



# Evaluating Others If They Stand Up for Their Moral Convictions, and Evaluating Ourselves If We Don't Stand Up

Florien M. Cramwinckel<sup>1,2,3</sup> · Kees van den Bos<sup>1</sup> · Eric van Dijk<sup>4</sup>

Accepted: 26 April 2021 / Published online: 1 June 2021  
© The Author(s) 2021

## Abstract

Acting on one's moral principles is not always easy. Upholding one's moral beliefs may run counter to one's social environment or situational demands. It may often cause people to remain silent on their convictions, while at the same time some may show the moral courage to speak out. How do people evaluate those who do stand up, and how does it affect their self-evaluations? In two experimental studies ( $N_s=207$  and  $204$ ), we investigated both types of evaluations. The studies demonstrate that people who failed to uphold their moral beliefs still had positive evaluations of others who showed moral courage. More specifically, pro-gay participants who went along with writing an anti-gay essay denouncing equal rights for sexual minorities had positive evaluations of another person who spoke up and refused this task. The failure to display moral courage had negative consequences for participants' self-concepts. In Experiment 1, we show that pro-gay participants' positive self-concepts were lowered after writing an anti-gay essay (vs. a pro-gay essay). In Experiment 2, we reveal that participants' positive self-concepts were lowered only when they were confronted with morally courageous behavior and their own failure to uphold their moral beliefs was visible to the experimenter.

**Keywords** Moral courage · Moral refuser · Gay rights · Moral conviction

---

Author affiliations changed for the first author. She was affiliated with Utrecht University<sup>1</sup> during the execution of the research and while preparing the original submission of this paper<sup>2</sup>. She is currently working as a senior behavioral expert at Online Dialogue<sup>3</sup>.

✉ Florien M. Cramwinckel  
f.m.cramwinckel@gmail.com

<sup>1</sup> Social, Health, and Organizational Psychology, Utrecht University, Utrecht, The Netherlands

<sup>2</sup> Department of Interdisciplinary Social Science: Youth Studies, Utrecht University, Utrecht, The Netherlands

<sup>3</sup> Online Dialogue, Sint Jacobsstraat 31, 3511 BL Utrecht, The Netherlands

<sup>4</sup> Social, Economic and Organizational Psychology, Leiden University, Leiden, The Netherlands

## Introduction

Gay rights are the focus of debate in many countries. Especially same-sex marriage has attracted attention. In 2001, the Netherlands legalized same-sex marriage. Since then, same-sex marriage has been legalized in 28 other countries, (<https://www.hrc.org/resources/marriage-equality-around-the-world>). In 2015, Ireland legalized same-sex marriage by popular vote ([www.marriagequality.ie](http://www.marriagequality.ie)). Later that year, the US supreme court ruled same-sex marriage as a right protected by the US constitution in all 50 states ([www.gaymarriage.procon.org](http://www.gaymarriage.procon.org)), marking the historic significance of this day for the gay rights movement.

In short, more and more people support marriage equality. For example, in the Netherlands, over eighty percent of the people support same-sex marriage (Netherlands Institute for Social Research [SCP], 2013). Thus, in most Western countries people are becoming more and more protective of equal rights for sexual minorities. However, this does not mean that the battle for equal rights for sexual minorities is over. In fact, some countries have recently decreased rights for those people who identify as Lesbian, Gay, Bisexual, Transgender and Queer or are questioning their sexual identity (LGBTQ). For instance, several rights for sexual and gender minorities in the USA have been restricted under the Trump administration (Gonzalez et al., 2018). Furthermore, in 2019, a leader of a religious political party in the Netherlands (SGP) signed the Nashville Declaration and denounced marriages between same-sex partners, as well as identifying as transgender (<https://sgp.nl/actueel/nieuws/reactie-van-der-staaij-op-nashville-verklaring>).

Apparently, equal rights for sexual and gender minorities can be fragile and should be continuously protected. Thus, even after implementation in law, equal rights need to be safeguarded by advocates who believe in the (moral) value of these rights. Upholding one's moral beliefs may, however, come with (social) costs (Halmburger et al., 2015; Hannah et al., 2011). This goes even for situations where the costs may be relatively small, such as a failure to be seen as a good participant in psychological research (Orne, 1962). As a result, many people may decide against standing up for their beliefs. The literature on moral courage, however, suggests that some may decide to stand up, even under these circumstances.

The key issue that we investigate is how people who fail to act as advocates respond when they see that others did actively defend gay rights. For this purpose, we asked pro-gay participants to write anti-gay essays after which we studied how those who complied to the request reacted when finding out that another participant refused to comply with the task, arguing that writing the essay was against their moral conviction. Our focus was on (a) how participants would evaluate this person, and (b) how participants would evaluate themselves.

## Evaluating Those Who Stand Up

The social psychological literature on reactions toward those who stand up for their convictions suggest two possible reactions. People may admire those who do stand

up, but they may also downgrade these others. A positive reaction may be expected, given that people generally value morality (e.g., Leach, Ellemers, & Barreto, 2007; Monin, 2007; Schwartz, 1992), and admire taking actions in such important domains (Van de Ven et al., 2019). Witnessing moral others may be an elevating experience (Pohling & Diessner, 2016). And since standing up may invite social repercussions, people also may feel—and appreciate—that it takes moral courage to stand up for one's beliefs (e.g., Greitemeyer et al., 2006; Greitemeyer, Osswald, & Frey, 2007; see also Dickter et al., 2012). Such appreciation might be shown irrespective of whether oneself stood up. In that case the evaluations of others would be based primarily on the basis of the behavior the other displayed. One could also reason that the appreciation of other's courageous behavior might be more pronounced for those who did not show such courage themselves.

People have, however, also been found to react negatively to those who stand up. The key issue in this line of reasoning is that people's evaluation of others who stand up may become more negative if they themselves did not stand up. Earlier work on do-gooder derogation, including our own, demonstrates that people who themselves are involved in the situation and remain silent may feel threatened by those who stand up. Cramwinckel, Van Dijk, Scheepers, & Van den Bos (2013), for example, showed that people who had just eaten meat in a task on taste perceptions, negatively evaluated moral vegetarians who—while participating in the same study—refused to eat the meat for the reason of it being immoral (see also Minson & Monin, 2012). In a similar vein, Monin et al. (2008) conducted a study in which participants first performed a police detection task in which they had to select the person they felt was most likely to have committed a burglary, while the information provided in the task incriminated an African-American target. When being informed that one other participant refused to answer the question arguing it was a racist task, participants who had completed the task predominantly evaluated this “moral rebel” negatively.

Based on these latter insights, one might expect a similar finding in our study, such that participants would evaluate a person refusing to write an anti-gay negatively. Note, however, that our current setup does differ from the cited studies above. As opposed to the meat tasting studies, for example, participants in our study showed behavior inconsistent with their conviction. We asked them to write an essay that was against their convictions, while participants (nonvegetarians) in Cramwinckel et al. (2013) showed behavior (tasting meat) that was consistent with their own conviction (they were not vegetarians), while the ‘moral refuser’ behaved inconsistent with the participants’ behavior (i.e., refused to eat meat). The Monin et al. (2008) study comes closer to the current study, but there too, the conflict may have been less apparent, as participants were not asked to perform a racist task. They were asked to perform a police task, which after taking part turned out to have an element that one of the other participants then identified as being racist. In our current study, the request was from the outset much more clearly inconsistent with the participants’ own attitudes: By asking people to write a speech against gay rights, participants would immediately realize they were doing something that ran counter to their conviction; a conviction that may be considered a key conviction. Would the negative evaluation of moral refusers also emerge under this condition? And if so, would such a negative evaluation also be contingent on one's own behavior such that

the negative evaluation would be especially observed among those who themselves had acted against their own conviction?

### Evaluating Oneself After Not Standing Up

The second issue we address is how people evaluate themselves after willingly having participated in a task that runs counter to their convictions and finding out that another person did stand up. The theory of moral self-regulation proposes that people's evaluations of themselves are based on whether their previous behavior matches their convictions (e.g., Zhong, Liljenquist, & Cain, 2009). In this process, people compare their behavior with their moral ideals. When people do not live up to their ideals, their self-image is threatened and needs to be restored. Based on this, we expect that people have more negative self-evaluations (e.g., being angry with themselves) after showing behavior that so clearly opposes their moral values than after showing behavior that aligns with their moral values. This implies that when people endorse equal rights for sexual minorities, they should experience more negative self-evaluations after writing an essay against gay rights (as opposed to having written an essay in favor of gay rights). Witnessing someone else who did stand up might aggravate such self-blame even further.

Importantly, rather than negatively adjusting one's self-concept, people can also engage in various types of defensive responding that protect the moral self-concept and prevent the need to adjust it (Cohen & Sherman, 2014). This means that the negative effect of not acting in line with one's moral beliefs on one's self-concept may be stronger in some situations than in others. Self-concept maintenance theory states that while people want to maintain a positive view of themselves as moral individuals, they can be tempted to display moral transgressions (Mazar et al., 2008). Whether people stick to their moral values or succumb to temptation depends on how well they are able to maintain a positive self-concept while transgressing. This is, among other things, influenced by how easy it is for people to see their behavior as moral (Mazar et al., 2008), and by the salience of their own behavior and moral standards (Diener & Wallbom, 1976). We examine these aspects in Experiment 2.

In all experiments, we aimed to collect a minimum of 50 participants per cell, and stopped data collection at the end of the week in which we achieved this aim. We used G\*Power (version 3.1.7.) to perform a sensitivity analysis to compute the minimum required effect size (given  $\alpha=0.05$ , power=0.80, and sample size=200, numerator df=1, number of groups=4, and number of covariates=0). Experiments 1 and 2 had 80% power to detect an effect size of  $f=0.199$ . We report all data exclusions, all manipulations, and all measures in the current studies. We used the program *statcheck* ([statcheck.io](https://statcheck.io)) to check for errors in statistical reporting, and found no inconsistent results. All stimulus materials, data, syntaxes and output can be viewed on the Open Science Framework ([https://osf.io/uzsav/?view\\_only=7d49b98ac0024fce84b6f6ba356493cf](https://osf.io/uzsav/?view_only=7d49b98ac0024fce84b6f6ba356493cf)). In all experiments, participants were recruited via posters in university buildings or were approached on campus grounds and invited to come to the laboratory. Potential participants could also come to the laboratory

without being actively approached, as multiple studies were running in the laboratory at the same time, and this was well known among students.

## Experiment 1

In this experiment, participants reported their own attitudes toward homosexuality and were asked to write an essay that either opposed equal rights for homosexuals (the anti-gay essay condition) or endorsed those rights (the pro-gay essay condition). We verified that most participants had relatively positive attitudes about homosexuals and disagreed with the statement that homosexuals should *not* have the same rights as heterosexuals. After having written an essay, participants read the bogus reaction of a person who refused to write an anti-gay essay because he/she considered being gay as normal and acceptable (i.e., the anti-essay refuser) or because he/she considered being gay as abnormal and unacceptable (i.e., the pro-essay refuser). Afterward, we measured participants' evaluations of the refuser and of themselves. We investigated whether participants would negatively (similar to previous studies on moral rebels) evaluate the refuser of the anti-gay essay, or whether they might positively evaluate the shown moral courage to stand up. As for the self-evaluations, we expected that participants would have lower positive self-evaluations after writing an anti-gay essay, rather than a pro-gay essay.

Although we were primarily interested in participants' reactions toward the refuser of the anti-gay essay, we included a condition with a refuser of the pro-gay essay condition. By doing so, we could rule out the possibility that reactions would be driven by the fact that the refuser was assigned a different task than the participant. We do realize that this condition is special because the pro-essay refuser makes a moral claim that is counter-normative in Dutch society. We discuss the possible implications of this control condition in the Discussion of this experiment, and there explain how we address this in Experiment 2 with a different control.

## Method

### Participants and Design

A total of 207 students at Utrecht University participated in this experiment in exchange for a monetary reward or course credits ( $M_{\text{age}} = 21.22$  years;  $SD_{\text{age}} = 3.74$ ; 130 women; 76 men; 1 participant failed to provide demographic information). The participants were randomly assigned to the conditions of our 2 (Own essay: pro-gay essay vs. anti-gay essay)  $\times$  2 (Refuser: pro-gay essay vs. anti-gay essay) experimental design. The number of participants within each condition varied between 50 and 53.

Twenty-nine additional participants were not included in the analyses: Eighteen participants participated more than once; we only included data from their first participation. Furthermore, we had two a priori criteria to exclude participants from the analyses: (a) refusing to write the essay (7 participants), and (b) writing an essay of one sentence or less (4 participants). These participants were excluded from the

research, resulting in a final N of 207. Besides the measures reported here, we also measured participants' moral identity (11 items,  $\alpha=0.73$ , Aquino & Reed, 2002) and social comparison orientation (11 items,  $\alpha=0.83$ , Gibbons & Buunk, 1999).

## Procedure

Participants were seated in cubicles. The complete experiment was run on a computer. We assessed participants' attitude regarding homosexuality by asking them to indicate on a continuous slider ranging from 0 to 100 (Haddock et al., 1993) the extent to which they felt negative or positive toward homosexuality.

Hereafter, the essay task started. Participants were asked to write an essay from the viewpoint of a passionate adversary of gay rights (the anti-gay essay condition) or from the viewpoint of a passionate advocate of gay rights (the pro-gay essay condition). In the anti-gay essay condition, participants were asked to defend the viewpoint that homosexuality is abnormal and unnatural and that gays should not have the same rights as straight people. In the pro-gay essay condition, participants were asked to defend the viewpoint that homosexuality is normal and natural, and that gays should have the same rights as straight people. They were asked to imagine as lively as possible that they agreed with the provided position. In both conditions, participants learned that their essay could be evaluated by another participant. In reality, the essays were not seen by other participants.

Given the fact that 87% of the Dutch population supports equal rights for lesbians and gay men (SCP, 2013), we anticipated that most participants in our sample would be relatively positive toward homosexuality. We verified this assumption by letting participants, before they wrote their essay, indicate to what extent they agreed with their assigned statement on a 101-point scale ranging from 0 (*disagree completely*) to 100 (*agree completely*).

Afterward, participants were exposed to the reaction of someone who had refused to write the essay out of moral concern. In the *anti-essay refuser* condition, the other participant refused to write the anti-gay essay because he/she thought homosexuality was normal and moral, and thus considered it to be morally wrong to write an anti-gay essay. In the *pro-essay refuser* condition, the other participant refused to write the pro-gay essay because he/she thought homosexuality was abnormal and immoral, and thus considered it to be morally wrong to write a pro-gay essay. In both conditions, participants were instructed to read the reaction of the other participant carefully and try to form an impression of this other person's character.

Subsequently, the dependent variables were collected. *Refuser evaluation* was measured with 45 items, assessing the extent to which the refuser was perceived as being nice and honest, among other things. *Self-evaluation* was measured with 23 items, such as the extent to which participants felt satisfied with themselves. See the Supplementary Information for all items (Online Resources 1 and 2). Answers were given on 7-point Likert scales, ranging from 1 (*not at all*) to 7 (*very much so*).

Hereafter, participants could note what they thought the research was about, and whether they had noticed something unusual during the experiment. Furthermore,

demographic variables were asked. Finally, participants were thanked, debriefed, and rewarded for their participation.

## Results

### Assumption Check

As expected, participants in the pro-gay essay condition agreed more with their assigned position ( $M=90.35$ ,  $SD=20.73$ ) than did participants in the anti-gay essay condition ( $M=12.75$ ,  $SD=23.07$ ),  $F(1, 203)=647.82$ ,  $p<0.001$ ,  $\eta^2=0.76$ . Furthermore, participants' attitudes toward homosexuality were on average positive ( $M=71.61$ ,  $SD=25.05$ ).

### Scale Construction

The first step in the data analysis was to create scales to measure evaluations of the refuser and the self. We performed exploratory factor analyses (EFAs) with oblique (Oblimin) rotation to create scales, and tested our hypotheses by conducting analyses of variance (ANOVAs) on these scales. The correlation between different factors can be seen in Table 1.

### Refuser Evaluation

We performed an EFA to investigate the underlying structure of the 45 items that measured refuser evaluation. The scree-plot showed a large indent after the first factor, and a nearly horizontal line after the third factor, suggesting a solution between one and three factors.

Hereafter, we performed factor analyses fixed with 1, 2, or 3 factors. The three-factor solution fit with recent research showing that evaluations of others can be

**Table 1** Correlations between factors in Experiments 1 and 2

	Morality	Competence	Sociability	Self-Evaluation
<i>Experiment 1</i>				
Morality		.593***	.892***	.015
Competence			.538***	.021
Sociability				.039
Self-Evaluation				
<i>Experiment 2</i>				
Morality		.646***	.806***	– .099
Competence			.601***	– .092
Sociability				– .104
Self-Evaluation				

\*  $p<.05$ , \*\*  $p<.01$ , \*\*\*  $p<.001$

divided into three domains, namely morality, competence, and sociability (e.g., Brambilla et al., 2011; Ellemers et al., 2008; Leach, et al., 2007). Furthermore, the three-factor solution explained 65.96% of variance and each factor had at least five items that loaded strongly ( $>|.55|$ ) on only that scale, suggesting good factor loadings (Tabachnick & Fidell, 2007). Therefore, we opted for the three-factor solution. These three factors were interpreted as *morality* (e.g., the extent to which the refuser is perceived as moral and just.  $\alpha=0.97$ , 14 items), *competence* (e.g., the extent to which the refuser is perceived as confident and strong.  $\alpha=0.91$ , 7 items), and *sociability* (e.g., the extent to which the refuser is perceived as pleasant and obnoxious [reverse coded].  $\alpha=0.95$ , 5 items). Only items with strong factor loadings ( $>|.55|$ ) and without cross loadings (i.e., factor loadings  $>|.55|$  on multiple factors) were included, which led to the exclusion of 19 items. Which items are included in which scales and how strongly they load on this scale is displayed in the Supplementary Information (Online Resource 3).

### Self-evaluation

We performed an EFA to investigate the underlying structure of the 23 items that measured self-evaluation. The scree plot showed a large indent after the first component, indicating the underlying structure of a single component. The first component explained 43.66% of the variance.

Hereafter, we performed factor analyses, fixed with 1, 2, or 3 factors. The one-factor solution had the most items with good factor loadings ( $>|.55|$ ) on this factor (Tabachnick & Fidell, 2007). Therefore, we opted for the one-factor solution, which we interpreted as *self-evaluation* ( $\alpha=0.95$ , 16 items). Only items that loaded strongly ( $>|.55|$ ) were included, which led to the exclusion of 7 items. Which items are included in the scale and how strongly they load on this scale is displayed in Online Resource 1 (Supplementary Information).

### Main Analyses

We performed ANOVAs with the Own Essay and Refuser manipulations as the independent variables, and our outcome measures as the dependent variables.

#### Refuser Morality

Results only showed a significant effect of refuser,  $F(1, 203)=305.53$ ,  $p<0.001$ ,  $\eta_p^2=0.60$ , indicating that participants perceived the refuser of the anti-gay essay as more moral ( $M=5.55$ ,  $SD=0.96$ , 95% CI [5.35, 5.76]), than the refuser of the pro-gay essay ( $M=3.03$ ,  $SD=1.10$ , 95% CI [2.83, 3.23]). We did not observe a significant main effect of own essay,  $F(1, 203)=0.06$ ,  $p=0.808$ ,  $\eta_p^2<0.001$ , nor did we observe a significant interaction effect,  $F(1, 203)=0.31$ ,  $p=0.579$ ,  $\eta_p^2<0.001$ . See Table 2 for a detailed overview of all statistics, including the non-significant results.



**Table 2** All statistics main analyses Experiment 1

		<i>Refuser Essay</i>																	
		Pro-gay essay refuser				Anti-gay essay refuser				Total									
		M	SD	LL	95% CI	UL	95% CI	M	SD	LL	95% CI	UL	95% CI	M	SD	LL	95% CI	UL	95% CI
<b>DV: Morality</b>	<i>Own Essay</i>																		
	Pro-gay essay	3.05	1.20	2.77	3.34	3.34	3.34	5.50	1.11	5.22	5.78	5.78	5.78	4.29	1.68	4.07	4.47	4.47	4.47
	Anti-gay essay	3.01	1.00	2.72	3.29	3.29	3.29	5.61	0.79	5.32	5.90	5.90	5.90	4.28	1.59	4.11	4.51	4.51	4.51
	Total	3.03	1.10	2.83	3.23	3.23	3.23	5.55	0.96	2.83	3.23	3.23	3.23	4.28	1.63	4.06	4.51	4.51	4.51
<b>DV: Competence</b>	<i>Own Essay</i>																		
	Pro-gay essay	4.58	1.45	4.26	4.91	4.91	4.91	5.74	1.24	5.42	6.06	6.06	6.06	5.17	1.46	4.93	5.39	5.39	5.39
	Anti-gay essay	5.01	1.04	4.69	5.33	5.33	5.33	5.75	0.91	5.42	6.08	6.08	6.08	5.37	1.04	5.15	5.61	5.61	5.61
	Total	4.80	1.28	4.57	5.02	5.02	5.02	5.74	1.09	5.51	5.97	5.97	5.97	5.27	1.27	5.09	5.44	5.44	5.44
<b>DV: Sociability</b>	<i>Own Essay</i>																		
	Pro-gay essay	2.72	1.28	2.34	3.09	3.09	3.09	4.49	1.40	4.12	4.86	4.86	4.86	3.61	1.60	3.34	3.86	3.86	3.86
	Anti-gay essay	2.80	1.29	2.43	3.17	3.17	3.17	4.44	1.48	4.06	4.82	4.82	4.82	3.61	1.61	3.36	3.89	3.89	3.89
	Total	2.76	1.28	2.49	3.02	3.02	3.02	4.47	1.43	4.20	4.73	4.73	4.73	3.61	1.60	3.39	3.83	3.83	3.83
<b>DV: Self-Evaluation</b>	<i>Own Essay</i>																		
	Pro-gay essay	5.34	0.89	5.07	5.61	5.61	5.61	5.62	0.80	5.36	5.88	5.88	5.88	5.48	0.85	5.29	5.67	5.67	5.67
	Anti-gay essay	5.06	1.03	4.80	5.33	5.33	5.33	4.93	1.15	4.66	5.20	5.20	5.20	5.00	1.09	4.81	5.19	5.19	5.19
	Total	5.20	0.97	5.01	5.39	5.39	5.39	5.28	1.04	5.09	5.46	5.46	5.46	5.24	1.00	5.11	5.38	5.38	5.38

## Refuser Competence

Results only showed a significant effect of refuser,  $F(1, 203)=33.21$ ,  $p<0.001$ ,  $\eta_p^2=0.14$ , indicating that participants ascribed higher competence ratings to the refuser of the anti-gay essay ( $M=5.74$ ,  $SD=1.09$ , 95% CI [5.51, 5.97]), than to the refuser of the pro-gay essay ( $M=4.80$ ,  $SD=1.28$ , 95% CI [4.57, 5.02]). We did not observe a significant main effect of own essay,  $F(1, 203)=1.74$ ,  $p=0.189$ ,  $\eta_p^2=0.01$ , nor of the interaction effect,  $F(1, 203)=1.62$ ,  $p=0.204$ ,  $\eta_p^2=0.01$ . See Table 2 for a detailed overview of all statistics, including the non-significant results.

## Refuser Sociability

Results only showed a significant effect of refuser,  $F(1, 203)=81.28$ ,  $p<0.001$ ,  $\eta_p^2=0.29$ , indicating that participants perceived the refuser of the anti-gay essay as more sociable ( $M=4.47$ ,  $SD=1.43$ , 95% CI [4.20, 4.73]), than the refuser of the pro-gay essay ( $M=2.76$ ,  $SD=1.28$ , 95% CI [2.49, 3.02]). We did not observe a significant main effect of own essay,  $F(1, 203)=0.01$ ,  $p=0.920$ ,  $\eta_p^2<0.001$ , nor did we observe a significant interaction effect,  $F(1, 203)=0.11$ ,  $p=0.737$ ,  $\eta_p^2=0.001$ . See Table 2 for a detailed overview of all statistics, including the non-significant results.

## Self-Evaluation

Results only showed a significant effect of own essay,  $F(1, 203)=12.71$ ,  $p<0.001$ ,  $\eta_p^2=0.06$ , indicating that participants experienced more positive self-evaluations after writing a pro-gay essay ( $M=5.48$ ,  $SD=0.85$ , 95% CI [5.29, 5.67]), than after writing an anti-gay essay ( $M=5.00$ ,  $SD=1.09$ , 95% CI [4.81, 5.19]). We did not observe a significant main effect of refuser,  $F(1, 203)=0.28$ ,  $p=0.596$ ,  $\eta_p^2<0.001$ , nor did we observe a significant interaction effect,  $F(1, 203)=2.30$ ,  $p=0.131$ ,  $\eta_p^2=0.01$ . See Table 2 for a detailed overview of all statistics, including the non-significant results.

## Discussion

The results of our first study showed that the participants had more positive evaluations of the refuser of the anti-gay essay (vs. the refuser of the pro-gay essay) on all important elements of person perception (e.g., Leach et al., 2007). More specifically, the refuser of the anti-gay essay was seen as more moral, more competent and more sociable than the refuser of the pro-gay essay. This means that participants appreciated the moral character of the person who refused to write an anti-gay essay. Thus, participants had positive evaluations of people who upheld their moral beliefs in a situation where most people failed to do so.

These results thus support the view that people may positively evaluate others who stand up for their moral convictions (e.g., Greitemeyer et al., 2006, 2007; Pohling & Diessner, 2016; Van de Ven et al., 2019; see also Dickter et al., 2012). It is relevant to note that these positive evaluations were not contingent on whether

or not one had personally stood up for one's convictions (i.e., we did not observe an interaction). This suggests that people can appreciate courageous decisions of others when they themselves did not stand up but also when they themselves did stand up.

Also note, that these findings do not match those obtained in previous research on reactions to moral rebels (e.g., Cramwinckel et al., 2013; Monin et al., 2008). While we do not consider our study as a critical experiment regarding negative versus positive evaluations, it is noteworthy that we did not find any indications that participants reacted negatively to the person standing up, even though they themselves did not. At the very least, these findings suggest that being involved and remaining silent does not always lead to a negative evaluation of those who do speak up. We will return to this in the General Discussion.

With regard to their self-evaluations, participants experienced more positive self-evaluations after writing an explicit pro-gay than after writing an anti-gay essay. Here too, we did not observe an interaction such that these self-evaluations were not contingent on other's behavior. This is in line with the theory of moral self-regulation, which posits that people compare their own behaviors with their own moral standards (e.g., Zhong et al., 2009). The consequences for one's self-concept may, however, also rely on the extent to which one's actions are visible to others. Such considerations may be especially relevant in a context of standing up. For example, research by Greene and Low (2014) demonstrates that although people were more likely to engage in a moral transgression after remembering a moral deed (i.e., engage in moral licensing) rather than an immoral deed, this only occurred when the behavior was private and not when it was public. This fits with literature on public self-awareness which demonstrates that situations where one's "attention is directed toward the implications of one's behavior for other's evaluations of the self" increase awareness of one's own behavior and self-image (Hirt et al., 2000, p. 1133). Thus, the salience of one's behavior to others also influences the consequences thereof for one's self-evaluation. Therefore, in our second experiment, we investigated whether self-evaluations suffer the most when people's behavior is visible to others and it is difficult for participants to view their own behavior as moral.

At this point it is relevant to also consider the control condition we used. The pro-gay essay refuser clearly deviated from important norms in the Dutch society where homosexuality is seen as acceptable. Given that deviating from societal norms is a strong source of dislike and antipathy (Marques & Yzerbyt, 1988; Skitka, 2010), participants may have disliked the refuser of the pro-gay essay because the refuser supported a deviant position in society. Differences in reactions to the refuser of the anti-gay issue versus the control, may thus at least partly reflect the negative view of the refuser in the control condition. To rule out this possibility, we included a different control condition in Experiment 2.

## Experiment 2

In our second experiment, we wanted to investigate the robustness of the Experiment 1's finding that people have positive evaluations of others who showed moral courage by refusing to write an anti-gay essay out of moral concern.

Furthermore, we wanted to zoom in on factors that may influence the effect of a failure to show moral courage on people's self-evaluations. In order to do so, we investigated whether people's self-evaluations after engaging in anti-gay behavior depends on whether or not they feel that their counter-attitudinal behavior is visible to others. There are several areas of research that indicate that the visibility of one's behavior to others is important, especially when engaging in immoral behavior. People have a strong motive to present themselves favorably to others (for a review, see, e.g., Geen, 1991) and therefore try to engage in praiseworthy behavior and refrain from undesirable behavior when their behavior is visible to others. For example, people primarily engage in moral transgressions in private settings and not in public settings (Greene & Low, 2014). Furthermore, cues of being watched lead to more pro-social behavior in real-life settings (Bateson et al., 2006) and people are especially likely to buy "green" products in public rather than private settings (Griskevicius, Tybur, & Van den Bergh, 2010).

In Experiment 1, people's own anti-gay behavior was salient to others because participants were informed that their essay could be read by the other participant. This increased public self-awareness may have made it more difficult for them to maintain a positive self-concept (e.g., Mazar et al., 2008), because of increased concern of how their behavior would be evaluated by others (e.g., Hirt et al., 2000). We therefore argue that engaging in anti-gay behavior within a context that is highly evaluative leads to lowered positive self-evaluations, while these negative consequences may be attenuated when the evaluative context is less salient. To test this prediction, we manipulated in Experiment 2 whether participants were informed that their anti-gay essays would be read by the experimenter (the strong evaluative context condition) or not (the mild evaluative context condition).

Furthermore, we manipulated how hard it was for participants to see their anti-gay behavior as moral. Building on the work from Mazar et al. (2008), we expected that people's positive self-concepts would be most negatively impacted when it would be relatively difficult to view one's own behavior as moral, while their positive self-concepts would remain relatively unaffected when it would be easier to view one's own behavior as moral. Participants therefore read the bogus reaction of another participant who refused to write the anti-gay essay out of moral concern or out of non-moral concern (i.e., because he/she had an injury and therefore could not type an essay). It should be harder for participants to view their own behavior as moral when someone else claimed that writing an anti-gay essay was immoral. We expected people's positive self-evaluations to be lower when they were confronted with a refuser who makes a moral claim (vs. a non-moral claim), but primarily when they were in a strong evaluative context.

Moreover, including the non-moral reason condition enabled us to rule out the alternative explanation from Experiment 1, that the positive evaluation of the refuser of the anti-gay essay was due to the deviant societal position of the refuser. More specifically, in Experiment 2, the reactions were both consistent with Dutch societal norms. In either case the other participant refused to engage in counter-normative behavior (i.e., writing the anti-gay essay), albeit for different reasons.

## Method

### Participants and Design

A total of 204 students at Utrecht University participated in exchange for a monetary reward or course credits ( $M_{\text{age}}=21.12$  years;  $SD_{\text{age}}=2.76$  years; 120 women. 193 participants self-identified as heterosexual, two as homosexual/lesbian, and two as bisexual). Participants were randomly assigned to one of the four conditions of our 2 (Refusal Reason: moral reason vs. non-moral reason)  $\times$  2 (Evaluative Context: strong vs. mild) experimental design.

Twenty-one additional participants completed this experiment but were not included in the analyses. Five participants participated more than once in Experiment 2, and we therefore only included their data from the first participation. Furthermore, as in Experiment 1, we excluded participants who refused to write the essay (15 participants) and/or wrote an essay of one sentence or less (1 participant). This resulted in a final N of 204. The number of participants per cell varied between 47 and 54.

### Procedure

The procedure was similar to Experiment 1. We therefore only elaborate on the differences. All participants were asked to write an essay from the viewpoint of a passionate adversary of gay rights (the anti-gay essay condition of Experiment 1).

Participants were asked to write a good essay. Following research by Van den Bos et al. (1999), half of the participants were informed that their essay would be evaluated by the experimenter and this message was repeated three times throughout the experiment (*strong evaluative context condition*). The other half of the participants was not given the information that the experimenter would evaluate their essay (*mild evaluative context condition*; Van den Bos et al., 1999).

Afterward, participants were confronted with the reaction of someone who had refused to write the anti-gay essay. In the *non-moral reason* condition, the other participant refused to write the pro-gay essay because he/she had injured his/her hand and could therefore not type an essay. The *moral reason* condition was exactly the same as the anti-gay essay refuser condition in Experiment 1.

*Refuser morality* (14 items,  $\alpha=0.96$ ), *refuser competence* (7 items,  $\alpha=0.91$ ), *refuser sociability* (5 items,  $\alpha=0.92$ ), and self-evaluation (16 items,  $\alpha=0.95$ ) were measured similarly as in Experiment 1. Answers were given on 7-point Likert scales, ranging from 1 (*not at all*) to 7 (*very much so*). We measured participants' fear of negative evaluation with a state version of the brief *Fear of Negative Evaluation scale* (12 items,  $\alpha=0.91$ ; Leary, 1983), consisting of items such as "Right now, I am afraid others will not approve of me". Answers were given on 7-point Likert scales, ranging from 1 (*Not at all characteristic of me, at this moment*) to 7 (*Extremely characteristic of me, at this moment*). Participants' attitudes toward lesbians, gay men and bisexuals (LGBs) were measured with an adapted version of

the *Attitudes toward Homosexuals scale* (21 items,  $\alpha=0.91$ ; Kite & Deaux, 1986), which included items such as “I would not mind having a lesbian/gay/bisexual friend”. Answers were given on 7-point Likert scales, ranging from 1 (*Disagree strongly*) to 7 (*Agree strongly*). We also assessed participants’ sexual orientation.

## Results

### Assumption Check

As expected, agreement with the assigned anti-gay essay was low ( $M=11.80$ ,  $SD=21.71$ ), which indicated that the assignment did not fit with participants moral beliefs. Agreement did not differ between the different conditions of our experiment (all  $ps > 0.290$ ). Participants’ attitudes toward homosexuality were on average positive ( $M=68.86$ ,  $SD=24.95$ ), and did not differ between different conditions of the experiment (all  $ps > 0.250$ ).

### Check on Attitudes

There was a significant and strong correlation between the one-item thermometer question that assessed attitudes toward homosexuality, and the 21-item extended questionnaire that assessed attitudes toward LGBs ( $r=0.67$ ,  $p < 0.001$ ), thereby validating our use of the one-item thermometer question as an assessment of participants’ attitudes toward homosexuality.

### Check on Evaluative Context

A one-way ANOVA indicated that the evaluative context manipulation had some influence on participants’ measured state fear of negative evaluation (Leary, 1983). Participants in the strong evaluative context condition scored somewhat higher on this scale ( $M=3.80$ ,  $SD=1.16$ ) than participants in the mild evaluative context condition ( $M=3.53$ ,  $SD=1.16$ ), but this effect was not significant,  $F(1, 202)=2.76$ ,  $p=0.099$ ,  $\eta^2=0.01$ .

### Main Analyses

We performed ANOVAs with the Refusal Reason and Evaluative Context manipulations as the independent variables.

#### Refuser Morality

Results only showed a significant effect of the refusal reason manipulation,  $F(1, 200)=221.25$ ,  $p < 0.001$ ,  $\eta_p^2=0.53$ , indicating that participants in the moral reason condition evaluated the refuser as more moral ( $M=5.62$ ,  $SD=0.93$ , 95% CI [5.44, 5.80]), than participants in the non-moral reason condition did ( $M=3.69$ ,  $SD=0.92$ ,

95% CI [3.51, 3.87]). We did not observe a significant main effect of evaluative context,  $F(1, 200)=0.38$ ,  $p=0.537$ ,  $\eta_p^2=0.002$ , nor did we observe a significant interaction effect,  $F(1, 200)=2.14$ ,  $p=0.145$ ,  $\eta_p^2=0.01$ . See Table 3 for a detailed overview of all statistics, including the non-significant results.

### Refuser Competence

Results showed a significant effect of the refusal reason manipulation,  $F(1, 200)=61.46$ ,  $p<0.001$ ,  $\eta_p^2=0.24$ , indicating that participants in the moral reason condition ascribed higher competence ratings to the refuser ( $M=5.78$ ,  $SD=1.16$ , 95% CI [ 5.44,5.90]), than did participants in the non-moral reason condition ( $M=4.47$ ,  $SD=1.21$ , 95% CI [4.09, 4.55]). We did not observe a significant main effect of evaluative context,  $F(1, 200)=1.13$ ,  $p=0.290$ ,  $\eta_p^2=0.01$ , nor did we observe a significant interaction effect,  $F(1, 200)=2.94$ ,  $p=0.088$ ,  $\eta_p^2=0.02$ . See Table 3 for a detailed overview of all statistics, including the non-significant results.

### Refuser Sociability

Results only showed a significant effect of the refusal reason manipulation,  $F(1, 200)=88.27$ ,  $p<0.001$ ,  $\eta_p^2=0.31$ , indicating that participants in the moral reason condition evaluated the refuser as more sociable ( $M=4.62$ ,  $SD=1.36$ , 95% CI [4.37, 4.85]), than participants in the non-moral reason condition ( $M=3.01$ ,  $SD=1.03$ , 95% CI [2.78, 3.25]). We did not observe a significant main effect of evaluative context,  $F(1, 200)=0.01$ ,  $p=0.921$ ,  $\eta_p^2<0.001$ , nor did we observe a significant interaction effect,  $F(1, 200)=0.71$ ,  $p=0.401$ ,  $\eta_p^2=0.004$ . See Table 3 for a detailed overview of all statistics, including the non-significant results.

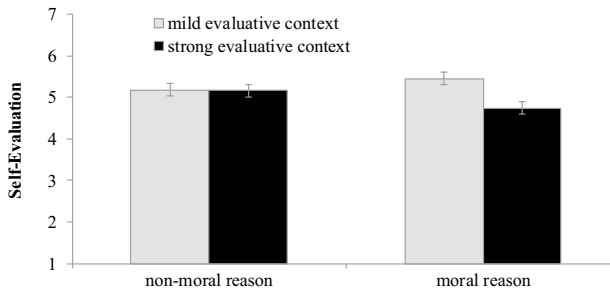
### Self-evaluation

Results showed a significant main effect of evaluative context,  $F(1, 200)=6.22$ ,  $p=0.013$ ,  $\eta_p^2=0.03$ , indicating that after writing an anti-gay essay, participants experienced higher (i.e., more positive) self-evaluations in a mild evaluative context ( $M=5.32$ ,  $SD=0.95$ , 95% CI [5.12, 5.52]), rather than a strong evaluative context ( $M=4.96$ ,  $SD=1.15$ , 95% CI [4.74, 5.16]). Furthermore, we observed a significant Refusal Reason x Evaluative Context interaction,  $F(1, 200)=5.41$ ,  $p=0.021$ ,  $\eta_p^2=0.03$ . Simple main effects demonstrated that participants in the non-moral reason condition experienced similarly high levels of self-evaluation, regardless of whether they were in a mild evaluative context ( $M=5.18$ ,  $SD=0.99$ ) or strong evaluative context ( $M=5.16$ ,  $SD=0.95$ ),  $F(1, 200)=0.01$ ,  $p=0.905$ ,  $\eta_p^2<0.001$ . Figure 1 shows these effects. Importantly, and in support of our expectations, participants in the moral reason condition had lowered self-evaluations in the strong evaluative context ( $M=4.74$ ,  $SD=1.32$ ) compared to the mild evaluative context ( $M=5.45$ ,  $SD=0.90$ ),  $F(1, 200)=11.47$ ,  $p=0.001$ ,  $\eta_p^2=0.05$ . Figure 1 illustrates these effects as well. We did not observe a significant main effect of Refuser,  $F(1, 200)=0.30$ ,  $p=0.587$ ,  $\eta_p^2=0.001$ . See Table 3 for a detailed overview of all statistics, including the non-significant main effect.

**Table 3** All statistics main analyses Experiment 2

	<i>Refusal Reason</i>																		
	Non-Moral Reason				Moral Reason				Total										
	M	SD	LL	95% CI	UL	95% CI	M	SD	LL	95% CI	UL	95% CI	M	SD	LL	95% CI	UL	95% CI	
<b>DV: Morality</b>																			
<i>Evaluative Context</i>																			
Mild evaluative context	3.55	0.97	3.30	3.81	3.81	3.81	5.67	0.87	5.42	5.42	5.42	5.92	5.92	4.63	1.40	4.44	4.44	4.79	4.79
Strong evaluative context	3.82	0.85	3.57	4.08	4.08	4.08	5.56	1.01	5.30	5.30	5.30	5.83	5.83	4.66	1.27	4.51	4.51	4.88	4.88
Total	3.69	0.92	3.51	3.87	3.87	3.87	5.62	0.93	5.44	5.44	5.44	5.80	5.80	4.64	1.34	4.46	4.46	4.83	4.83
<b>DV: Competence</b>																			
<i>Evaluative Context</i>																			
Mild evaluative context	4.25	1.25	3.93	4.57	4.57	4.57	5.83	1.05	5.51	5.51	5.51	6.14	6.14	5.05	1.39	4.81	4.81	5.26	5.26
Strong evaluative context	4.71	1.12	4.38	5.03	5.03	5.03	5.72	1.29	5.38	5.38	5.38	6.06	6.06	5.19	1.30	4.98	4.98	5.45	5.45
Total	4.47	1.21	4.25	4.71	4.71	4.71	5.78	1.16	5.54	5.54	5.54	6.00	6.00	5.12	1.35	4.93	4.93	5.30	5.30
<b>DV: Sociability</b>																			
<i>Evaluative Context</i>																			
Mild evaluative context	2.93	1.08	2.60	3.27	3.27	3.27	4.67	1.36	4.35	4.35	4.35	4.99	4.99	3.82	1.50	3.57	3.57	4.03	4.03
Strong evaluative context	3.09	0.98	2.76	3.43	3.43	3.43	4.54	1.38	4.20	4.20	4.20	4.89	4.89	3.79	1.39	3.58	3.58	4.06	4.06
Total	3.01	1.03	2.78	3.25	3.25	3.25	4.61	1.36	4.37	4.37	4.37	4.85	4.85	3.80	1.45	3.61	3.61	4.00	4.00
<b>DV: Self-Evaluation</b>																			
<i>Evaluative Context</i>																			
Mild evaluative context	5.19	0.99	4.90	5.47	5.47	5.47	5.45	0.90	5.17	5.17	5.17	5.73	5.73	5.32	0.95	5.12	5.12	5.52	5.52
Strong evaluative context	5.16	0.95	4.87	5.45	5.45	5.45	4.74	1.32	4.44	4.44	4.44	5.04	5.04	4.96	1.15	4.74	4.74	5.16	5.16
Total	5.17	0.97	4.97	5.38	5.38	5.38	5.12	1.16	4.89	4.89	4.89	5.30	5.30	5.15	1.06	5.00	5.00	5.29	5.29





**Fig. 1** Refusal Reason by Evaluative Context Interaction on self-evaluation in Experiment 2. Error bars represent standard errors

## Discussion

We again observed that people had positive evaluations of others who refused to write an anti-gay essay out of moral concern—now compared to someone who refused to write an anti-gay essay out of non-moral concern. We thus did not only observe this positive evaluation when contrasted with an extremely deviant position in society (as in Experiment 1) but also when contrasted with a non-deviant position in society. This is important because it fits with the idea that we investigate a situation where people appreciate the character of those who—out of moral reasons—refuse to engage in behavior that they consider to be wrong. In two experiments, we now observed that people have positive evaluations of others who show moral courage by refusing to write an anti-gay essay out of moral reasons.

Furthermore, we provided further understanding on how a failure to stand up may impact one's evaluation of oneself. Our findings extend self-concept maintenance theory (Mazar et al., 2008). This theory states that people are less likely to engage in unethical behavior when people are self-aware and it is hard to categorize their behavior as moral, because this makes it harder to maintain a positive self-concept. We demonstrate that this may also work the other way around: After engaging in anti-gay behavior, people's evaluations of themselves were only lowered when they believed their essays would be read by the experimenter and they read a moral reason of someone else refusing to engage in the same task. We realize that we should be cautious in our interpretations. For example, one might also wonder whether demand characteristics and/or self-presentation concerns might have played a role when participants believed their essay could be read by the experimenter. Note, however, that our manipulation only pertained to whether the essay would be read by the experimenter. In none of our conditions did we make any suggestion that the answers to the questions would be read by the experimenter. It therefore seems unlikely that the pattern of findings we observed could be explained by demand characteristics or self-presentation concerns. In this respect, it may also be noted that we did not observe a main effect of evaluative context, meaning that only the fact that participants thought their essay would be read by the experimenter was not enough to lower their self-evaluations. We do have to acknowledge that our check

on the evaluative context did not show a significant effect of our manipulation of the fear for being evaluated. Some caution in interpreting the effects is therefore warranted. The data did show, however, that our manipulation affected how they evaluated themselves. With some caution, we suggest that the knowledge that their essays would be read by the experimenter may have made them more self-conscious and thereby increased awareness of their own behavior and moral standards. Combined, these findings show that engaging in anti-gay behavior that does not fit with one's moral values can have negative consequences for one's evaluation of oneself, but only under specific conditions.

## General Discussion

In two studies, we investigated how people react when they do not stand up for their convictions, while someone else does. We demonstrated that people who failed to uphold their moral beliefs still had positive evaluations of others who stand up. More specifically, pro-gay participants who went along with writing an anti-gay essay denouncing equal rights for sexual minorities had positive evaluations of others who had refused this task. We observed this on the evaluation dimensions morality, competence and sociability. Given the conceptual differences we analyzed these dimensions in our studies separately. It may be noted, however, that in their evaluations, people may hold similar views on different dimensions. A positive evaluation of morality may well come with a positive evaluation of sociability, and that is indeed what our data showed as reflected in the strong positive correlations between the evaluations (Table 1). Also note, however, that the presence of different dimensions could potentially lead to a differentiated pattern. For example, when seeing someone else taking a more moral stance, people might positively value the other high on agency, but lower their ratings of the other's likability (see, e.g., Cramwinckel et al., 2013, Experiment 2). In the current paper, we did not observe this differentiated pattern, but perhaps this possibility could be examined in future research.

In Experiment 1, we demonstrated that participants had more positive evaluations of the other person when they read a refusal to write an anti-gay essay, rather than a refusal to write a pro-gay essay. Furthermore, participants reported more positive self-evaluations after having written a pro-gay, rather than an anti-gay essay. In Experiment 2, all participants wrote an anti-gay essay. As in Experiment 1, people had positive evaluations of the other participant who refused to write an anti-gay essay out of moral concern. We also showed that participants' self-evaluations were lowered when they read a moral reason rather than a non-moral reason, but only when they believed that their own anti-gay essay would be read by the experimenter.

By studying how people react to others who stand up for their convictions, our studies add to the literature by addressing an issue that may potentially keep people from standing up: fear for negative social repercussions. We observed that in this particular setting, standing up was not evaluated negatively. In fact, non-morally courageous participants had positive evaluations of others, on the three most important dimensions of person perception (Leach et al., 2007). This insight may prove to be relevant in future studies and theorizing on obstacles that people perceive in

upholding their moral beliefs. Perhaps some of these obstacles (e.g., the fear for negative repercussions) may not be as high as feared, or may not even be the main obstacle for people to inhibit morally courageous behavior.

Another relevant contribution is the focus on people's evaluation of one-self. How moral courage or the failure to show moral courage impacts people's self-evaluations has not been investigated thoroughly. However, a deeper understanding of how people's self-images are influenced by these behaviors may be very relevant in better understanding what drives people to show or not show moral courage. This is also related to the theory of moral self-regulation (e.g., Zhong et al., 2009). More specifically, people's self-concepts were more negative after writing an anti-gay essay rather than a pro-gay essay (Experiment 1). Interestingly, we also present evidence that people's self-concepts are flexible and only affected under specific conditions. In an extension of self-concept maintenance theory (Mazar et al., 2008), we demonstrate that people's self-concepts are only negatively impacted when their own behavior is salient and it is hard to view their behavior as moral (Experiment 2). This implies that people may often remain silent without negative repercussions to their self-image. This may, however, also be contingent on the type of behavior that people display. In the current studies, people not standing up meant that participants went along with the requested task. This may have allowed participants to construe their decision in terms of inaction ("I failed to stand up for equal rights by not refusing to write this essay") rather than in terms of action ("I actively undermined equal rights by writing this essay"). As research on the omission bias has shown, reactions to immoral inactions are often milder than to immoral actions (e.g., Cushman et al., 2006; Ritov & Baron, 1992). An interesting implication of this observation would be to assess what would happen if participants in future studies would have the opportunity to go along with our request in more active way. For example, in future research, participants could be asked to actively indicate that they are willing to write the essay? It could be assessed whether this alternative way of standing up could have an impact on participants' actual behaviors and self-evaluations. Perhaps this future research would reveal that making compliance an active decision leads to stronger effects on self-evaluations. For future research it may be interesting to see whether the action-inaction difference is also relevant to understanding reactions to (not) standing up.

Future studies could also address another issue, namely whether it matters whether people already experience a moral conflict before finding out that somebody else steps up, or whether the stepping-up of others makes people realize that they in fact showed questionable behavior. Having actively shown immoral behavior (as opposed to inactively) may increase the likelihood that people already experience a moral conflict. As we discussed in our introduction, following up on an explicit request to write an essay that is against one's conviction may also generate more of a conflict than acting against one's convictions in more ambiguous settings. We have no data that speaks directly to this matter. But this may be an important issue for future research to address as well. Such studies could, for example, shed more light on the issue of when people derogate those who stand up (as found in the moral rebel literature) and when (as in the current study) people positively evaluate those who stand up.

The findings we obtained here may be relevant for the literature on defensive responding (e.g., Cohen & Sherman, 2014) by highlighting a situation where a failure to show moral behavior does not lead to defensive responding (i.e., negatively evaluating the other person). This is important because one could argue that people will only change their behavior when they feel motivated to compensate their previous mischiefs with improved future behavior (e.g., Zhong et al., 2009). Thus, it can be argued that more negative self-evaluations will motivate compensation behavior. This will especially be the case in situations where no defensive responding occurs (i.e., no derogation of the other person), because defensive responses are an alternative way of dealing with self-threat that will prevent behavioral change (e.g., Cohen & Sherman, 2014). Thus, we suggest that compensation behavior will be particularly likely when the need for compensation is high (i.e., lowered self-evaluations) and defensive responding is low (i.e., positive other-evaluations). This would imply, for instance, that although people could be easily enticed to write an anti-gay essay, they might afterward donate more time or money to a pro-gay charity in order to restore their moral self-worth.

Related to this is our observation that being confronted with someone who refuses to engage in counter-attitudinal behavior may sometimes teach people how to stick to their convictions. Maybe witnessing a refusal to engage in counter-attitudinal behavior can liberate people to act in line with their own moral beliefs as well. This connects to literature on moral elevation that demonstrates that witnessing moral behavior by others can induce moral or pro-social behavior (e.g., Aquino et al., 2011; Pohling & Diessner, 2016; Schnall & Roper, 2012; Schnall et al.; 2010). It may be interesting to systematically investigate under which conditions people learn from the exemplary behavior of the morally courageous.

One potential limitation of our current work is that we did not incorporate hierarchical or power differences. In some situations, speaking up means going against one's superiors. This could, for example, be relevant to the literature on retaliation against whistleblowers. Although it would be in the interest of companies and society at large when people would blow the whistle on organizational misconduct, this often does not happen. And when it does, those who dare to blow the whistle are often targets of negative reactions and retaliation (e.g., Mesmer-Magnus & Viswesvaran, 2005; Yeargain & Kessler, 2009). Investigating reactions to moral courage by people who hold different positions may an interesting avenue for future research.

Another potential limitation concerns the generalizability of the findings. Our samples consisted of participants who were higher educated, relatively young and overwhelmingly heterosexual. It is possible that some of our findings would be different in different populations. For example, people who identify as LGBTQ+ may refuse to engage in such an experimental task to a larger degree than we observed in the current studies, and/or be more self-critical when they do engage in such an experimental task. It would therefore be interesting to explore how our findings hold in other populations.

To conclude, our experiments indicate that behaving in accordance with one's moral beliefs remains a challenge. Most people can be easily persuaded to engage in counter-attitudinal behavior, in some cases even without any repercussions to their personal self-concepts. However, our results also paint a more optimistic picture.

We show that people have positive views of others who refuse to compromise their moral values and uphold these values when others fail to do so. This suggests that negative repercussions for those who defend moral causes may not be as severe as feared.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11211-021-00367-y>.

**Acknowledgements** The research reported in this paper was supported by a Graduate Program research grant from the Netherlands Organization for Scientific Research (NWO, 022.003.040) awarded to the Kurt Lewin Institute. We thank Anneke Bruin, Simone van Oudvorst, and Marian van Sprakelaar for their help with the data collection.

#### Declarations

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Aquino, K., McFerran, B., & Laven, M. (2011). Moral identity and the experience of moral elevation in response to acts of uncommon goodness. *Journal of Personality and Social Psychology, 100*(4), 703–718. <https://doi.org/10.1037/a0022540>
- Aquino, K., & Reed, A. (2002). The self-importance of moral identity. *Journal of Personality and Social Psychology, 83*(6), 1423–1440. <https://doi.org/10.1037//0022-3514.83.6.1423>
- Bateson, M., Nettle, D., & Roberts, G. (2006). Cues of being watched enhance cooperation in a real-world setting. *Biology Letters, 2*(3), 412–414. <https://doi.org/10.1098/rsbl.2006.0509>
- Brambilla, M., Rusconi, P., Sacchi, S., & Cherubini, P. (2011). *European Journal of Social Psychology, 41*, 135–143. <https://doi.org/10.1002/ejsp.744>
- Cohen, G. L., & Sherman, D. K. (2014). The psychology of change: Self-affirmation and social psychological intervention. *Annual Review of Psychology, 65*, 333–371. <https://doi.org/10.1146/annurev-psych-010213-115137>
- Cramwinckel, F. M., Van Dijk, E., Scheepers, D., & Van den Bos, K. (2013). The threat of moral refusers for one's self-concept and the protective function of physical cleansing. *Journal of Experimental Social Psychology, 49*(6), 1049–1058. <https://doi.org/10.1016/j.jesp.2013.07.009>
- Cushman, F., Young, L., & Hauser, M. (2006). The role of conscious reasoning and intuition in moral judgment: Testing three principles of harm. *Psychological Science, 17*(2), 1082–1089. <https://doi.org/10.1111/j.1467-9280.2006.01834.x>
- Dickter, C. L., Kittel, J. A., & Gyurovski, I. I. (2012). Perceptions of non-target confronters in response to racist and heterosexual remarks. *European Journal of Social Psychology, 42*, 112–119. <https://doi.org/10.1002/ejsp.855>
- Diener, E., & Wallbom, M. (1976). Effects of self-awareness on antinormative behavior. *Journal of Research in Personality, 10*(1), 107–111. [https://doi.org/10.1016/0092-6566\(76\)90088-X](https://doi.org/10.1016/0092-6566(76)90088-X)

- Ellemers, N., Pagliaro, S., Barreto, M., & Leach, C. W. (2008). Is it better to be moral than to be smart? The effects of morality and competence norms on the decision to work at group status improvement. *Journal of Personality and Social Psychology*, 95(6), 1397–1410. <https://doi.org/10.1037/a0012628>
- Geen, R. G. (1991). Social motivation. *Annual Review Psychology*, 42, 377–399. <https://doi.org/10.1146/annurev.ps.42.020191.002113>
- Gibbons, F., & Buunk, B. (1999). Individual differences in social comparison: Development of a scale of social comparison orientation. *Journal of Personality and Social Psychology*, 76(1), 129–142. <https://doi.org/10.1037/0022-3514.76.1.129>
- Gonzalez, K. A., Ramirez, J. L., & Paz Galupo, M. (2018). Increase in GLBTQ minority stress following the 2016 US presidential election. *Journal of GLBT Family Studies*, 14(1–2), 130–151. <https://doi.org/10.1080/1550428X.2017.1420849>
- Greene, M., & Low, K. (2014). Public integrity, private hypocrisy, and the moral licensing effect. *Social Behavior and Personality: an International Journal*, 42(3), 391–400. <https://doi.org/10.2224/sbp.2014.42.3.391>
- Greitemeyer, T., Fischer, P., Kastenmüller, A., & Frey, D. (2006). Civil courage and helping behavior: Differences and similarities. *European Psychologist*, 11(2), 90–98. <https://doi.org/10.1027/1016-9040.11.2.90>
- Greitemeyer, T., Osswald, F., & Frey, D. (2007). Civil courage: Implicit theories, related concepts, and measurement. *The Journal of Positive Psychology*, 2(2), 115–119. <https://doi.org/10.1080/17439760701228789>
- Griskevicius, V., Tybur, J. M., & Van den Bergh, B. (2010). Going green to be seen: Status, reputation, and conspicuous consumption. *Journal of Personality and Social Psychology*, 98(3), 392–404. <https://doi.org/10.1037/a0017346>
- Haddock, G., Zanna, M. P., & Esses, V. M. (1993). Assessing the structure of prejudicial attitudes: The case of attitudes toward homosexuals. *Journal of Personality and Social Psychology*, 65(6), 1105–1118. <https://doi.org/10.1037/0022-3514.65.6.1105>
- Halmburger, A., Baumert, A., & Schmitt, M. (2015). Anger as a driving factor of moral courage in comparison with guilt and global mood: A multimethod approach. *European Journal of Social Psychology*, 45, 39–51. <https://doi.org/10.1002/ejsp.2071>
- Hannah, S. T., Avolio, B. T., & Walumbwa, F. O. (2011). Relationships between authentic leadership, moral courage, and ethical and pro-social behaviors. *Business Ethics Quarterly*, 21(4), 555–578. <https://doi.org/10.1016/j.leaqua.2011.12.004>
- Hirt, E. R., McCrea, S. M., & Kimble, C. E. (2000). Public self-focus and sex differences in behavioral self-handicapping: Does increasing self-threat still make it “just a man’s game”? *Personality and Social Psychology Bulletin*, 26(9), 1131–1141. <https://doi.org/10.1177/01461672002611009>
- Kite, M. E., & Deaux, K. (1986). Attitudes toward homosexuality - assessment and behavioral consequences. *Basic and Applied Social Psychology*, 7(2), 137–162. [https://doi.org/10.1207/s15324834basp0702\\_4](https://doi.org/10.1207/s15324834basp0702_4)
- Leach, C. W., Ellemers, N., & Barreto, M. (2007). Group virtue: The importance of morality (vs. competence and sociability) in the positive evaluation of in-groups. *Journal of Personality and Social Psychology*, 93(2), 234–249. <https://doi.org/10.1037/0022-3514.93.2.234>
- Leary, M. R. (1983). A brief version of the fear of negative evaluation scale. *Personality and Social Psychology Bulletin*, 9(3), 371–375. <https://doi.org/10.1177/0146167283093007>
- Marques, J. M., & Yzerbyt, V. Y. (1988). The black sheep effect: Judgmental extremity towards ingroup members in inter- and intra-group situations. *European Journal of Social Psychology*, 18, 287–292. <https://doi.org/10.1002/ejsp.2420180308>
- Mazar, N., Amir, O., & Ariely, D. (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research*, 45(6), 633–644. <https://doi.org/10.1509/jmkr.45.6.633>
- Mesmer-Magnus & Viswesvaran. (2005). Whistleblowing in organizations: An examination of correlates of whistleblowing intentions, actions, and retaliation. *Journal of Business Ethics*, 62, 277–297. <https://doi.org/10.1007/s10551-005-0849-1>
- Minson, J. A., & Monin, B. (2012). Do-gooder derogation: Disparaging morally motivated minorities to defuse anticipated reproach. *Social Psychological and Personality Science*, 3(2), 200–207. <https://doi.org/10.1177/1948550611415695>
- Monin, B. (2007). Holier than me? threatening social comparison in the moral domain. *Revue Internationale De Psychologie Sociale-International Review of Social Psychology*, 20(1), 53–68.

- Monin, B., Sawyer, P. J., & Marquez, M. J. (2008). The rejection of moral rebels: Resenting those who do the right thing. *Journal of Personality and Social Psychology*, 95(1), 76–93. <https://doi.org/10.1037/0022-3514.95.1.76>
- Netherlands Institute for Social Research (SCP). (2013). *Acceptance of lesbian, gay, bisexual and transgender individuals in the Netherlands 2013*. Retrieved from [http://www.scp.nl/english/Publications/Publications\\_by\\_year/Publications\\_2013/Acceptance\\_of\\_lesbian\\_gay\\_bisexual\\_and\\_transgender\\_individuals\\_in\\_the\\_Netherlands\\_2013](http://www.scp.nl/english/Publications/Publications_by_year/Publications_2013/Acceptance_of_lesbian_gay_bisexual_and_transgender_individuals_in_the_Netherlands_2013)
- Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 17, 776–783. <https://doi.org/10.1037/h0043424>
- Pohling, R., & Diessner, R. (2016). Moral elevation and moral beauty: A review of the empirical literature. *Review of General Psychology*, 20(4), 412–425.
- Ritov, I., & Baron, J. (1992). Status-quo and omission biases. *Journal of Risk and Uncertainty*, 5, 49–61.
- Schnall, S., & Roper, J. (2012). Elevation puts moral values into action. *Social Psychological and Personality Science*, 3(3), 373–378. <https://doi.org/10.1177/1948550611423595>
- Schnall, S., Roper, J., & Fessler, D. M. T. (2010). Elevation leads to altruistic behavior. *Psychological Science*, 21(3), 315–320. <https://doi.org/10.1177/0956797609359882>
- Schwartz, S. H. (1992). Universals in the content and structure of values: Theory and empirical tests in 20 countries. In M. Zanna (Ed.), *Advances in experimental social psychology*. (Vol. 25, pp. 1–65). Academic Press.
- Skitka, L. J. (2010). The psychology of moral conviction. *Social and Personality Psychology Compass*, 4, 267–281. <https://doi.org/10.1111/j.1751-9004.2010.00254.x>
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. (5th ed.). Pearson.
- Van den Bos, K., Bruins, J., Wilke, H., & Dronkert, E. (1999). Sometimes unfair procedures have nice aspects: On the psychology of the fair process effect. *Journal of Personality and Social Psychology*, 77(2), 324–336. <https://doi.org/10.1037/0022-3514.77.2.324>
- Van de Ven, N., Archer, A. T. M., & Engelen, B. (2019). More important and surprising actions of a moral exemplar trigger strong admiration and inspiration. *The Journal of Social Psychology*, 159(4), 383–397. <https://doi.org/10.1080/00224545.2018.1498317>
- Yeargain, J. W., & Kessler, L. L. (2009). Organizational hostility toward whistleblowers. *Journal of Legal, Ethical and Regulatory Issues*, 13(1), 87–92.
- Zhong, C. B., Liljenquist, K., & Cain, D. M. (2009). Moral self-regulation. Licensing and compensation. In D. De Cremer (Ed.), *Psychological perspectives on ethical behavior and decision making*. (pp. 75–43). Charlotte, NC: Information Age Publishing.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.