von Grundherr, Michael; Jauernig, Johanna; Uhl, Matthias

To condemn is not to punish: An experiment on hypocrisy

Games

Provided in Cooperation with:
MDPI – Multidisciplinary Digital Publishing Institute, Basel

Suggested Citation: von Grundherr, Michael; Jauernig, Johanna; Uhl, Matthias (2021) : To condemn is not to punish: An experiment on hypocrisy, Games, ISSN 2073-4336, MDPI, Basel, Vol. 12, Iss. 2, pp. 1-13, https://doi.org/10.3390/g12020038

games

MDPI

Article

To Condemn Is Not to Punish: An Experiment on Hypocrisy

Michael von Grundherr [1], Johanna Jauernig [2] and Matthias Uhl [3,*]

[1] Research Center for Neurophilosophy and Ethics of Neuroscience, LMU Munich, 80539 Munich, Germany; mvg@lrz.uni-muenchen.de
[2] Leibniz Institute of Agricultural Development in Transition Economies (IAMO), 06120 Halle (Saale), Germany; jauernig@iamo.de
[3] Faculty of Computer Science, Technische Hochschule Ingolstadt, 85049 Ingolstadt, Germany
* Correspondence: matthias.uhl@thi.de

Abstract: Hypocrisy is the act of claiming moral standards to which one's own behavior does not conform. Instances of hypocrisy, such as the supposedly green furnishing group IKEA's selling of furniture made from illegally felled wood, are frequently reported in the media. In a controlled and incentivized experiment, we investigate how observers rate different types of hypocritical behavior and if this judgment also translates into punishment. Results show that observers do, indeed, condemn hypocritical behavior strongly. The aversion to deceptive behavior is, in fact, so strong that even purely self-deceptive behavior is regarded as blameworthy. Observers who score high in the moral identity test have particularly strong reactions to acts of hypocrisy. The moral condemnation of hypocritical behavior, however, fails to produce a proportional amount of punishment. Punishment seems to be driven more by the violation of the norm of fair distribution than by moral pretense. From the viewpoint of positive retributivism, it is problematic if neither formal nor informal punishment follows moral condemnation.

Keywords: corporate hypocrisy; punishment; moral judgment; experimental ethics; behavioral ethics; moral identity

Citation: von Grundherr, M.; Jauernig, J.; Uhl, M. To Condemn Is Not to Punish: An Experiment on Hypocrisy. *Games* **2021**, *12*, 38. https://doi.org/10.3390/g12020038

Academic Editors: Rainer Michael Rilke, Stefania Bortolotti and Ulrich Berger

Received: 29 January 2021
Accepted: 19 April 2021
Published: 26 April 2021

> *"Today we have to learn all over again that love for the sinner and love for the person who has been harmed are correctly balanced if I punish the sinner in the form that is possible and appropriate."*

Pope Benedict XVI, Light of the World: The Pope, the Church, and the Signs of the Times—A Conversation with Peter Seewald

1. Introduction

Hypocrisy consists of the exercise of moral pretense: the claim of moral standards that one's own behavior does not withstand. The literature on ethics often seems to assume that people condemn hypocrisy on moral grounds. Isserow and Klein [1] (p. 192) state that "hypocrites tend to invite moral opprobrium—we condemn them, and usually quite harshly". Cases of hypocrisy are regularly reported in the mass media. In 2020, IKEA was accused of selling furniture made of wood which was illegally felled in the Ukrainian Carpathians, while relying on the Forest Stewardship Council as the world's leading green labelling system for timber [2]. However, although the behavior of IKEA provoked criticism in the media and from NGOs such as Earthsight, doubts may be raised as to whether customers actually punished the company through boycotts. After all, a press release from October announced, "strong IKEA retail sales of EUR 35.2 billion for the financial year 2020, despite the economic and public health challenges posed by COVID-19" [3]. Although comparisons of a company's sales, before and after accusations of hypocrisy, may serve as an index of its condemnation, sales figures are usually confounded by several other factors such as the COVID-19 pandemic, thereby rendering them a poor measure of retributive action or sentiment.

Previous studies have focused on the psyche of the individual hypocrite, conducting research on the psychological mechanisms underlying hypocritical behavior. Much less empirical attention has been paid to the moral evaluation of hypocrisy, which is a particularly compelling subject, from an ethical perspective. This study's first aim was therefore to develop an experiment to test observers' reactions to hypocrisy in a controlled context. Given that we were more interested in ethical judgements based on conscious deliberation or reflections than on responses produced by emotional bias, we analyzed the attitudes of impartial observers who were not affected personally by the salient hypocrisy.

Moral outrage becomes most relevant upon its eventual operationalization into punishment. Wrongdoers are much more likely to adapt their behavior consequent to punishment than the mere experience of being blamed (see, e.g., [4]). Whether people tend to punish behavior that they find morally reprehensible is therefore important, because a reluctance to punish behavior regarded as morally inappropriate may support an argument to fill this gap institutionally. (Likewise, institutionalized punishment may render individual punishment obsolete.) Furthermore, there are instances of norm violation in which no justiciable offense is present; therefore, punishment can only be exercised by stakeholders. Various examples of norm violation that are not illegal can be found in white-collar crime. Practices such as sophisticated ways of tax evasion or even bribery are not always a criminal offence but are clearly condemned by society (for a discussion, see [5]). In these instances, punishment cannot be "outsourced" to judicial institutions. Arguably, this could be so in the cited case of IKEA, where a negligence in the monitoring of the supply chain may not constitute clear-cut illegal conduct. In practice, punishment is usually costly, which means that there is often a motivational hurdle that must be surmounted to impose punishment. Those who are outraged, however, may still be reluctant to punish, even if punishment does not cost them much. One reason could be that they do not want to perceive themselves as vindictive or do not want to be perceived that way by others. In this paper, we seek to test whether moral blame translates proportionally into punishment in the absence of a motivational hurdle, i.e., if punishment is not costly, because this constitutes a conservative upper bound for the inclination to punish morally reprehensible behavior.

Ryberg [6] (p. 12) emphasizes the "intuitive appeal" of the moral claim that punishment is proportionate to crime. This intuitive claim is substantiated by theories of retributive justice. These theories are often based on Kant's notion of justice according to which perpetrators are moral agents whose moral agency and autonomy must be acknowledged and hence, a measured response is in order [7]. This measured response means that the punishment must be in proportion to what was done, and the right kind of punishment is neither too harsh, nor too lenient. Mackie [8] discusses a retributive theory of punishment that is explicitly distinct from ethical theories that justify punishment in terms of its desirable deterrent effect on the punished individual (special prevention) and on potential copycats (general prevention). In this line of argument, he introduces the principle of positive retributivism, which claims that one who is guilty ought to be punished, even if, for instance, no one ever finds out about the punishment. Thus, for positive retributivists, the observable gap between judgment and punishment constitutes an intrinsic ethical defect. According to Mackie, although the principle of positive retributivism may seem less compelling than its negative variant (which asserts that one who is not guilty must not be punished), it bears immediate appeal and apparent authority.

So far, there is little empirical research investigating the proportionality of punishment and blame. What has been studied is moral evaluation of deception and honesty [9,10], punishment of deception and honesty [11], and reward and punishment of honesty and deception [12]. These authors do find asymmetries between reward and punishment, yet their interpretations are not without criticism (see, e.g., [13]). To fill this research gap, it is the key contention of this paper to better understand how different forms of hypocrisy are morally evaluated and punished.

This paper proceeds as follows: in the second section, we provide some theoretical background on the relevance of self-deception and moral identity for the evaluation and

punishment of hypocrisy; we outline our research questions in the third section; in the fourth section, we discuss our experimental setup; results are presented in the fifth section; and the sixth section represents the conclusion of the paper.

2. Theoretical Background: The Role of Self-Deception and Moral Identity in Judging Hypocrisy

Several authors have outlined the importance of self-deception for hypocritical behavior. A prominent strand in the classical psychological literature on hypocrisy explains instances of hypocrisy as being motivated by self-deceptive attempts to maintain a consistent and favorable self-image [14]. In a series of experiments using the "coin paradigm", Batson and colleagues gave participants an allocation task, for which they could decide in a fair or a selfish way. To help make the allocation, a coin toss was introduced as a fairness norm. It was absolutely clear to the deciders that affected third parties would never learn how the allocation was made (by a fair coin toss or selfishly). Yet, the vast majority of deciders pretended to have followed the coin toss (by fiddling with the coin, tossing it multiple times, until the desired result was achieved, etc.). Due to the experimental setup, no one could be deceived by the coin fiddling as it was the private knowledge of the coin fiddler. Therefore, only the decider—him- or herself—could be affected by means of self-deception. Additionally, the use of the coin in a deceptive way did in fact make a difference for deciders, as was shown by a set of post-experimental questions, in which their moral self-regard was elicited. Deciders which used the coin but rigged the result so that it profited themselves felt significantly more moral than those who did not use the coin but decided selfishly. Thus, even though the fairness norm was violated, by using a device that represents the fairness norm (the coin), participants felt better about themselves. Hence, Batson and colleagues conclude that self-deception is at work. In the economics literature, a similar concept of self-deception has been used and experimentally studied, e.g., by Gneezy et al. [15], Dana et al. [16], and many others.

Furthermore, self-deception may facilitate the deception of others and thereby the hypocritical communication on which such deception relies. The evolutionary biologist Robert Trivers argued that self-deception emerged in service of deluding others, effectively reducing "the subtle signs of self-knowledge that may give us away" [17].

The causal link between self-deception and hypocrisy may lead to a regular co-occurrence of self-deception and hypocrisy. In the coin paradigm studies, both terms, hypocrisy and self-deception, coincide. They both fulfill the definition used by Batson and colleagues, according to which moral hypocrisy consists of appearing moral without paying the costs. This definition includes instances where the psychological costs of *perceiving oneself* as a norm violator, and also the psychological and maybe also physical costs of being *perceived by others* as a norm violator. In what follows, we refer to the former as self-deception and to the latter as hypocrisy.

In recent studies, the coin paradigm was modified to disentangle whether participants who violated a fairness norm were motivated by their internal self-image, or by managing the impression they make on others. Findings revealed that experimental subjects are motivated mainly to manipulate the impression they make on others, and not their internal self-image [18].

If we want to better understand the moral evaluation of hypocrisy, the empirical findings described above beg the question to which extent self-deception induces moral resentment. Following Trivers' reasoning, self-deception can be regarded as indirectly harmful to others, because it makes the act of deceiving others easier, by enabling the hypocrite to cover his moral transgressions and thereby avoid socially desirable punishment. Various philosophers believe self-deception corrupts the conscience or threatens moral agency [19,20]. For instance, Darwall [21] (pp. 424–425) argues that it "threatens the very capacity for [moral] judgment". Moreover, it may be indicative of a weak will, which renders the agent likelier to break moral norms to gain personal advantages.

However, it is implausible that observers rate self-deceit on the same level as full-blown hypocrisy, as it is unlikely to qualify as an equally obvious moral evil. People

may also be doubtful about the degree of responsibility that should be borne by self-deceivers. Some philosophers, mainly those who hold the view that self-deceivers are not intentionally deceiving themselves, argue against automatically blaming self-deceivers for their state, as self-deception is often a subconscious process that can be hard to detect and control [22]. Finally, self-deception may also be considered indicative of (albeit weak) moral self-awareness, because a need to deceive oneself only arises if the agent, at least on some level, is aware of the wrongness of his or her actions [23]. Reasoning along these lines may even be positively acknowledged by others.

As self-deception may or may not in itself trigger moral resentment, and may co-occur with hypocritical behavior, it is important to factor out the potential influence of perceived self-deception on moral evaluation to be able to identify the reactions that hypocrisy alone triggers. We therefore compare moral reactions to full-blown hypocrisy (which may include instances of self-deception) and to pure self-deception.

Given that we are explicitly interested in people's perception of hypocrisy, it seems worthwhile to investigate whether varying degrees of moral sensitivity can account for disparate moral evaluation. Ample evidence suggests that people's moral judgments and behaviors vary in intensity, because of differences in dispositional factors. As hypocrisy is not attributable to a specific domain of attitudes (such as social, environmental, etc.), we account for such considerations by testing observers' moral consciousness more generally. More precisely, we measured the self-importance of moral identity [24], which has been shown to act as an important moderating factor in many moral contexts. Identity theory assumes that people are motivated to achieve consistency between the perceptions of their own behavior and their identity standards [25].

A high level of self-importance of moral identity is also likely to trigger more socially appropriate and stricter moral judgment [26]. We therefore include a moral identity measure in our analysis to determine the extent to which self-importance of moral identity impacts the moral evaluation and punishment of hypocrisy. We assume that people with a stronger moral identity react more strongly to hypocrisy, evaluate hypocrites more negatively, and punish them more harshly than people with less pronounced moral identities.

3. Research Questions

In the context of our study, we seek to answer three research questions. First, we are interested in a comparison between the moral evaluation of hypocrisy and self-deception on the one hand and open egoism on the other. It also seems interesting whether mere self-deception, where external victims are absent, is regarded as morally inferior to open egoism. This is due to the important role that self-deception plays as a prerequisite for hypocrisy according to the psychological literature.

Research Question 1. *Is hypocrisy considered more reprehensible than self-deception, which in turn is considered more reprehensible than openly egoistic behavior?*

Second, we ask whether hypocrisy and self-deception are punished more severely than openly egoistic behavior. This question seems particularly relevant, if there is to be a prospect of stimulating a change in the future behavior of hypocrites, because moral condemnation that lacks sanctioning consequences is unlikely to induce these changes. We also ask how full-blown hypocrisy compares to mere self-deception in terms of punishment.

Research Question 2. *Is hypocrisy punished more severely than self-deception, which is in turn punished more severely than openly egoistic behavior?*

Finally, we ask whether evaluation and punishment of hypocrisy and self-deception are fully proportionate. Although we may find that hypocrisy and self-deception are evaluated as being morally worse, thereby prompting harsher punishment than openly egoistic behavior, the level of punishment may still fall short of the level of condemnation.

Research Question 3. *Is the punishment of hypocrisy fully proportionate to its moral evaluation?*

4. Experiment Design

The experiment consisted of two parts. In the first part, we elicited our respondents' moral evaluations and actual punishments of certain types of behavior that lab participants would later be able to display. In the second part, we made respondents' punishment decisions consequential by actually applying them to our lab participants. Respondents were able to be sure that this would be the case. The first part was conducted online with SoSci Survey [27] and made available to the participants on www.soscisurvey.com. Participants of both parts of the experiment were informed ex ante about the general structure of the experiment and no one was deceived.

In this first part, there were 380 participants, 344 of whom finished the study, of which 216 were female. Their average age was 38.41 years (sd = 13.63). Thirty-two percent of participants were students. Participants were randomly assigned to the hypocrisy condition (176 participants) or to the self-deception condition (168 participants). Shortly after the conclusion of the first part of the experiment, the second part was conducted in the economic research lab of a German university. Participants were undergraduate students from various disciplines. Thirty-two participants—16 in each condition—took part.

Respondents' task was to evaluate the behavior of lab participants. Respondents learned that lab participants would be paired up randomly and play a dictator game, in which a randomly determined dictator receives 80 experimental currency units (ECU, 1 ECU = 10 Eurocent) that he or she can share with a randomly determined recipient. The dictator can give the recipient 20 ECU (Option 1) or 60 ECU (Option 2) and keep the rest. For our analysis, we labeled those who selected Option 1 "egoists", because they kept the larger share for themselves, and those who went for Option 2 "altruists", because they kept the smaller share for themselves.

An Option 3 existed, wherein the dictator was permitted to toss a (virtual) coin to decide whether the receiver receives 20 or 60 ECU. Respondents learned that a dictator who selected Option 3 was unexpectedly confronted with three further options concerning the handling of the tossed coin. First, the dictator had the option to disclose the outcome of the coin toss and carry out the consequences as tossed (Option 3a). We label the type of individual who selects Option 3a "fair", because he or she actually gives him- or herself and the recipient equal chances of receiving the higher payoff.

Second, a dictator who had opted for the coin now unexpectedly had the opportunity to ignore the coin (which may be unfavorable to him or her) and just give 20 ECU to the recipient (Option 3b). We label a dictator who selects Option 3b a "selfish deceiver", as he or she pretends to make a fair allocation while, in fact, being certain that he or she can keep the larger share of money for him- or herself. Finally, a dictator who had opted for the coin now unexpectedly had the opportunity to ignore the coin (which may be unfavorable to the recipient) and just give 60 ECU to the recipient (Option 3c). Prima facie, this behavior may seem downright strange, as it conceals an altruistic act. One potential motivation for such behavior, however, may be to spare the recipient the uncomfortable feeling of having accepted a charitable gift. We therefore call this type an "altruistic deceiver".

In Options 3b and 3c, participants did *not* observe the outcome of the coin toss to then decide whether to go with it or overrule it. Instead, after having chosen the coin to determine the distribution, they were presented with a surprise stage, in which they found out that even though they had chosen the coin, now they still had the option to ignore the potentially unfavorable coin altogether and distribute the money directly. We implemented the coin toss in this way for the following reason: we wanted to avoid eliciting moral luck or lack thereof. If participants had tossed the coin, had waited for the result and had then overruled it, we would not have known what had happened in those cases in which participants had benefitted from the coin toss result. Either the participant could have been fair and stuck with the coin result, no matter what the result was, or he or she could have waited to see whether the coin showed the desired result, and if not, would have changed the result by neglecting the result of the coin toss. In all cases of a favorable coin toss result, it would have been unclear whether the participants would have stuck with the outcome or

not. Yet, it was crucial for our study to elicit moral evaluation and punishment of behavior that could be clearly identified while not losing too many observations.

It is worth noting that, in the experiment, the respective types were neutrally labeled as Type A to E to avoid moral priming. Furthermore, the fair option of the coin toss was made salient to the participants by a sentence indicating that most people regard using the coin as fair [14]. The experimental conditions that we implemented are a modification of those used by Lönnqvist et al. [18], who in turn based their study on that of Batson et al. [14].

In the hypocrisy condition, the recipient ultimately learned whether the dictator opted for the coin to attribute the amounts. The recipient, however, did not learn whether the dictator felt committed to his or her choice of the coin and actually let the coin decide the distribution or not, when presented with the additional options. In this condition, a dictator choosing Option 3b fulfils the definition of a hypocrite, as he or she publicly feigns use of a fair procedure, while factually making an egoistic decision behind the curtain.

In the self-deception condition, the recipient only learned about the ultimate distribution of the 80 ECU but remained completely ignorant as to whether the dictator had opted for the coin or not. Therefore, a participant who chooses the coin in the first place but then decides directly in a selfish (or altruistic) manner can only deceive him- or herself about the fairness of the action just like in the original coin paradigm by Batson and colleagues. Whatever reason he or she might come up with to justify the behavior, it can only be targeted at him- or herself, because no one was ever informed about the decision to take the coin in the first place.

Thus, the two conditions diverged only in the description of recipients' information about the dictators' use or waiving of the coin toss. Respondents were explicitly told either that recipients were informed about whether dictators had used the coin (hypocrisy condition) or whether they stayed ignorant of that fact (self-deception condition). Otherwise, descriptions were identical.

The participants in the online part were then asked to evaluate how good or bad the behavior of the egoist, the altruist, the fair, the hypocrite/self-deceiver and the altruistic deceiver was on a continuous scale from 0 to 100. After that, we elicited participants' punishment decisions for each behavioral type. Based on Falk et al. [28], punishment consisted of imposing tedious tasks. In line with Falk et al. [28], the tedious task consisted of a 150-field matrix consisting of the digits zero and one. Participants had to count how many ones there were in the matrix. The task was solved, when the correct number was typed in. Just as in the evaluation, participants could administer the punishment on a continuous scale, which for reasons of partibility ranged from 0 to 10. Participants were informed that one of them was randomly drawn and his or her punishment decisions were implemented, by administering the chosen task number for each behavioral type to real lab participants, who had acted out the respective behavior. Note, that we intentionally implemented punishment free of charge. Since the evaluating participants had no stakes in the future behavioral decisions, by making punishment costly to them, we could not expect proportionality of blame (without costs) and punishment (with costs). We would, most likely, only have observed the well-researched difference between stated and revealed preferences. Yet, our aim was to study the proportionality of blame and punishment.

Finally, respondents completed the Self-Importance of Moral Identity Questionnaire (SMI-Q). Participants attained a mean score of 6.01 (sd = 0.83) on the internalization scale of the SMI-Q. To analyze the effects of this score, we split the sample at the median (6.20).

5. Results

5.1. Power Analysis

In our design, detecting differences between the evaluation of behaviors requires within-subject comparisons, while detecting differences in the evaluation of hypocrisy vs. self-deception requires comparisons between subjects in different experimental groups. Based on pilot data, we consider relevant differences in the standardized means to be

around 0.33 (10 scale points at $sd = 30$ in the case of evaluation, 1 point at $sd = 3$ in the case of punishment).

A power analysis with the R-package Superpower [29] that accounts for multiple hypothesis testing was conducted for a 5 (type of behavior) × 2 (information condition) × 2 (moral identity) ANOVA. It shows that a power of 0.80 requires a sample size of at least 75 per cell (300 in total), if we want to detect (a) an omnibus main effect of behavior, (b) an omnibus interaction effect between behavior and information condition and (c) an omnibus interaction effect of behavior and moral identity. A further analysis with the R-package pwr [30] indicates that a sample-size of 75 (one-sample, two-sided *t*-test) or 150 (two-sample, two-sided *t*-test) per group is necessary to reach a power of 0.80 in selective post hoc tests required by our research questions.

5.2. Moral Evaluation

The analysis examined the effects of behavior type (egoist, altruist, fair, hypocrite/self-deceiver, altruistic deceiver), information condition (hypocrisy, self-deception) and self-importance of moral identity (high, low) on moral evaluation. Thus, we analyzed differences in judgments using a 5 (type of behavior) × 2 (information condition) × 2 (moral identity) repeated measures ANOVA, with type of behavior as the repeated measure. The analysis yielded a large main effect for type of behavior ($F(4, 1360) = 133.00$, $p < 0.001$, $\eta_p^2 = 0.28$) and a small effect for information condition ($F(1, 340) = 6.55$, $p = 0.011$, $\eta_p^2 = 0.018$). A significant, albeit small, interaction effect also occurred between type of behavior and moral identity score ($F(4, 1360) = 3.48$, $p = 0.008$, $\eta_p^2 = 0.01$). Figure 1 shows the moral evaluations of the five decision types in both conditions.

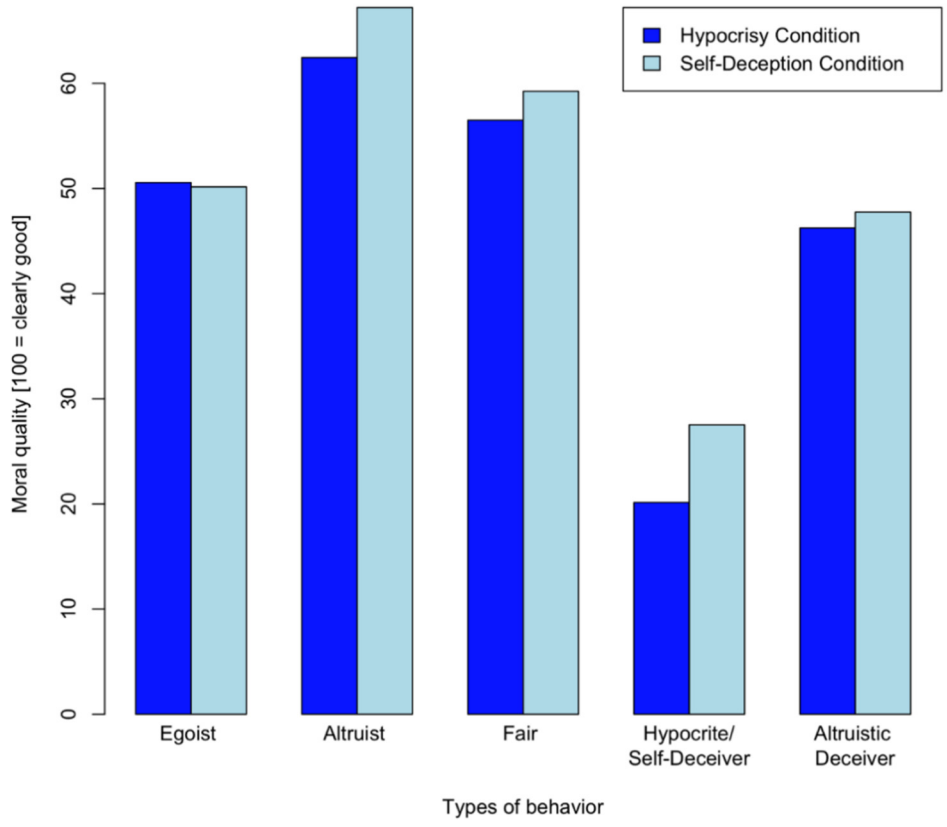

Figure 1. Evaluation of decision types.

Post hoc tests (see Table 1) revealed that hypocrisy and self-deception were determined to be the worst of all decision types. Open egoism was evaluated as being significantly worse than altruistic behavior in both conditions and significantly worse than fair behavior in the self-deception condition. Participants with high moral identity scores generally

followed the same pattern, but the hypocrisy condition reflected two differences. First, a marginally significant trend indicated that they may evaluate hypocrites even more negatively than other participants (*mean* = 16.88, *sd* = 16.63 vs. *mean* = 22.045, *sd* = 19.80, *t* (152.97) = 1.852, *p* = 0.066). Second, they made a clear distinction between the evaluation of the fair agent and the evaluation of the egoistic agent in the hypocrisy condition, but did not clearly rate the altruistic agent as being better than the fair one.

Table 1. Evaluation of decision types.

	Hypocrisy Condition		**Self-Deception Condition**	
	All	**High MI**	**All**	**High MI**
Behavior	mean (sd)	mean (sd)	mean (sd)	mean (sd)
Egoist	50.57(23.50) ^{a,d}	44.98(25.20) ^{a}	50.15(26.16) ^{a}	48.58(29.98) ^{a}
Altruist	62.45(26.88) ^{b}	64.37(27.43) ^{b}	67.12(25.21) ^{b}	72.74(27.51) ^{b}
Fair	56.49(27.23) ^{a,b}	62.83(26.50) ^{b}	59.24(27.32) ^{c}	59.68(28.41) ^{c}
Hypocrite/Self-Deceiver	20.14(18.81) ^{c}	16.88(16.63) ^{c}	27.51(19.41) ^{d}	28.63(21.35) ^{d}
Altruistic Deceiver	46.24(29.05) ^{d}	45.84(30.33) ^{a}	47.76(27.54) ^{a}	52.36(29.65) ^{a}

Note. MI = moral identity score. Values in one column with different superscripts differ significantlyin post hoc test at the 0.05 level (paired *t*-test with holm adjustment).

Next, we compared the moral evaluations of hypocrisy and self-deception (between conditions). Hypocrisy was evaluated more negatively than self-deception (*mean* = 20.14, *sd* = 18.81 vs. *mean* = 27.51, *sd* = 19.41, $t(339.93) = -3.58$, $p < 0.001$, Cohen's $d = 0.39$). Note that this difference in moral evaluation is not a general trend: any other differences in moral evaluation between the two conditions (except for the evaluation of open altruists) are insignificant.

Result 1. *Full-blown hypocrisy was considered morally worse than mere self-deception, which in turn was considered morally worse than openly egoistic behavior.*

Further explorative tests revealed that the finding that deceiving others is seen more critically than deceiving oneself is mainly driven by participants with high moral identity (moral identity score above the sample median) because they caused a significant and medium–large difference between the moral evaluation of the hypocrite and the self-deceiver (*mean* = 16.88, *sd* = 16.63 vs. *mean* = 28.63, *sd* = 21.35, $t(132.24) = -3.61$, $p < 0.001$, Cohen's $d = 0.61$), but participants with lower moral identity made a small and insignificant difference (*mean* = 22.04, *sd* = 19.80 vs. *mean* = 26.68, *sd* = 17.88, $t(204.6) = -1.77$, $p = 0.078$, *Cohen's d* = 0.24). People who have a higher sense of their own moral identity, therefore, seem more concerned about the deception of others than the deception of oneself.

5.3. Punishment

The assignment of tedious tasks to a decision maker served as our proxy for punishment. Figure 2 shows the number of tedious tasks assigned to the decision types in both conditions. At first, it may seem odd that the altruist and the fair type were punished at all. In fact, throughout various experimental paradigms, a base amount of antisocial behavior is found. This was most clearly shown by Abbink and Sadrieh [31] with their joy-of-destruction game. In this experiment, participants had no monetary stakes whatsoever in the game, but still in roughly 40% of decisions, participants caused detriment to other participants, just because they could. Similar results have been replicated by, e.g., Jauernig et al. [32,33] and Jauernig and Uhl [34]. These findings show that antisocial behavior is something that frequently occurs in lab experiments; therefore, we should interpret differences between ascribed punishment and not the absolute amount of punishment ascribed to each behavior type. The analysis examined the effects of behavior type (egoist, altruist, fair, hypocrite, altruistic deceiver), information condition (hypocrisy, self-deception) and self-importance of moral identity (high, low) on the level of punishment. Thus, we analyzed differences in judgments using a 5 (type of behavior) × 2 (information

condition) × 2 (moral identity) repeated measures ANOVA with type of behavior as the repeated measure. The analysis yielded a large main effect for type of behavior ($F(4, 1360) = 179.25$, $p < 0.001$, $\eta_p^2 = 0.35$). The omnibus effect of information condition was not significant. There was a small but significant interaction effect between information condition and behavior ($F(4, 1360) = 4.72$, $p < 0.001$, $\eta_p^2 = 0.14$) and a small interaction effect between type of behavior and moral identity score ($F(4, 1360) = 10.86$, $p < 0.001$, $\eta_p^2 = 0.031$).

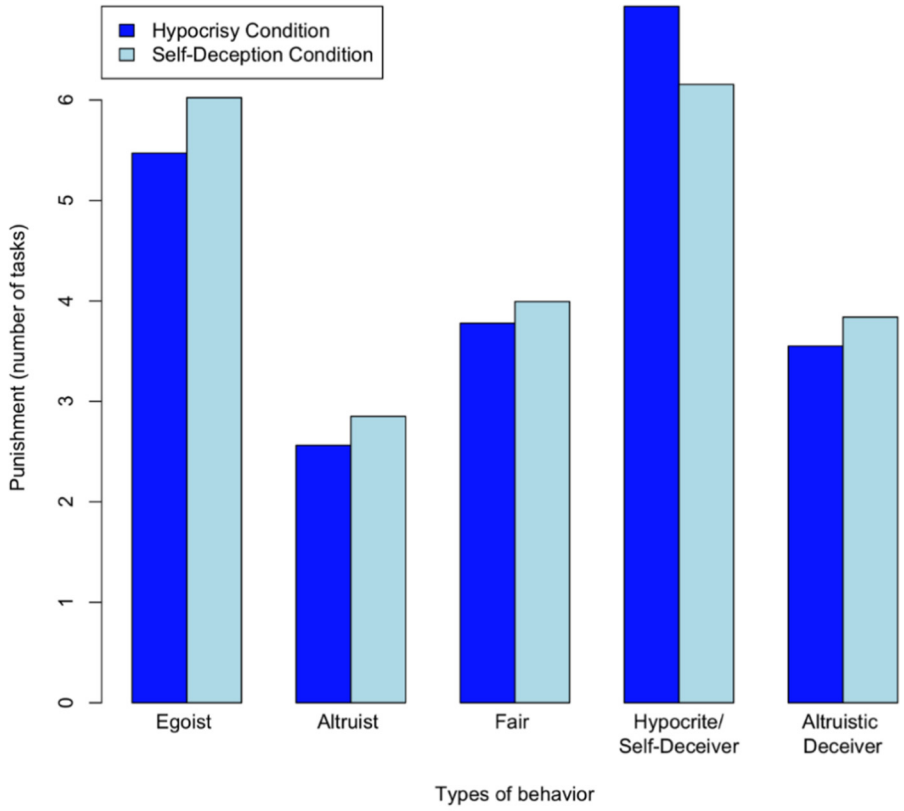

Figure 2. Punishment of decision types (assignment of tedious tasks).

Post hoc tests (see Table 2) revealed that hypocrisy was punished more strongly than all other types of behavior in the hypocrisy condition. In particular, hypocrisy was punished more rigidly than open egoism. Open egoism and self-deception, however, did not receive significantly different punishments in the self-deception condition.

Table 2. Punishment of decision types (assignment of tedious tasks).

	Hypocrisy Condition		**Self-Deception Condition**	
	All	**High MI**	**All**	**High MI**
Behavior	mean (sd)	mean (sd)	mean (sd)	mean (sd)
Egoist	5.47(3.02) ^{a}	6.46(3.20) ^{a}	6.02(3.08) ^{a}	6.71(3.20) ^{a}
Altruist	2.56(2.45) ^{b}	2.60(2.65) ^{b}	2.85(2.73) ^{b}	2.50(2.51) ^{b}
Fair	3.78(2.89) ^{c}	3.51(3.17) ^{c}	3.99(2.91) ^{c}	3.82(2.67) ^{c}
Hypocrite/Self-Deceiver	6.93(2.94) ^{d}	7.58(2.62) ^{d}	6.15(2.92) ^{a}	6.43(2.99) ^{a}
Altruistic Deceiver	3.55(2.84) ^{c}	3.58(2.88) ^{c}	3.84(2.89) ^{c}	4.43(2.75) ^{c}

Note. MI = moral identity score. Values in one column with different superscripts differ significantly in post hoc test at the 0.05 level (paired *t*-test with holm adjustment).

Altruistic deception was not punished more strongly than fair behavior in either condition. Given that altruistic deception was evaluated more negatively than fairness,

punishment seems to be driven less strongly than moral evaluation by the mere involvement of deception.

Hypocrisy was punished more severely than self-deception (*mean* = 6.93, *sd* = 2.94 vs. *mean* = 6.15, *sd* = 2.92, $t(341.48) = 2.46$, $p = 0.014$, *Cohen's d* = 0.26). Note that this difference in punishment is not a general trend, because the degree of punishment is generally higher in the self-deception condition. These findings indicate that deceiving others is punished more strongly than deceiving oneself. The milder moral evaluation of self-deceivers, relative to hypocrites, therefore, also results in milder punishment.

Result 2. *Full-blown hypocrisy was punished more than mere self-deception, which in turn was punished more than openly egoistic behavior.*

Further explorative tests revealed that the latter effect seems to be driven by the decisions of participants with high self-importance of moral identity. These participants identified a clear and significant difference between the punishment meted out to the hypocrite and that meted out to the self-deceiver (*mean* = 7.58, *sd* = 2.62 vs. *mean* = 6.43, *sd* = 2.99, $t(134.87) = 2.40$, $p = 0.017$, Cohen'*sd* = 0.41), but participants who placed a lower value on their moral identity identified a smaller and less significant difference (*mean* = 6.55, *sd* = 3.07 vs. *mean* = 5.94, *sd* = 2.86, $t(203.77) = 1.46$, $p = 0.146$, Cohen'*sd* = 0.20). As already revealed in the moral evaluation pattern, people for whom moral identity is more central are more concerned about punishing those who deceive others than those who deceive themselves.

We find that more negative evaluation of hypocrites, relative to self-deceivers, is also reflected in harsher punishment. This effect is moderated by people with a strong moral identity.

5.4. Drivers of Moral Evaluation and Punishment

In a final step, we attempted to analyze various drivers of moral evaluation and punishment in the hypocrisy condition. To situate the results in an interpretable quantitative relationship, we recoded the behavior types according to Table 3, pitting a process quality (false moral pretense by the dictator) against an outcome quality (expected payoff distribution). This recoding enabled us to determine whether punishment was more strongly influenced by the objective fairness of the outcome than by moral evaluation, which was what the pilot data had suggested. In addition, we centered (z-transformed) the variables "payoff" and number of "tasks" as a gradual measure of punishment, to make the results comparable. We built multinomial linear regression models to predict moral evaluation and punishment. Individuals were modeled as random effects to account for the repeated measures design. Table 4 presents parameter estimates, confidence intervals and *p*-values for the resulting multinomial linear regression model. Significant main effects of payoff for the dictator and false pretense emerged for evaluation and punishment. An interaction effect between the two factors was only significant in the evaluation condition.

Table 3. Recoded decision types.

Decision Type	Procedure	False Pretense	Expected Payoff in ECU	
			Self	Other
Egoist	direct	no	60	20
Altruist	direct	no	20	60
Fair	coin	no	40	40
Hypocrite/Self-Deceiver	direct	yes	60	20
Altruistic Deceiver	direct	yes	20	60

Table 4. Moral evaluation and punishment of decision types (hypocrisy condition).

Variable	Evaluation			Punishment		
	Beta	**95%-CI**	***p*-Value**	**Beta**	**95%-CI**	***p*-Value**
Dictator's Payoff	−0.010	[−0.014, −0.006]	0.000	0.022	[0.018, 0.026]	0.000
False Pretense	−0.314	[−0.598, −0.031]	0.030	0.279	[0.053, 0.504]	0.015
False Pretense x Payoff	−0.012	[−0.019, −0.006]	0.000	0.003	[−0.002, 0.008]	0.257

Note. 95%-CI = 95%-confidence interval. Individual participants were modeled as random effects.

The significance of differences between coefficients in the models was tested via integration of interaction terms (with moral evaluation vs. punishment as a binary dummy variable). The beta value of false pretense did not significantly differ between moral evaluation and punishment ($p > 0.1$). The dictator's payoff was significantly more important for punishment than for moral evaluation ($p < 0.001$). In turn, the interaction between payoff and pretense was significantly more important for evaluation than for punishment ($p = 0.048$).

Result 3. *The evaluation and punishment of hypocrisy were not fully proportionate.*

Although false moral pretense (hypocritical communication) has a very negative impact on moral evaluation, it does not fully translate into harsher punishment. Subjects' punishment is determined more by the violation of the fairness norm that applies to the distribution of money than by false pretense. The results of the regression analysis indicate that false pretense (or hypocritical communication) increases punishment in a simple additive way, and it amplifies the negative moral evaluation in a multiplicative way.

6. Conclusions

We are among the first to provide empirical data on moral reactions to hypocrisy. We find that observers clearly morally condemn hypocrisy. Even purely self-deceptive behavior—seen by many as a precondition for successful hypocrisy—is regarded much more critically than outright egoism. The deception of others adds to this perception of wrongness. Analogously, hypocrites are punished more severely than self-deceivers. In particular, subjects with a high regard for their moral identity are more concerned about the deception of others, relative to self-deception. Interestingly, however, the moral condemnation of hypocrites does not translate proportionally into meting out punishment for their behavior. Although false pretense plays a major role in determining moral evaluation, punishment is driven more significantly by the violation of the fairness norm regarding the unequal outcome.

Although reactions to hypocrisy can be frequently observed in the context of public discourse, it is reasonable to expect that moral outrage fails to translate proportionally into punishment. Lönnqvist et al. [35] have already provided evidence for the claim that people abstain from punishing hypocrites if there is reasonable doubt about the actual guilt of the hypocrite. Our results indicate that even if no such reasonable doubt exists, hypocrisy is morally condemned, but not punished in accordance with the fact of the condemnation, even if punishment is free of charge and no one else can be expected to do the punishing. As the psychological drivers of moral judgment of this transgression seem to differ from those that evoke the behavioral response, further research is needed to better understand this phenomenon.

This may also explain why IKEA's sales failed to plummet, despite the indignation provoked by the moral pretense of this supposedly green furnishing house. If there is a broad societal consensus on the moral reprehensibility of hypocrisy, behavioral reluctance to actually translate such blame into punishment may signal a need for an institutional compensation of this reluctance. If it is unlikely that the regulatory framework will impose sanctions as in the case of IKEA where the diffusion of responsibility along the supply chain may render judicial consequences unlikely, other institutional bodies could be in demand. The explicit inclusion of a measure of hypocrisy in the certification of a company

to qualify for sustainable investment funds could be such an institutional measure. This is also likely to have implications for non-listed companies such as IKEA, which has been considering an initial public offering for some time. The alternative is that hypocrites may learn that it is safe to regard deception as a dominant and viable strategy. When dealing with instances which are morally reprehensible but not justiciable, it is left to the outrage of stakeholders to translate this outrage into punishment such as boycotts. Yet, this translation does not always work: research shows that what we resent as citizens, when, e.g., demanding animal welfare programs, does not necessarily translate in our consumer behavior, e.g., buying certified animal products [36,37]. This citizen–consumer duality needs to be taken into consideration. With respect to people's ethical intuitions, it would be interesting to study whether their approving or disapproving of positive retributivism predict the proportionality of their evaluation, relative to the punishments they are willing to mete out.

This study is subject to various limitations. Our results about the proportionality of blame and punishment of hypocritical behavior can only be tentatively applied to real cases of corporate hypocrisy. It would therefore be desirable to replicate the findings of this study in other contexts and with different methodologies. Further research should also account for exogenous sanctioning institutions and investigate whether those institutions promote stakeholder punishment, for instance, because their existence makes the norm violation even more salient, or whether they crowd out stakeholder punishment. Investigating the interaction between exogenous institutions and stakeholder preferences is an important part of behavioral ethics as it helps to ensure the efficacy of these institutions.

Author Contributions: Conceptualization, M.v.G., J.J. and M.U.; methodology, M.v.G., J.J. and M.U.; software, M.v.G.; validation, J.J. and M.U.; formal analysis, M.v.G.; investigation, M.v.G., J.J.; data curation, M.v.G.; writing—original draft preparation, M.v.G., J.J., M.U.; writing—review and editing, M.v.G., J.J. and M.U.; visualization, M.v.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Ethical approval was obtained from the review board of experimenTUM, the social-science research lab of the Technical University of Munich.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Data can be accessed via https://github.com/grumiq/hypocrisy.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Isserow, J.; Klein, C. Hypocrisy and Moral Authority. *J. Ethics Soc. Philos.* **2017**, *12*, 191–222. [CrossRef]
2. Earthsight. FLATPÅCKED FÖRESTS: IKEA's Illegal Timber Problem and the Flawed Green Label behind It. Available online: https://www.earthsight.org.uk/flatpackedforests-en (accessed on 12 March 2021).
3. Ingka Group. Solid Sales Performance When Life at Home Has Never Been More Important. Available online: https://www.ingka.com/news/solid-sales-performance-when-life-at-home-has-never-been-more-important/ (accessed on 12 March 2020).
4. Klepper, S.; Nagin, D. The deterrence effect of perceived certainty and severity of punishment revisited. *Criminology* **1989**, *27*, 721–746. [CrossRef]
5. Green, S.P. The concept of white-collar crime in law and legal theory. *Buffalo Crim. Law Rev.* **2004**, *8*, 1–34. [CrossRef]
6. Ryberg, J. *The Ethics of Proportionate Punishment: A Critical Investigation*; Kluwer Academic Publishers: Norwell, MA, USA, 2004.
7. Ward, T.; Salmon, K. The ethics of punishment: Correctional practice implications. *Aggress. Violent Behav.* **2009**, *14*, 239–247. [CrossRef]
8. Mackie, J.L. Morality and the retributive emotions. *Crim. Justice Ethics* **1982**, *1*, 3–10. [CrossRef]
9. Abbink, K.; Irlenbusch, B.; Renner, E. The moonlighting game: An experimental study on reciprocity and retribution. *J. Econ. Behav. Organ.* **2000**, *42*, 265–277. [CrossRef]
10. Offermann, T. Hurting hurts more than helping helps. *Eur. Econ. Rev.* **2002**, *46*, 1423–1437. [CrossRef]
11. Baumeister, R.F.; Bratslavsky, E.; Finkenauer, C.; Vohs, K.D. Bad is Stronger than Good. *Rev. Gen. Psychol.* **2001**, *5*, 323–370. [CrossRef]

12. Wang, C.S.; Galinsky, A.D.; Murnighan, J.K. Bad drives psychological reactions, but good propels behavior responses to honesty and deception. *Psychol. Sci.* **2009**, *20*, 634–644. [CrossRef]
13. Hindriks, F. Normativity in Action: How to Explain the Knobe Effect and its Relatives. *Mind Lang.* **2014**, *29*, 51–72. [CrossRef]
14. Batson, C.D.; Kobrynowicz, D.; Dinnerstein, J.L.; Kampf, H.C.; Wilson, A.D. In a very different voice: Unmasking moral hypocrisy. *J. Personal. Soc. Psychol* **1997**, *72*, 1335–1348. [CrossRef]
15. Gneezy, U.; Saccardo, S.; Serra-Garcia, M.; van Veldhuizen, R. Bribing the self. *Games Econ. Behav.* **2020**, *120*, 311–324. [CrossRef]
16. Dana, J.; Weber, R.A.; Kuang, J.X. Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Econ. Theory* **2007**, *33*, 67–80. [CrossRef]
17. Trivers, R. *Social Evolution*; Benjamin/Cummings: Menlo Park, CA, USA, 1985.
18. Lönnqvist, J.-E.; Irlenbusch, B.; Walkowitz, G. Moral hypocrisy: Impression management or self-deception? *J. Exp. Soc. Psychol.* **2014**, *55*, 53–62. [CrossRef]
19. Butler, J. Upon Self-Deceit. In *The Works of Bishop Butler*; White, D., Ed.; Rochester: New York, NY, USA, 1726.
20. Van Leeuwen, N.; LaFollette, H. Self-Deception. In *The International Encyclopedia of Ethics*; Blackwell: Oxford, UK, 2013.
21. Darwall, S. Self-Deception, Autonomy, and Moral Constitution. In *Perspectives on Self-Deception*; McLaughlin, B.P., Oksenberg Rorty, A., Eds.; University of California Press: Berkeley, CA, USA, 1988; pp. 407–430.
22. Levy, N. Self-Deception and Moral Responsibility. *Ratio* **2004**, *17*, 294–311. [CrossRef]
23. Sie, M. Moral Hypocrisy and Acting for Reasons: How Moralizing Can Invite Self-Deception. *Ethical Theory Moral Pract.* **2015**, *18*, 223–235. [CrossRef]
24. Aquino, K.; Reed, A. The self-importance of moral identity. *J. Pers. Soc. Psychol.* **2002**, *83*, 1423–1440. [CrossRef] [PubMed]
25. Stets, J.E.; Carter, M.J. A Theory of the Self for the Sociology of Morality. *Am. Sociol. Rev.* **2012**, *77*, 120–140. [CrossRef]
26. Reed, A.; Aquino, K.; Levy, E. Moral Identity and Judgments of Charitable Behaviors. *J. Mark.* **2007**, *71*, 178–193. [CrossRef]
27. Leiner, D.J. SoSci Survey, Version 2.5. 00-i1142. 2018. Available online: https://www.soscisurvey.de/ (accessed on 20 April 2020).
28. Abeler, J.; Falk, A.; Goette, L.; Huffman, D. Reference Points and Effort Provision. *Am. Econ. Rev.* **2011**, *101*, 470–492. [CrossRef]
29. Lakens, D.; Caldwell, A.R. Simulation-Based Power-Analysis for Factorial ANOVA Designs. Available online: https://doi.org/10.31234/osf.io/baxsf (accessed on 28 May 2019).
30. Champley, S. Pwr: Basic Functions for Power Analysis. R Package Version 1.3-0. Available online: https://CRAN.R-project.org/package=pwr (accessed on 15 December 2020).
31. Abbink, K.; Sadrieh, A. The pleasure of being nasty. *Econ. Lett.* **2009**, *105*, 306–308. [CrossRef]
32. Jauernig, J.; Uhl, M.; Luetge, C. Competition-induced punishment of winners and losers: Who is the target? *J. Econ. Psychol.* **2016**, *57*, 13–25. [CrossRef]
33. Jauernig, J.; Uhl, M.; Luetge, C. Voluntary agreements between competitors: Trick or truth? *J. Bus. Econ.* **2017**, *87*, 1173–1191. [CrossRef]
34. Jauernig, J.; Uhl, M. Spite and preemptive retaliation after tournaments. *J. Econ. Behav. Organ.* **2019**, *158*, 328–336. [CrossRef]
35. Lönnqvist, J.-E.; Rilke, R.M.; Walkowitz, G. On why hypocrisy thrives: Reasonable doubt created by moral posturing can deter punishment. *J. Exp. Soc. Psychol.* **2015**, *59*, 139–145. [CrossRef]
36. Verbeke, W. Stakeholder, citizen and consumer interests in farm animal welfare. *Anim. Welf.* **2009**, *18*, 325–333.
37. Bennett, R.M.; Anderson, J.; Blaney, R.J.P. Moral Intensity and Willingness to Pay Concerning Farm Animal Welfare Issues and the Implications for Agricultural Policy. *J. Agric. Environ. Ethics* **2002**, *15*, 187–202. [CrossRef]