963721414566449>
- Heinze, G., Ploner, M., Jiricka, L., & Steiner, G. (2023). *logistf: Firth's Bias-Reduced Logistic Regression* (Version R package version 1.26.0) [Computer software]. <https://CRAN.R-project.org/package=logistf>
- Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, 21(16), 2409–2419. <https://doi.org/10.1002/sim.1047>
- Jetten, J., & Hornsey, M. J. (2014). Deviance and dissent in groups. *Annual Review of Psychology*, 65, 461–468. <https://doi.org/10.1146/annurev-psych-010213-115151>
- Judd, C. M., McClelland, G., & Ryan, C. S. (2017). *Data analysis: A model comparison approach to regression, ANOVA, and beyond*. Routledge.

- Kaas, L., & Manger, C. (2012). Ethnic Discrimination in Germany's Labour Market: A Field Experiment. *German Economic Review*, 13(1), 1–20. <https://doi.org/10.1111/j.1468-0475.2011.00538.x>
- Kallgren, C. A., Reno, R. R., & Cialdini, R. B. (2000). A Focus Theory of Normative Conduct: When Norms Do and Do not Affect Behavior. *Personality and Social Psychology Bulletin*, 26(8), 1002–1012. <https://doi.org/10.1177/01461672002610009>
- Langner, O., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., Hawk, S. T., & van Knippenberg, A. (2010). Presentation and validation of the Radboud Faces Database. *Cognition & Emotion*, 24(8), 1377–1388. <https://doi.org/10.1080/02699930903485076>
- Liégeois, A., Codou, O., Rubens, L., & Priolo, D. (2017). « Faites ce que je dis, pas ce que je fais »: Synthèse et perspectives du paradigme de l'hypocrisie induite. *Psychologie Française*, 62(2), 177–194. <https://doi.org/10.1016/j.psfr.2015.12.001>
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review*, 98(2), 224–253. <https://doi.org/10.1037/0033-295X.98.2.224>
- Mauduy, M., Bagneux, V., & Sénémeaud, C. (2023). Fostering victim-defending behaviors among school bullying witnesses: A longitudinal and experimental test of two new strategies for changing behavior. *Social Psychology of Education*, 26(1), 263–274. <https://doi.org/10.1007/s11218-022-09745-z>
- Mauduy, M., & Mange, J. (2023). Subjective Norm Rather Than Social Norm in the Induced-Hypocrisy Paradigm: A Test in the Context of School Bullying Victim Support. *International Review of Social Psychology*, 36(1), 14. <https://doi.org/10.5334/irsp.776>
- Mauduy, M., Priolo, D., Margas, N., & Sénémeaud, C. (2022). When Combining Injunctive and Descriptive Norms Strengthens the Hypocrisy Effect: A test in the Field of Discrimination. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2022.989599>
- Mauduy, M., Priolo, D., Margas, N., & Sénémeaud, C. (2023). Does the First Step of the Induced-Hypocrisy Paradigm Really Matter? An Initial Investigation Using a Meta-Analytic Approach. *Social Psychological and Personality Science*. <https://doi.org/10.1177/19485506231188164>

- Mauduy, M., Priolo, D., Margas, N., & Sénémeaud, C. (2024, juillet). *Psychological-need-based model of dissonance : Une nouvelle modélisation intégrative de l'implication des besoins psychologiques fondamentaux dans les processus de dissonance cognitive*. 15^{ème} Congrès Internationale de Psychologie Sociale (CIPS), Bruxelles, Belgique.
- McGrath, A. L. (2018). Encouraging ecological behaviour through induced hypocrisy and inconsistency: A commentary. *Journal of Environmental Psychology, 56*, 1–2.
<https://doi.org/10.1016/j.jenvp.2018.01.005>
- Michinov, N., Dambrun, M., Guimond, S., & Méot, A. (2005). Social dominance orientation, prejudice, and discrimination: A new computer-based method for studying discriminatory behaviors. *Behavior Research Methods, 37*(1), 91–98. <https://doi.org/10.3758/BF03206402>
- Monteith, M. J., Ashburn-Nardo, L., Voils, C. I., & Czopp, A. M. (2002). Putting the brakes on prejudice: On the development and operation of cues for control. *Journal of Personality and Social Psychology, 83*(5), 1029–1050. <https://doi.org/10.1037/0022-3514.83.5.1029>
- Monteith, M. J., Parker, L. R., & Burns, M. D. (2016). The self-regulation of prejudice. In T. D. Nelson (Ed.), *Handbook of prejudice, stereotyping, and discrimination* (pp. 409–432). Psychology Press.
- Oudou, P., Darke, P., & Voisin, D. (2019). Promoting pro-environmental behaviours through induced hypocrisy. *Recherche et Applications En Marketing (English Edition), 34*(1), 74–90.
<https://doi.org/10.1177/2051570718813848>
- Paluck, E. L., Porat, R., Clark, C. S., & Green, D. P. (2021). Prejudice Reduction: Progress and Challenges. *Annual Review of Psychology, 72*(1), 533–560. <https://doi.org/10.1146/annurev-psych-071620-030619>
- Peterson, A. A., Haynes, G. A., & Olson, J. M. (2008). Self-Esteem Differences in the Effects of Hypocrisy Induction on Behavioral Intentions in the Health Domain. *Journal of Personality, 76*(2), 305–322. <https://doi.org/10.1111/j.1467-6494.2007.00487.x>
- Pirlott, A. G., & MacKinnon, D. P. (2016). Design approaches to experimental mediation. *Journal of Experimental Social Psychology, 66*, 29–38. <https://doi.org/10.1016/j.jesp.2015.09.012>

- Plant, E. A., & Devine, P. G. (1998). Internal and external motivation to respond without prejudice. *Journal of Personality and Social Psychology*, 75(3), 811–832. <https://doi.org/10.1037//0022-3514.75.3.811>
- Priolo, D., & Liégeois, A. (2008). Prôner une norme et la transgresser: Le rôle des normes sociales dans le paradigme de l'hypocrisie induite. *Les Cahiers Internationaux de Psychologie Sociale, Numéro 79(3)*, 19. <https://doi.org/10.3917/cips.079.0019>
- Priolo, D., Milhabet, I., Codou, O., Fointiat, V., Lebarbenchon, E., & Gabarrot, F. (2016). Encouraging ecological behaviour through induced hypocrisy and inconsistency. *Journal of Environmental Psychology*, 47, 166–180. <https://doi.org/10.1016/j.jenvp.2016.06.001>
- Priolo, D., Pelt, A., Bauzel, R. S., Rubens, L., Voisin, D., & Fointiat, V. (2019). Three Decades of Research on Induced Hypocrisy: A Meta-Analysis. *Personality and Social Psychology Bulletin*, 45(12), 1681–1701. <https://doi.org/10.1177/0146167219841621>
- Puhr, R., Heinze, G., Nold, M., Lusa, L., & Geroldinger, A. (2017). Firth's logistic regression with rare events: Accurate effect estimates and predictions? *Statistics in Medicine*, 36(14), 2302–2317. <https://doi.org/10.1002/sim.7273>
- R Core Team. (2021). *R: A Language and environment for statistical computing* [Computer software]. <https://cran.r-project.org>
- Sasaki, S., Okuyama, N., Ogaki, M., & Ohtake, F. (2017). Education and Pro-Family Altruistic Discrimination Against Foreigners: Five-Country Comparisons. *SSRN Electronic Journal*, 1002. <https://doi.org/10.2139/ssrn.2965327>
- Sénémeaud, C., Mange, J., Fointiat, V., & Somat, A. (2014). Being hypocritical disturbs some people more than others: How individual differences in preference for consistency moderate the behavioral effects of the induced-hypocrisy paradigm. *Social Influence*, 9(2), 133–148. <https://doi.org/10.1080/15534510.2013.791235>
- Sidanius, J., & Pratto, F. (2001). *Social Dominance: An Intergroup Theory of Social Hierarchy and Oppression*. Cambridge University Press.

- Smith, J. R., & Louis, W. R. (2008). Do as we say and as we do: The interplay of descriptive and injunctive group norms in the attitude-behaviour relationship. *British Journal of Social Psychology, 47*(4), 647–666. <https://doi.org/10.1348/014466607X269748>
- Son Hing, L. S., Li, W., & Zanna, M. P. (2002). Inducing hypocrisy to reduce prejudicial responses among aversive racists. *Journal of Experimental Social Psychology, 38*(1), 71–78. <https://doi.org/10.1006/jesp.2001.1484>
- Steffens, N. K., Schuh, S. C., Haslam, S. A., Pérez, A., & van Dick, R. (2015). “Of the group” and “for the group”: How followership is shaped by leaders’ prototypicality and group identification. *European Journal of Social Psychology, 45*(2), 180–190. <https://doi.org/10.1002/ejsp.2088>
- Stone, J., & Cooper, J. (2001). A Self-Standards Model of Cognitive Dissonance. *Journal of Experimental Social Psychology, 37*(3), 228–243. <https://doi.org/10.1006/jesp.2000.1446>
- Stone, J., & Fernandez, N. C. (2008). To Practice What We Preach: The Use of Hypocrisy and Cognitive Dissonance to Motivate Behavior Change. *Social and Personality Psychology Compass, 2*(2), 1024–1051. <https://doi.org/10.1111/j.1751-9004.2008.00088.x>
- Stone, J., & Fernandez, N. C. (2011). When thinking about less failure causes more dissonance: The effect of elaboration and recall on behavior change following hypocrisy. *Social Influence, 6*(4), 199–211. <https://doi.org/10.1080/15534510.2011.618368>
- Thøgersen, J. (2006). Norms for environmentally responsible behaviour: An extended taxonomy. *Journal of Environmental Psychology, 26*(4), 247–261. <https://doi.org/10.1016/j.jenvp.2006.09.004>
- Thøgersen, J. (2014). The mediated influences of perceived norms on pro-environmental behavior. *Revue d’Economie Politique, 124*(2), 179–193. <https://doi.org/10.3917/redp.242.0179>
- Thorman, D., Whitmarsh, L., & Demski, C. (2020). Policy Acceptance of Low-Consumption Governance Approaches: The Effect of Social Norms and Hypocrisy. *Sustainability, 12*(3), 1247. <https://doi.org/10.3390/su12031247>

- Vansteenkiste, M., Ryan, R. M., & Soenens, B. (2020). Basic psychological need theory: Advancements, critical themes, and future directions. *Motivation and Emotion, 44*(1), 1–31. <https://doi.org/10.1007/s11031-019-09818-1>
- Vautier, S., Raufaste, E., & Cariou, M. (2003). Dimensionality of the Revised Life Orientation Test and the status of filler items. *International Journal of Psychology, 38*(6), 390–400. <https://doi.org/10.1080/00207590344000222>
- Vosgerau, J., Simonsohn, U., Nelson, L. D., & Simmons, J. P. (2019). 99% impossible: A valid, or falsifiable, internal meta-analysis. *Journal of Experimental Psychology: General, 148*(9), 1628–1639. <https://doi.org/10.1037/xge0000663>
- Yoo, J., & Jin, Y. (2020). Comparative impact of fear appeals and induced hypocrisy advertising in encouraging intent to quit smoking: Applying self-construal theory to consumers' attitudes. *Global Health Promotion, 27*(1), 51–58. <https://doi.org/10.1177/1757975918783425>

Data availability statement

All data, analysis code, and materials are available on OSF at https://osf.io/k53mq/?view_only=d4752d9b66e240f987390e51e6dc6fcb.

Ethics statement

All participants provided written consent prior and after taking part in the studies. This research was conducted in accordance with the principles outlined in the Declaration of Helsinki (2008) and the ethical guidelines set forth by the American Psychological Association Ethical Principles of Psychologists and Code of Conduct (APA, 2019) for the ethical treatment of human participants.

Funding

The work reported was supported by a RIN Doctorant of Normandie Région, France.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Author contributions statements

Author 1: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Resources; Software; Validation; Visualization; Writing – Original Draft; Writing – Review and Editing

Author 2: Conceptualization; Methodology; Resources; Software; Supervision; Writing – Review and Editing

Author 3: Conceptualization; Methodology; Resources; Supervision; Writing – Review and Editing

Author 4: Conceptualization; Funding acquisition; Methodology; Resources; Project administration; Supervision; Writing – Review and Editing