DISCUSSION PAPER SERIES

DP15645

SELF-SIGNALING IN MORAL VOTING

Lydia Mechtenberg, Grischa Perino, Nicolas Treich,
Jean-Robert Tyran and Stephanie Wang

PUBLIC ECONOMICS

ISSN 0265-8003

SELF-SIGNALING IN MORAL VOTING

Lydia Mechtenberg, Grischa Perino, Nicolas Treich, Jean-Robert Tyran and Stephanie Wang

Discussion Paper DP15645
Published 06 January 2021
Submitted 01 January 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

SELF-SIGNALING IN MORAL VOTING

Abstract

This paper presents a two-wave survey experiment on self-image concerns in moral voting. We elicit votes on the so-called Horncow Initiative. This initiative required subsidization of farmers who refrain from dehorning. We investigate how non-consequentialist and non-deontological messages changing the moral self-signaling value of a Yes vote affect selection and processing of consequentialist information, and reported voting behavior. We find that a message enhancing the self-signaling value of a Yes vote is effective: voters agree more with arguments in favor of the initiative, anticipate more frequently voting in favor, and report more frequently having voted in favor of the initiative.

JEL Classification: C93, D72, D91

Keywords: moral bias, voting, multi-wave field experiment, information avoidance

Lydia Mechtenberg - lydia.mechtenberg@uni-hamburg.de
University of Hamburg

Grischa Perino - grischa.perino@uni-hamburg.de
University of Hamburg

Nicolas Treich - nicolas.treich@inrae.fr
Toulouse School of Economics

Jean-Robert Tyran - jean-robert.tyran@univie.ac.at
University of Vienna and CEPR

Stephanie Wang - swwang@pitt.edu
University of Pittsburgh

Acknowledgements
We thank Claudia Schwirplies for helpful comments. Nicolas Treich acknowledges support from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program), and IDEX-AMEP and FDIR chairs at the Toulouse School of Economics (TSE-P). Lydia Mechtenberg acknowledges support from the EU-Consortium DEMOS under grant agreement ID 822590 (Horizon 2020).

Self-Signaling in Moral Voting

Lydia Mechtenberg, Grischa Perino, Nicolas Treich,
Jean-Robert Tyran, and Stephanie Wang[1]

January 4, 2021

Abstract

This paper presents a two-wave survey experiment on self-image concerns in moral voting. We elicit votes on the so-called Horncow Initiative. This initiative required subsidization of farmers who refrain from dehorning. We investigate how non-consequentialist and non-deontological messages changing the moral self-signaling value of a *Yes* vote affect selection and processing of consequentialist information, and reported voting behavior. We find that a message enhancing the self-signaling value of a *Yes* vote is effective: voters agree more with arguments in favor of the initiative, anticipate more frequently voting in favor, and report more frequently having voted in favor of the initiative.

JEL: C93, D72, D91

Keywords: moral bias, voting, multi-wave field experiment, information avoidance

[1] Mechtenberg (corresponding author): University of Hamburg, Lydia.Mechtenberg@uni-hamburg.de. Perino: University of Hamburg, Grischa.Perino@uni-hamburg.de. Treich: University Toulouse Capitole, INRAE, Toulouse School of Economics, nicolas.treich@inrae.fr, Tyran: University of Vienna, University of Copenhagen and CEPR (London), Jean-Robert.Tyran@univie.ac.at, Wang: University of Pittsburgh, swwang@pitt.edu. We thank Claudia Schwirplies for helpful comments. Nicolas Treich acknowledges support from ANR under grant ANR-17-EURE-0010 (Investissements d'Avenir program), and IDEX-AMEP and FDIR chairs at the Toulouse School of Economics (TSE-P). Lydia Mechtenberg acknowledges support from the EU-Consortium DEMOS under grant agreement ID 822590 (Horizon 2020).

1. Introduction

This paper tests two different pathways through which moralizing political campaigns can use self-image concerns of voters to bias votes. Both in Europe and the U.S., moral arguments are becoming increasingly prevalent in the political discourse (Sandel 2006, Haidt 2012, Enke 2020). This development is challenging for economists. A substantial part of the moral claims used in political campaigns clash with the consequentialist reasoning that economists prefer, even without being deontological (i.e., duty- or rights-based) in nature. Moral claims in political discourse take the form of praise or condemnation, suggesting, for instance, that voters' choices between conservative and progressive politics reveal whether they are good Christians. Moralizing campaigners do not always inform the public about the true moral consequences of their requests; and not all of these non-consequentialist moralizing campaigns argue in favor of generalizable duties or rights either. Instead, some of them simply appeal to the voters' desire to feel good about themselves (Bénabou and Tirole 2011). If voters are liable to such appeals, then voting outcomes may frequently contradict consequentialist and even deontological thinking.[2] Indeed, the moral benefits of some such outcomes may be outweighed by their moral costs. In a large (N >1000) two-wave survey experiment conducted around an animal-welfare ballot in Switzerland, we test if and how one can use moralizing campaigns that circumvent both consequentialist and deontological arguments to bias voters' information collection, information processing, and voting behavior.

It is well-understood that most people, at least to some extent, like acting moral. In the voting context, this is of particular importance since with low pivotality, voters tend to express their moral beliefs more than their material interests (Feddersen et al. 2009). However, acting moral relates to the true moral values of the acts in a rather opaque way. While utilitarianism requires that good deeds have good consequences, at least to the best of the agent's knowledge, people confronted with moral choices often deliberately neglect harmful consequences of their acts. At least in the laboratory, they tend to avoid information on whether their choices harm others, as documented by Dana et al. (2007) and Ehrich and Irwin (2005). Such information avoidance is hard to rationalize by assigning a specific type of morality to the subjects. For instance, though a deontological morality could explain following a moral rule even in a situation where its application generates harmful consequences, it is not clear why a deontologist would avoid knowing these consequences.

In an attempt to explain such willful ignorance, Bénabou and Tirole (2011) provide a theory of moral acts as identity investments: subjects infer what their character is worth from their past acts. Hence, they choose their acts with an eye on their future self-image and *want* to believe that

2 See also Bénabou et al. (2020). The concept of consequentialism is used here in its broadest sense, i.e., encompassing any approach that derives the moral value of an act from the moral value of its consequences. Hence, consequentialism in its broadest sense does not necessarily require that the values of the moral consequences can be quantified and compared across individuals.

these acts are morally good, or at least not morally bad.[3] Thus, apart from true (consequentialist) other-regarding preferences or deontological principles, self-image concerns – i.e., people's desires to feel good about themselves – are a potential explanation of seemingly moral choices. These concerns motivate a self-signaling game in which beliefs about one's own character are purposefully manipulated. For instance, people may support a campaign for a quick nuclear phase-out or against the use of animals in medical science just in order to make themselves believe that they have a morally good character. This phenomenon has important implications for information selection and processing. An act chosen without knowledge of its harmful consequences on others may keep one's self-image intact. However, that same act committed in full knowledge of its harmful social effects would impair the self-image of anyone who is not a radical deontologist actually following a general moral principle. Hence, a self-signaling motivation for moral choices directly leads into the "moral wiggle room" (Dana et al. 2007): subjects in the laboratory avoid costless information on whether a preferred act has pro- or anti-social consequences; and they do so in order to be able to both commit this act and sustain their moral self-image. Hence, self-image concerns can explain the purposeful use of information-avoidance strategies.

It is largely an open question whether people are both able and willing to uphold such strategies outside the laboratory. Information is harder to avoid in everyday life than in the laboratory. Voters encounter political ads in movies and on their Facebook account and receive political messages from their family, friends, and neighbors, many of them about consequences of the policies in question. However, there is another possible informational strategy that people could use to manipulate their beliefs about the moral value of their acts, in particular their votes: They could downplay information that contradicts the supposed morality of their choices, and they could overweigh information supporting this supposed morality.[4] Such information-processing strategies of self-manipulation are as consistent with the theory of Bénabou and Tirole (2011) as pure information avoidance is. Similar to confirmation bias (Rabin and Schrag 1999), biased moral beliefs resulting from biased information-processing strategies could survive all contrary campaigns that impose consequentialist fact-based information on voters. If people do indeed use biased information-processing to improve their moral self-image, campaigners will be heavily tempted to play the moral card to trigger these strategies in voters and win them over. Their opponents will be at a loss of what to do against this – other than engaging in the same strategy and thereby escalating political polarization (Garrett and Bankert 2020).

In this paper, we investigate how self-image concerns affect information selection and processing – and, through these, intended and actual votes. We do so in the context of a popular vote on an animal-welfare policy in Switzerland. The so-called Horncow Initiative demanded writing the

3 See also Bénabou and Tirole (2006).

4 For a recent theory that models such strategies as the use of narratives, see Bénabou et al. (2020).

dignity of horned animals into the constitution and to cross-subsidize farmers with horned cattle that refrain from dehorning. Both the campaigners for the initiative and their opponents used consequentialist moral arguments. This ballot provides the ideal setting for our experimental research since it embodies a strong moral dimension and has a negligible material impact on the voters. We conduct a pre-registered and IRB-approved randomized online experiment prior to the vote and elicit voting behavior afterwards.[5] Our main treatments manipulate the moral self-signaling value of voting for the policy required by the initiative. These manipulations are implemented without referring to moral principles or attempting to affect beliefs on the policy's consequences for the animals concerned, i.e., without manipulating consequentialist or deontological beliefs about the true moral value of a *Yes* or *No* vote. Within an informational intervention designed to kindle self-image concerns, we provide subjects with the truthful, if simplified, message that good-hearted people tend to be good to animals, too. Thereby, we weakly increase the moral self-signaling value of voting for the policy. We enable subjects by this treatment to use a *Yes* vote as a means to self-signal both being good to animals and being good *in general*, which supposedly feels better than the more specific belief of being good to animals. Importantly, our intervention does not provide additional information on the policy's effectiveness or conformity with moral duties or rights and is hence irrelevant for truly moral attitudes. In an opposite informational intervention that is designed to restrict the moral self-signaling value of a *Yes* vote, we tell our subjects that being good to animals does not necessarily imply being good to humans. Both interventions are compared with the benchmark treatment in which no extra message is provided. We then study how our moral information interventions affect the willingness to read arguments pro and contra the policy, trust in these arguments, voting intentions, and reported voting behavior.

Furthermore, we add a social dimension. Even pure self-signaling must be sustainable in the individual's social setting. Communication with peers can either counteract the strategies used to protect one's self-image or, on the contrary, complement them. In particular, communication with like-minded people has been documented to have an insulating effect, preventing opinion change (Hüning, Mechtenberg, and Wang 2020). This suggests that informational strategies sustaining moral ignorance – and hence a positive self-image – may be more successful when flanked by the opportunity to communicate with a like-minded person and less successful with an opposite-minded person. We hence provide such opportunities to our participants and test their effects on information selection.

At the end of our second wave after the vote, we match our subjects into pairs of either like-minded or opposite-minded voters who may chat about how they voted and why. In our first

5 Ethical approval has been granted by the dean of the social-science faculty at Hamburg University. The form can be obtained from the authors by request. The experiment was pre-registered at the AEA RCT Registry (https://doi.org/10.1257/rct.3551-1.0) under a different title.

experimental wave, prior to the vote, we manipulate whether subjects anticipate being thus matched. Hence, we manipulate the anticipated degree to which our experiment provides social insulation to the participant's prior attitude toward the Horncow Initiative. We then test whether the anticipated degree of social insulation affects information selection.

We find that increasing the moral self-signaling value of voting *Yes* through our informational intervention has indeed significant effects: First, it enhances the intensity of subjects' agreement with arguments stating that the policy at stake would indeed benefit the animals concerned, while the intensity of their agreement with arguments disputing this remains unchanged. Surprisingly, however, our intervention does not affect the choice of which type of arguments to read. Hence, we find that in the field, acting moral is sustained by biased information processing. This evidence specifies one important new channel through which "motivated bias" as in the modelling paradigm of Bénabou and Tirole (2002, 2006, 2011) is generated and sustained in real-world situations. Second, our intervention increases the number of subjects intending to vote *Yes* and, third, even the number of subjects who report having actually voted *Yes* after the ballot. Decreasing the moral self-signaling value of voting *Yes* has no effect. We also do not find that anticipating communication with another voter, like-minded or not, has any effect on reading or trusting arguments, or on intended or reported voting behavior. Hence, we do not find evidence for the importance of social insulation in moral voting contexts. Our findings support and specify the theory of Bénabou and Tirole (2011): people, through their choice of actions, signal to themselves a moral value of their character. In our experiment, this works through biased information processing rather than biased information selection. In sum, the motivation of gaining a positive moral self-image indeed seems to be an important determinant of acting moral through a vote.

Our findings are relevant for a variety of issues. If, as we argue, people want to believe being moral rather than be moral, playing the morality card during political, social, or economic campaigns can both be helpful and dangerous. It can be helpful since it can mobilize voters, citizens, and consumers to support a good cause, like moral standards in production. It can be dangerous since it biases at least some of them, motivating them to believe more strongly in the beneficial consequences of a suggested measure, ignoring counterarguments. For instance, it is tempting to play the moral card in support of rent control ("be moral, help the poor"), subsidies for windmills or bans on plastic bags ("be moral, save the climate"), or unilateral disarmament ("be moral, keep peace"), without going into depths about the complex consequences of these requests. Arguments to the effect that the suggested policies might not have the promised beneficial consequences will be less convincing when directed toward voters or consumers targeted by a moralizing campaign.

Contribution to the literature

This paper contributes to three different and hitherto unconnected strands of literature. The first strand is the theoretical and experimental literature on willful or moral ignorance. As modelled by Grossman and van der Weele (2017) building upon Bodner and Prelec (2003) and Bénabou and Tirole (2006, 2011), subjects signaling morality toward their future selves have an incentive to avoid learning whether their actions harm others.[6] They are driven into ignorance by their fear of losing their positive self-image. Out of this fear, they can either avoid information that would undermine their self-signaling strategies, or they can bias their information processing. Validating the former prediction, subjects in the economic laboratory tend to exhibit a positive willingness to pay for remaining ignorant about harmful consequences, as documented by Dana, Weber, and Kuang (2007), Ehrich and Irwin (2005), Grossman and van der Weele (2017), and Serra-Garcia and Szech (2019).[7]

We complement this literature in three different ways. First, to the best of our knowledge we are the first to investigate self-signaling strategies in the voting context.[8] In this context, self-signaling is particularly cheap since pivotality and the individual cost of voting are both low. Second, and relatedly, we choose a new angle: we shift the focus away from ignorance about the darker spots on one's character which dominates the existing literature. Instead, we focus on ignorance about the potential non-existence of bright spots. That is, our focus is not on ignorance about the harmful effects of one's *egoistic* choices but on ignorance about the doubtfulness of the supposed beneficial effects of one's – seemingly – *moral* choices, as in Niehaus (2020). The distinction between the two is important for the following reason: it is intuitive that egoistic people, with high costs of being moral but a desire to think well of themselves, shy away from information on potentially harmful consequences of their choices. By contrast, it is much harder, and more depressing, to believe that a substantial number of so-called altruists are in fact only engaging in self-signaling – with doubtful consequences on those targeted by their apparent morality. If this were true, it would cast a cloud over one of the main pillars of behavioral economics, social preferences. Since we are investigating a moral choice of low individual cost – voting on a morally relevant ballot – we only show that some people who act moral engage in cheap self-signaling, which is particularly easy to do in the voting context. Hence, we do not argue that moral choices that are associated with more substantial costs, such as, for instance, organ donations before death, can also be traced back to moral self-signaling. Nonetheless, the consequences of moral self-signaling can be substantial in the voting context, even if the signaling itself is cheap: If a seemingly moral cause that has in fact no beneficial consequences

6 For related theory papers, see Nyborg (2011) and Hestermann et al. (2020). The latter is explicitly concerned with animal welfare.

7 Andreoni (2017) documents information avoidance in the field.

8 For a survey experiment that investigates social image concerns, see DellaVigna et al. (2017).

can generate a crowd of supporters despite the costless availability of arguments against its moral value, then substantial rents can be re-distributed or even destroyed under the cloak of morality.[9]

The third way in which we complement the literature on moral ignorance is that we choose a new method: we investigate moral self-signaling in an online survey experiment, not in the laboratory.[10] While we hence lose some control, in particular in having to rely on self-reported rather than directly observed behavior, we arguably gain in external validity. In particular, our sample is heterogeneous with respect to age, education, income, and relatedness to the issue of the ballot – more so than a student sample could be. In addition, the measured behavior itself-how our subjects vote - takes place outside the experiment.

Moreover, we contribute to the literature on expressive voting. This literature is based on the idea that voters have two types of preferences, one about the material outcomes of the votes and one about expressing their opinions or emotions (see, among others, Brennan and Lomasky, 1993; Brennan and Hamlin, 1998; Tyran 2004; Feddersen et al. 2009; and Shayo and Harel 2012). The concept of expressive voting overlaps with the idea of moral voting bias that supposedly becomes larger when the pivot probability declines – a phenomenon that has been shown to occur in the laboratory (Feddersen et al. 2009) and that comes as no surprise if one believes in moral-self signaling. However, though intuitive, to the best of our knowledge the connection between moral self-signaling and moral bias in voting has never been made explicitly, or even supported by evidence, in the literature. We intend our paper to be the first step toward revealing the interconnectedness of these two important concepts.

The final strand of the literature related to our contribution are the numerous field intervention studies in political science and, recently, political economics, that use manipulations of real-world electoral campaigns to investigate voting behavior (see, among others, Zaller, 1992; Gerber and Green, 2000; Gerber et al., 2001; and Kendall et al. 2015).[11] Methodologically speaking, our study takes an intermediate place between this literature and the literature that reports laboratory experiments on determinants of moral choices. Our study is innovative with regard to both in that we investigate votes as moral self-signaling devices, and in that we study moral self-signaling in a real-world voting context.

9 This said, we reserve judgment on the true moral value of the Horncow Initiative. It is our informational interventions, not the Horncow Initiative, that we declare as void of morally relevant consequentialist content.

10 For other studies outside the laboratory, see Andreoni (2017) for a field experiment in the context of charitable giving and Freddi (2017) for field evidence from a natural experiment.

11 See Gerber and Green (2006) for an overview.

2. Procedures and predictions

2.1 Experimental design

On November 25, 2018, the Swiss voted on the proposal of a grass-root initiative colloquially called "Horncow Initiative". This initiative demanded to pin down the dignity of horned animals in the Swiss constitution. In addition, they asked for subsidizing farmers who do not cauterize their animals' horns. These subsidies, they requested, should be financed by cutting agricultural subsidies elsewhere and should hence be without effect on taxes or prices. We chose this ballot for its near-absence of substantial economic impact: The cost and consumption effects that the initiative's proposal would have on most voters in case of success would be negligible, and their self-interest would not be touched.[12] Hence, voters' instrumental concerns would be mainly altruistic, i.e., directed toward the proposal's true consequences on animal welfare. This provides an incentive to gain as objective, unbiased information about these consequences as possible. Moral self-signaling concerns, by contrast, make it attractive to remain ignorant about the proposal's potential negative consequences, or the potential absence of positive consequences. Hence, instrumental moral concerns and moral self-signaling concerns would be conflicting, which provides the ideal setting for a study of votes as potential devices of moral self-signaling, allowing us to separate these from consequentialist concerns. Deontological concerns would not necessarily conflict with self-signaling concerns but would be orthogonal to a change in the self-signaling value of a *Yes* vote. Hence, our treatments also allow us separating self-signaling from deontological voting. What we cannot do is separating deontological and consequentialist voting motives.

We conducted a pre-registered and IRB-approved two-waves survey experiment timed before and after the ballot.[13] In the second wave, we re-contacted only subjects that completed the first wave. The first wave was implemented in the two weeks prior to the final day of the ballot, and the second wave a few days after. We restricted our experiment to the German-speaking part of Switzerland. The Swiss standing LINK Institute panel was employed for recruitment, and written consent was obtained from all subjects as part of wave 1. Only truthful information was given to them. Subjects were informed as part of their consent that the survey in wave 1 might vary across participants. We screened out early voters who had voted already before the start date of wave 1 and participants not eligible to vote.

In the first wave, we conducted nine randomized versions of one survey. All versions elicited, among relevant demographics, (1) variables measuring *information selection*, (2) a variable measuring *information processing*, and (3) the *intended vote*. In addition, we elicited control variables such as the *PriorAttitude* toward the initiative's proposal and prior informedness

12 There are of course non-negligible cost effects on farmers. We elicit if our subjects are farmers or related to farmers and control for this.

13 The experiment was pre-registered at the AEA RCT Registry (https://doi.org/10.1257/rct.3551-1.0).

(*Informed*). A full list of variables and their explanations is relegated to Table A.1 in the appendix.

Information selection. We measure information selection as follows. A booklet that the Swiss government sent to all Swiss voters several weeks before our experiment started contained three arguments in favor and three arguments against the initiative's proposal. We used these and one other argument widely circulating in the media to create a balanced information menu: Three arguments in favor of the proposal claimed that dignity and physical well-being of animals and justice among farmers would improve, should the initiative be approved in the ballot. Three arguments against the proposal addressed these same three goals and argued that none of them would be reached in case of the proposal's success. (See Table 1 for the precise formulation of the six arguments.)[14] Our subjects had to choose which arguments to read: all six, only the three in favor, only the three against, or none at all. At the point of choice, they did not know that we were offering them only arguments they already were highly likely to know from the official booklet or the media. Thereby, without biasing their information set, we could measure the willingness of those predisposed in favor of the initiative to avoid negative information on the proposal's potential consequences, either by avoiding information in general or by reading only the supportive arguments. Similarly, we could measure the willingness of all voters to avoid reading arguments that contradict their prior attitude toward the initiative.

Information processing. Even if acquired information was unbiased, it might be processed in a biased way by our subjects, as modeled in most of the work of Bénabou and Tirole (e.g., 2002, 2006, 2011). In our survey, we therefore asked for each type of argument, supportive or unsupportive of the proposal, how much the subjects who read it agreed with it ranging from 'not at all' to 'fully'. This allows us to compute a measure of change relating the *PriorAttitude*, i.e. the degree to which participants reported to be leaning in favor or against the initiative before being exposed to treatment and information, to the degree to which they agree with pro or contra arguments post treatment. To this end we normalize all variables and take the difference between ex-post and ex-ante variables normalizing the result to the interval [-1,1]. The resulting change measures are *ΔAgreementPRO* and *ΔAgreementCON*. Moral self-signalers have to believe in the moral value of a *Yes* vote. Hence, the more they want to self-signal, the more they have an incentive to process arguments in a biased way, assigning more weight to the supportive type.

Intended and actual votes. Voting plans tend to function as commitment devices (Nickerson and Rogers 2010). We hence elicit the immediate effect of our treatment variations explained below on planned voting behavior by asking our subjects whether they intend to turn out and, if they do, how they intend to vote. In combination with the variable *PriorAttitude*, this allows us to measure changes in how subjects evaluate the proposal after being treated in the experiment.

14 Table 1 shows the English translation of the original German arguments.

After the ballot, we re-contacted all subjects who completed the first wave and elicited their actual vote by asking whether, and how, they voted.

Table 1: Arguments of the endogenous information-acquisition mechanism.

Arguments for the Horncow Initiative	**Arguments Against the Horncow Initiative**
Dehorning violates the dignity of animals and is tantamount to a mutilation. It must mean something if nature gave horns to cows. For instance, these horns help the cows sorting out their hierarchy within their herd.	*It is well possible that the Horncow Initiative does not improve the dignity of animals. The reason is that in order to get subsidized, farmers could resolve to fixate their animals (e.g., by tethering). Their motive: Wounds caused by horns lower profits but may be prevented not only by dehorning but also by resolute fixation of the cattle, i.e., by limiting their range of motion to the greatest extent. Hence, farmers who nowadays dehorn their animals could, in case of the initiative's success, switch to permanent tethering of their cattle.*
Horns are organs well supplied with blood. Dehorning cows requires cauterizing the sockets of the horns to prevent them growing. This is a substantial medical intervention. Even though this intervention is legally required to be conducted under anaesthetization, many calves suffer from pain after cauterization, partly for long time.	**It is well possible that the Horncow Initiative does not prevent cruelty to animals. Resolute limiting of their range of motion in the stable or wounds caused by horns of other cows could result from subsidizing farmers with horned cattle. Possibly cows suffer more from tethering (or, alternatively, wounds caused by skirmishes with other horned cows in the stable) than from the dehorning.*
Since horned animals need more space and care from their farmers, a compensation for farmers holding horned animals is justified. Hence, farmers holding horned animals should be subsidized. Since the initiative does not demand a legally banning dehorning animals, the farmers' freedom of choice is preserved.	*Subsidizing farmers with horned animals may put those farmers at a disadvantage who breed hornless cattle. Even nowadays there are such farmers in Switzerland. There is no scientific evidence that cattle that is born hornless is "less natural" or suffers more than horned cattle. Hence, one should not put farmers who breed hornless cattle at a disadvantage.*

Note: *This argument has been taken from the media. The Swiss booklet sent to all Swiss voters mentioned an argument almost identical to the second in the right column here.

Treatments. Our treatments are depicted in Figure 1. First, we vary the self-signaling value of a *Yes* vote: HIGH and LOW treatments differ from NEUTRAL treatments in that in both former types, we give subjects true information that we expect will enhance (in HIGH) or lower (in LOW) the salient moral value of a *Yes* vote. In HIGH, we cite evidence for the positive correlation between cruelty towards animals and cruelty towards humans and conclude that good-hearted people tend to be good to animals. In LOW, we cite evidence indicating that the correlation between empathy with animals in need and empathy with humans in need is less than perfect. Note that neither information touches the question whether the success of the initiative would improve the animals' situation. (See Appendix A.I for the precise wording and the

scientific foundation of our interventions.) These informational interventions are implemented after we elicit prior attitudes and prior informedness but before subjects have to make their choice of arguments to read. In the NEUTRAL treatments, we refrain from any such informational intervention.

Second, and orthogonal to this variation, the treatments BUBBLE, CONFRONT and NOA vary what and how much subjects were told about communication in the second wave: in BUBBLE (CONFRONT), we told them that they would chat with someone of similar (different) prior attitude toward the Horncow Initiative after the second survey. In NOA (for "no anticipation"), we only told them that they would be re-invited for a second survey. Hence, BUBBLE and CONFRONT but not NOA induce anticipation of a social situation in which subjects may discuss their votes. While BUBBLE lets subjects expect social insulation of their prior opinion, CONFRONT promises confrontation with a different opinion.

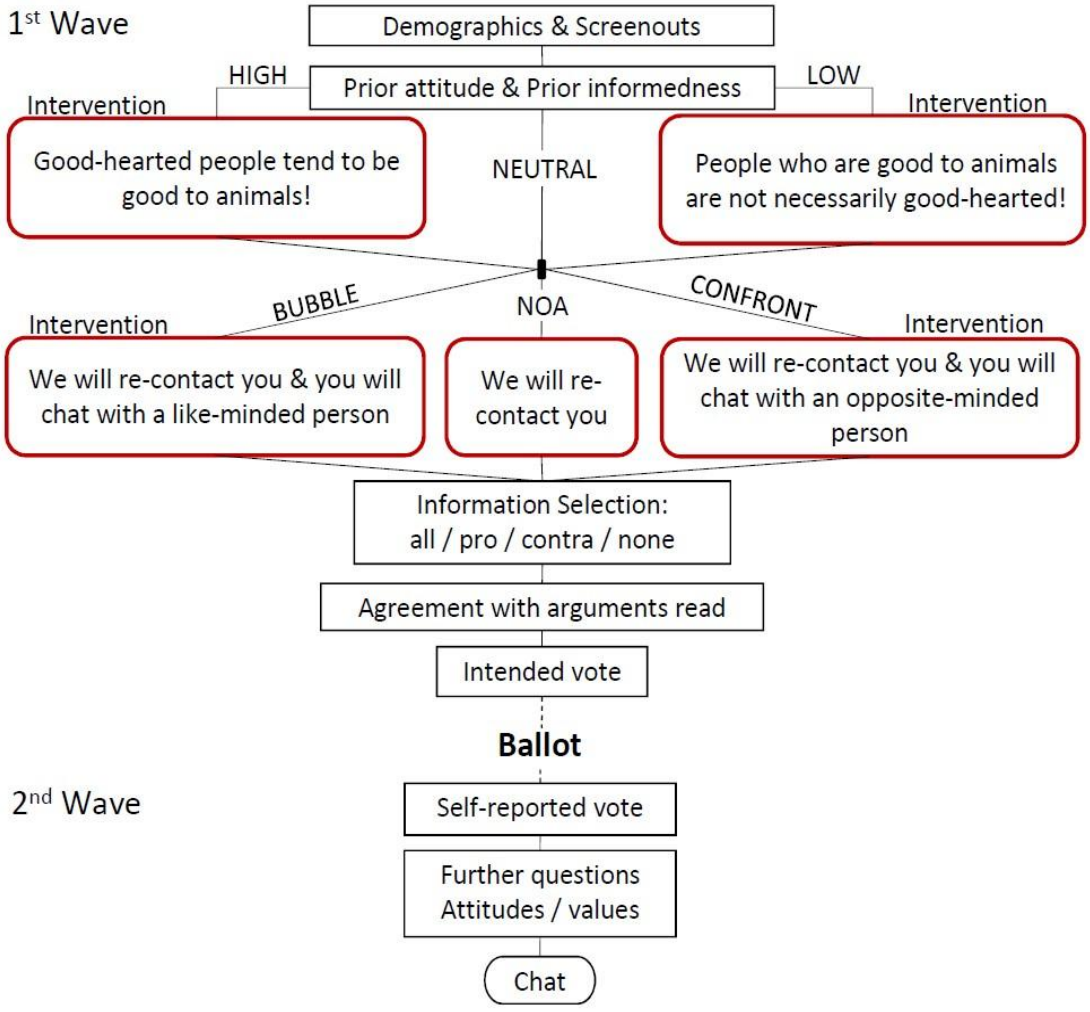
The second wave of our experiment included a short survey identical for all subjects and a partner-chat in which, depending on whether the subject was in BUBBLE or CONFRONT, they chatted with an (ex ante) like-minded or opposite-minded partner. (Participants in NOA were randomly assigned to either a like-minded or an opposite-minded partner.) We re-invited all 2,112 subjects who completed the first survey and had 1,057 completing the second. Apart from eliciting self-reported actual votes, the second survey contained questions on whether consequences (*GoodEffect*) or intentions (*GoodIntent*) are more important when morally evaluating – and rewarding – a particular action.

Two pathways. Before we state our predictions, we clarify one important distinction that separates two mutually exclusive pathways toward voting in order to self-signal morality in our experiment: the distinction between information selection and biased information processing. While we can measure information avoidance by whether subjects neglected arguments, we can measure biased information processing by the agreement with arguments in favor or against the initiative only *for those who read both* types of arguments. Hence, if we find sizeable information avoidance, we will not be able by design to find biased information processing, because biased information selection implies that the sample of those reading both types of arguments is biased. Treatment effects on agreement with arguments in favor or against the initiative could not be reliably attributed to biased information processing in this case.

Apart from this technicality, there is another design feature that marks the two types of informational bias as openings into two separate pathways of behavior. To see this, consider first a subject supportive of the initiative who ponders which arguments to read: all, or only those in favor, or only those against, or none. Our treatment interventions are designed to influence her choice. However, since we composed the set of arguments based on the booklet sent to all Swiss voters by the government, with the exception of one argument that was prominent in the media,

the subject's information set should not depend on which reading choice she would finally make. Therefore, by design subjects who exhibit biased information selection in our experiment should not be driven by that to morally biased voting behavior, be it planned or actual. Hence, biased information selection in our experiment should not lead to biased voting, compared to how our subjects would have voted without our treatment interventions.

Figure 1: Treatments



However, it is easy to see that the situation is different when considering biased information processing. Here, we were unable to preclude by design that the bias on the informational stage translates into biased voting. A subject reading both types of arguments but influenced by our treatment intervention HIGH (LOW) to put more (less) weight on those in favor of the initiative may well become more (less) likely to vote *Yes*. Hence, we get the two potential pathways *A (*for *Avoid)* and *B (*for *Bias)* below.

Pathway A: In treatment HIGH (LOW), subjects become weakly more (less) likely to skip arguments against the initiative, compared to NEUTRAL, respectively. But they remain unaffected in their planned and actual voting behavior.

Pathway B: In treatment HIGH (LOW), subjects become weakly more (less) likely to overweigh arguments in favor of the initiative, relative to those against, than in NEUTRAL, respectively. Hence, relative to NEUTRAL, subjects in the former treatments become weakly more (less) likely to plan to vote *Yes* and to actually vote *Yes.*

Both pathways, *A* and *B*, are rooted in the theoretical literature discussed above, in particular in the work of Bénabou and Tirole. While the experimental literature on information avoidance has already documented subjects using pathway *A* in the laboratory, pathway *B* still lacks empirical evidence.[15]

Predictions

Below, we state the predictions for both pathways. If pathway A is clearly refuted, we will get a large enough subsample of subjects who read all arguments, which allows us to test pathway B if that sample turns out to be unbiased.

Pathway A. We now state all hypotheses relating to pathway A.

Hypothesis H1.A (Self-Image and Information Selection).

(a) HIGH increases direct avoidance of arguments against the Horncow Initiative, compared to NEUTRAL for those not initially opposing the Initiative.

(b) LOW decreases direct avoidance of arguments against the Horncow Initiative, compared to NEUTRAL for those not initially opposing the Initiative.

Hypothesis H2.A (Social Dimension and Information Avoidance).

(a) BUBBLE increases direct avoidance of arguments opposing the participant's own prior attitude, compared to NOA.

(b) CONFRONT decreases direct avoidance of arguments opposing the participant's own prior attitude, compared to NOA.

Pathway B. We now state all hypotheses relating to pathway B.

Hypothesis H1.B (Self-Image and Information Processing).

(a) HIGH increases the agreement with arguments supportive of the Horncow Initiative, compared to NEUTRAL for those who have read both types of arguments.

15 In the pre-registration of this study, we only mentioned pathway A.

(b) LOW decreases the agreement with arguments supportive of the Horncow Initiative, compared to NEUTRAL for those who have read both types of arguments.

Conditional Hypothesis H2.B (Self-Image and Votes).

(a) If H1.B (a) is true, then HIGH increases the likelihood of (i) intended and (ii) reported *Yes* votes.

(b) If H1.B (b) is true, then LOW decreases the likelihood of (i) intended and (ii) reported *Yes* votes.

We now proceed to testing these hypotheses. We correct for multiple-hypotheses testing using the Romano-Wolf correction.

3. Results

3.1 Data and descriptive statistics

The first wave of the survey was completed by 2,112 participants in German-speaking parts of Switzerland recruited from the standing LINK Institute panel that is representative for the Swiss adult population. The second wave was completed by by 1,057 participants. Summary statistics of both waves can be found in Table 2. Attrition was not random. Among those completing wave 1 but not wave 2 there were significantly more ($p = 0.0000$) women and subjects were less well informed, more emotional about and more inclined toward supporting the initiative, compared to subjects completing both waves. Despite the overrepresentation of women among those dropping out, the share of women in the final sample is still above the national average (54.0 percent in the sample vs. 50.4 percent in the population). The share of participants that supported (opposed) the initiative in the final sample are comparable to those that participated in the ballot (see Table A.2). With respect to farmers in general, farmers with horned animals, and age, there was no significant difference between the samples. Unless stated otherwise, we report results for the subsample that completed both waves of the survey. For all outcome variables elicited in wave 1, we also report results for all participants completing wave 1 to check whether sample attrition is a relevant driver.

Table 2: Summary statistics of survey waves 1 and 2

	Wave 1			Wave 2		
Variable	**Obs.**	**Mean**	**std. dev.**	**Obs.**	**mean**	**std. dev.**
Female	2,108	.592	.492	1,054	.540	.499
age (categories)	2,112	3.836	1.623	1,057	3.855	1.666
Farmer	2,109	.0123	.110	1,056	.0123	.110
FarmHorn	2,112	.00473	.0687	1,057	.00473	.0686
Informed	2,098	.0686	1.830	1,056	.3570	1.627
PriorAttitude	1,825	4.023	2.062	1,057	3.741	1.967
Emotions	1,964	.1996	1.884	1,031	-.1077	1.698

Table 3: Information selection across treatments (percent of observations)

Read	**HIGH**	**LOW**	**NEUTRAL**	**BUBBLE**	**CONFRONT**	**NOA**	**ALL**	No. obs,
Both	77.50	79.50	79.73	82.81	79.28	76.91	78.90	834
None	16.67	14.29	13.60	12.50	13.94	16.36	14.85	157
Only PRO	4.17	4.04	4.27	2.73	5.58	4.18	4.16	44
Only CONTRA	1.67	2.17	2.40	1.95	1.20	2.55	2.08	22
Opposing	78.3	81.7	81.9	84.0	80.5	79.1	80.6	852
No. obs.	360	322	375	256	251	550		1,057

Note: 'Opposing' refers to the set of arguments that oppose the prior attitude towards the initiative expressed by the participant prior to exposure to treatments. The first four rows are mutually exclusive and exhaustive, i.e. add up to the full sample completing both waves of the survey. The fifth row overlaps with rows 1, 3 and 4.

3.2 Pathway A: information selection

Strategic avoidance of information would be most obvious if treatments induced one-sided information selection. Hypothesis 1.A implies that the share of those only reading the PRO arguments should be higher (lower) in HIGH (LOW) than in NEUTRAL. Hypothesis 2.A implies that the share of those reading both arguments should be lower (higher) in BUBBLE (CONFRONT). Table 3 indicates that neither is the case. Mann-Whitney tests confirm that the distributions are not significantly different across treatments (Table 4). Logit regressions in Tables A.3 and A.4 confirm this. Hence, there is no clear treatment effect on direct information avoidance, i.e., Hypotheses 1.A and 2.A are not confirmed.

Table 4: Information Avoidance Non-Parametric Tests

Outcome Variable	**Treatments**			**Mann-Whitney** (*p*-values)	*N*
AvoidanceCONTRA	HIGH	vs.	NEUTRAL	0.309	735
	HIGH	vs.	NEUTRAL & LOW	0.280	1,057
	LOW	vs.	NEUTRAL	0.876	697
	LOW	vs.	NEUTRAL & HIGH	0.704	1,057
ReadOpposingAttitute	BUBBLE	vs.	NOA	0.102	806
	BUBBLE	vs.	NOA & CONFRONT	0.116	1,057
	CONFRONT	vs.	NOA	0.652	801
	CONFRONT	vs.	NOA & BUBBLE	0.953	1,057

Note: While the impact of BUBBLE is close to being significant at the 10%-level, the direction of the impact is the opposite of that conjectured in Hypothesis H2.A a).

In sum, testing Hypotheses 1.A and 2.A does not provide any evidence for pathway A: we do not find that voters use more information avoidance to sustain a moral self-image when salience of morality increases (HIGH) or less information avoidance when salience of morality decreases (LOW). Neither do we find that they avoid more information in a harmonious social setting and less in a confrontational setting (BUBBLE / CONFRONT) than under social insulation (NOA). Hence, either moral self-signaling and the degree of social insulation are irrelevant motivations in our voting context, or some or all of these motivations work through pathway B rather than pathway A, i.e., via biased information processing rather than biased information selection. We hence turn to testing pathway B.

3.3 Pathway B: information processing

Because the four interventions HIGH/LOW and BUBBLE/ANT did not induce any response in terms of information selection, the samples of participants reading PRO or both PRO and CONTRA arguments are likely to be unbiased by treatments. Table 5 confirms this. Only BUBBLE comes close to having a significant impact on sample composition. Hence, should BUBBLE turn out to be a significant driver of biased information processing or voting, then we would need to treat that result with caution.

Testing the hypotheses for pathway B (H1.B – H2.B) requires some care in choosing the identification strategy. *PriorAttitude*, the variable capturing a participant's attitude towards the initiative before any of the treatment interventions took place, is highly correlated with both post-treatment agreement with PRO arguments (Pearson's r = -0.6104, p = 0.0000) and with anticipated (Pearson's r = -0.7374, p = 0.0000) and reported voting (Pearson's r = 0.7235, p =

0.0000). This is not surprising, as exposure to the survey and treatment interventions are unlikely to fundamentally uncouple a participant's preferences and voting behavior from her position at the beginning of the survey. While *PriorAttitude* has immense explanatory power for the outcome variables of interest in pathway B, it also is arguably correlated with the error term as it is highly likely that it is causally affected by unobserved variables that also causally affect the outcome variables of interest. The facts that both coefficients and significance levels of treatment dummies are highly sensitive to the inclusion of *PriorAttitude* as a control variable and that *t*-values for *PriorAttitude* are an order of magnitude higher than those of other explanatory variables (see Table A.5 in the appendix) point into the same direction.

Table 5: Test for sample selection (Logit)

	(1)		(2)	
	Read PRO & CON		Read PRO	
HIGH	-0.022	(0.468)	-0.015	(0.576)
LOW	-0.009	(0.774)	-0.006	(0.838)
BUBBLE	0.061	(0.056)	0.048	(0.094)
CONFRONT	0.019	(0.537)	0.038	(0.189)
Informed	0.016	(0.064)	0.018	(0.030)
PriorAttitude	0.006	(0.356)	0.014	(0.013)
Farmer	0.115	(0.514)	0.083	(0.574)
FarmerHorn	-0.222	(0.336)	-0.047	(0.826)
Age categ.	*Yes*		*Yes*	
N	1052		1052	

Notes: Table shows estimates from Logit regressions. Dependent variables: Dummies if PRO & CON or PRO arguments have been read, respectively. Marginal effects are presented. *p*-values in parentheses unadjusted for multiple hypothesis testing.

To address this problem while still using the highly relevant information contained in *PriorAttitude*, we apply a diff-in-diff approach where *PriorAttitude* serves as the reference point. Because both *PriorAttitude* and the outcome variables *AgreementPRO*, *IntendedVote* and *ReportedVote* are measured in different but intuitively compatible categorical scales we normalize each of them to the interval (-1,1) and (0,1), respectively, before taking differences. The latter are again normalized to the interval (-1,1). This has the added advantage that the distributions of the resulting variables *ΔAgreementPRO*, *ΔIntendedVote* and *ΔReportedVote* are now close to continuous and we therefore use OLS instead of (ordered) logit regressions which eases both interpretation of coefficients and multiple hypothesis testing. For the latter, we use the

Romano-Wolf (Romano and Wolf 2005a,b, 2016 and Clarke et al. 2019) procedure based on 10,000 replications to calculate *p*-values adjusted for twenty hypotheses (five outcome variables times four treatment variables) and eight hypotheses (outcome variables *AvoidanceCONTRA, ΔIntendedVote, ΔReportedVote, ΔAgreementPRO* for treatments HIGH and LOW). Hence, this procedure takes into account that we use two outcome variables, *ΔIntendedVote* and *ΔReportedVote*, to test for hypotheses H2.B(a) and H2.B(b). As the implementation in STATA does not allow us to specify outcome-variable specific estimation methods we use OLS for all. For brevity, we only report adjusted *p*-values in those instances where the unadjusted *p*-values point towards a significant effect.

Table 6: Information Processing Non-Parametric Tests

Outcome Variable	**Treatments**			**Mann-Whitney** (*p*-values)	***N***
ΔAgreementPRO					
Read PRO & CON	HIGH	vs.	NEUTRAL	0.010	573
Read PRO & CON / Wave 1	HIGH	vs.	NEUTRAL	0.005	961
Read PRO	HIGH	vs.	NEUTRAL	0.026	604
Read PRO & CON	LOW	vs.	NEUTRAL	0.676	548
Read PRO & CON / Wave 1	LOW	vs.	NEUTRAL	0.808	924
Read PRO	LOW	vs.	NEUTRAL	0.813	577
ΔAgreementCON					
Read PRO & CON	HIGH	vs.	NEUTRAL	0.242	572

Non-parametric tests of the impact of treatments on the agreement with the arguments in favor of the initiative and on intended and reported voting are reported in Table 6. Irrespective of the sample, HIGH significantly increases the agreement with arguments in favor of the initiative relative to participants' attitude before the intervention. This is in line with hypothesis H1.B(a). There are no significant effects on agreement with arguments opposing the initiative nor of the LOW treatment on any of the outcome variables. Hypotheses H1.B(b) is not confirmed.

Table 7 uses regression analysis to confirm the above findings controlling for different sets of exogenous variables that were all elicited before treatment interventions and multiple hypothesis testing. Regressions (1) and (2) show for the sample of participants that read both PRO and CONTRA arguments that HIGH significantly increases agreement with the PRO arguments.

Table 7: Biased information processing

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			ΔAgreementPRO			*ΔAgreementCON*	
	Read PRO & CON		PRO&CON/ Wave 1	Read PRO		Read PRO & CON	
HIGH	0.061	0.056	0.055	0.054	0.051	-0.016	-0.017
	(0.005)	(0.010)	(0.001)	(0.011)	(0.016)	(0.513)	(0.483)
Romano-*Wolf*	(0.014)	(0.037)		(0.031)	(0.052)		
LOW	-0.002	-0.005	0.004	0.001	-0.000	0.016	0.015
	(0.914)	(0.824)	(0.810)	(0.979)	(0.994)	(0.524)	(0.543)
BUBBLE		-0.006	-0.003		-0.004		-0.012
		(0.777)	(0.853)		(0.863)		(0.638)
CONFRONT		-0.028	-0.025		-0.024		0.006
		(0.219)	(0.160)		(0.278)		(0.801)
Informed		-0.012	-0.008		-0.008		-0.007
		(0.064)	(0.116)		(0.209)		(0.329)
Farmer		0.061	0.025		0.064		-0.109
		(0.534)	(0.735)		(0.513)		(0.323)
FarmHorn		0.253	0.041		0.225		0.026
		(0.155)	(0.729)		(0.164)		(0.910)
Age categ.	*No*	*Yes*	*Yes*	*No*	*Yes*	*No*	*Yes*
_cons	0.119	0.187	0.150	0.120	0.166	0.007	0.054
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.688)	(0.279)
N	825	825	1390	869	868	823	823
R^2	0.013	0.046	0.024	0.009	0.038	0.002	0.016
F	5.277	2.820	2.407	4.090	2.410	0.788	0.922
aic	120.7	116.0	257.4	126.2	125.3	295.0	307.6
bic	134.9	186.7	335.9	140.5	196.8	309.2	378.3

Notes: OLS regressions. Dependent variables: *ΔAgreementPRO* and *ΔAgreementCON* capture the difference between reported convincingness of PRO/CONTRA arguments post treatment and reported prior attitude (before treatment). Both variables are normalized to values between [-1,1] with positive numbers indicating a shift in attitude toward the respective set of arguments. *p*-values in parentheses unadjusted for multiple hypothesis testing unless specified otherwise; *Romano-Wolf p*-values in (2) and (5) corrected for 20 hypotheses (outcome variables *AvoidanceCONTRA, ReadOpposingAttitute, ΔIntendedVote, ΔReportedVote, ΔAgreementPRO*) for treatments *HIGH, LOW, BUBBLE* and *CONFRONT*) and in (1) and (4) corrected for 8 hypotheses (outcome variables *AvoidanceCONTRA, ΔIntendedVote, ΔReportedVote, ΔAgreementPRO*) for treatments *HIGH and LOW,* each based on 10,000 replications.

To test for robustness and reduce the risk of issues with sample selection, we repeat the analysis with two further samples. The sample in regression (3) consists again of those having read the arguments of both sides but based on all participants that completed wave 1 of the survey rather than on those that completed both waves. This adds another 565 observations, namely those that did not complete wave 2. Regressions (4) and (5) cover all participants completing both waves that read the PRO arguments, i.e. compared to the sample of regressions (1) and (2) they include

the 44 participants that did not read the CONTRA arguments. The treatment effect of HIGH is robust to both variations in the sample and to the correction for multiple hypothesis testing. Agreement with arguments against the initiative, however, is not affected by any of the treatments (regressions (6) and (7) in Table 7). In sum, we do find robust evidence for the use of biased information processing when morality becomes more salient, as in HIGH. If moral self-signaling is behind this bias, then HIGH will affect intended and reported voting behavior, too, as we hypothesized for pathway B. We hence now turn to investigating intended and reported votes.

3.4 Pathway B: voting behavior

Intended voting relative to participants' attitude towards the initiative before the intervention (*ΔIntendedVote*) exhibits the same pattern as the agreement with PRO arguments. The impact of HIGH on *ΔReportedVote* is only significant at the 10%-level. This gives some initial support of Hypothesis H2.B (a).

Table 8: Information Processing Non-Parametric Tests

Outcome Variable	**Treatments**			**Mann-Whitney** (*p*-values)	*N*
ΔIntendedVote					
	HIGH	vs.	NEUTRAL	0.001	712
Wave 1	HIGH	vs.	NEUTRAL	0.025	1206
	LOW	vs.	NEUTRAL	0.670	674
Wave 1	LOW	vs.	NEUTRAL	0.362	1144
ΔReportedVote					
	HIGH	vs.	NEUTRAL	0.064	529
	LOW	vs.	NEUTRAL	0.423	516

Figure 2 splits the share of reported *YES* votes by treatment and three categories of *PriorAttitude*. In all categories the share of *YES* votes is higher in the HIGH treatment than in NEUTRAL but the difference is statistically significant (5%-level) only among those initially opposed to the initiative. The latter group contains more than half of all participants in the HIGH and NEUTRAL treatments that reported their vote in wave 2 of the survey. Note that the attitude categories in Figure 2 are based on measurements prior to exposure to treatments and that the HIGH treatment induced a bias in favor of the PRO arguments relative to participants' prior attitude. Hence, the stronger impact on votes for those initially opposing the initiative is not surprising for two reasons. First, identifying a PRO-bias in a group that already reports agreement ex-ante is much harder than in a group that initially is more skeptical. Second, with a constant but small share of

additional *YES* votes induced by the treatment, they are much more likely to occur and found significant, the larger the number pre-treatment *NO* voters.[16]

Figure 2: Share of reported *YES* votes by categories of prior attitude

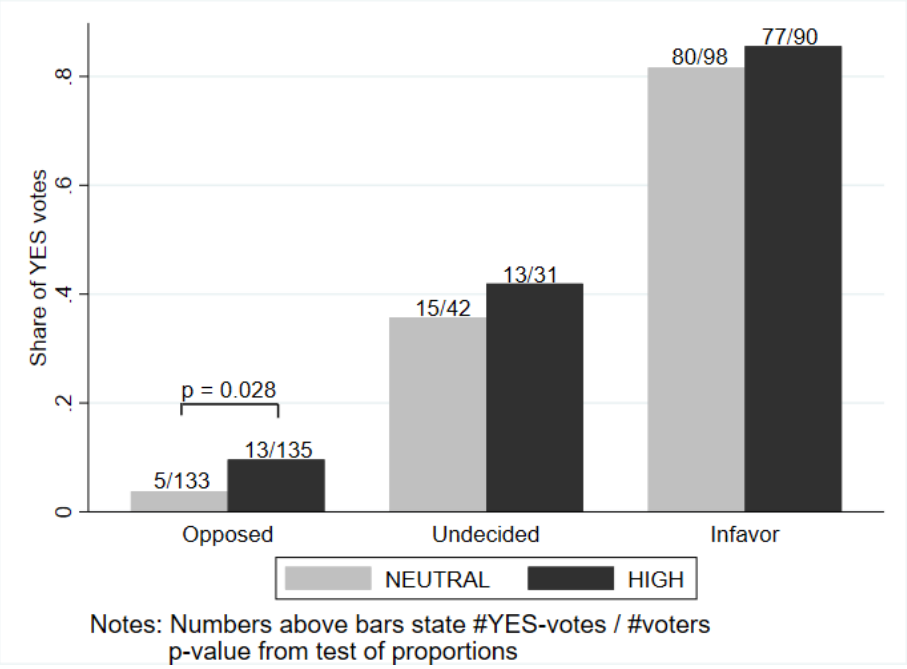


Table 9 presents additional results of OLS regressions on *ΔIntendedVote* and *ΔReportedVoting*. The sample of regressions (1) to (3) contains all participants that completed both waves of the survey whereas regression (4) also includes those that did not complete wave 2 of the survey which increases the sample size by 718 participants relative to (2). The sample in regressions (5) and (6) is smaller as it includes only those who have completed both waves of the survey and reported to have participated in the ballot.

All specifications show a significant impact of HIGH on intended and reported voting. Significance is robust to multiple hypothesis testing. In sum, we clearly find evidence for moral self-signaling along pathway B: Exposure to the intervention that raised the salience of moral self-signaling by voting in favor of the initiative, while having no impact on information selection, did increase agreement with PRO arguments and intended and reported actual voting in favor of the Horncow Initiative. Overall, this provides strong support for the modelling paradigm developed by Bénabou and Tirole (2002, 2006, 2011) and the entire literature building on them. As it seems, in the field the preferred strategy to keep up a positive moral self-image can also be not to entirely avoid information that would undermine moral self-signaling strategies but to assign higher weights to information that helps rationalizing such strategies.

16 For the sample presented in Figure 2, the difference in the share of *NO* votes between the NEUTRAL and the HIGH treatment is: 5.87 percent for those initially opposed, 6.23 percent for those initially undecided and 3.93 percent for those initially in favor.

Table 9: Impact on intended/reported voting (OLS)

	(1)	(2)	(3)	(4)	(5)	(6)
		ΔIntendedVote		*ΔInt.V./wave 1*		*ΔReportedVote*
HIGH	0.041	0.046	0.044	0.025	0.059	0.070
	(0.003)	(0.001)	(0.003)	(0.025)	(0.046)	(0.018)
Romano-Wolf	(0.012)	(0.005)			(0.092)	(0.043)
LOW	-0.004	-0.002	-0.002	0.003	0.018	0.030
	(0.783)	(0.889)	(0.873)	(0.786)	(0.555)	(0.322)
BUBBLE		0.023	0.023	0.013		0.016
		(0.107)	(0.114)	(0.242)		(0.603)
Romano-Wolf		(0.366)				(0.843)
CONFRONT		0.003	0.002	-0.014		0.026
		(0.849)	(0.869)	(0.210)		(0.399)
Informed		-0.009	-0.009	0.010		-0.013
		(0.023)	(0.030)	(0.001)		(0.136)
Farmer		0.065	0.066	0.037		0.146
		(0.360)	(0.358)	(0.488)		(0.294)
FarmHorn		0.104	0.105	0.059		0.257
		(0.342)	(0.339)	(0.468)		(0.284)
GoodIntent			0.006			
			(0.179)			
HIGHxGoodIntent			-0.005			
			(0.552)			
Age categ.	*No*	*Yes*	*Yes*	*Yes*	*No*	*Yes*
_cons	-0.001	-0.064	-0.024	-0.053	-0.161	-0.240
	(0.907)	(0.210)	(0.282)	(0.468)	(0.000)	(0.000)
N	1021	1020	987	1738	772	771
R^2	0.012	0.032	0.034	0.017	0.005	0.033
F	6.195	2.400	2.109	2.138	2.076	1.824
Aic	-525.4	-521.1	-493.3	-789.3	528.8	529.0
Bic	-510.6	-447.2	-410.1	-707.4	542.8	598.6

Notes: Dependent variables: *ΔIntendedVote* and *ΔReportedVote* capture the difference between reported planned/actual vote and reported prior attitude (before treatment). Both variables are normalized to values between [-1,1] with positive numbers indicating a shift in attitude toward support of the horncow initiative. *p*-values in parentheses unadjusted for multiple hypothesis testing unless specified otherwise; *Romano-Wolf p*-values in (2) and (6) corrected for 20 hypotheses (outcome variables *AvoidanceCONTRA, ReadOpposingAttitute, ΔIntendedVote, ΔReportedVote, ΔAgreementPRO*) for treatments *HIGH, LOW, BUBBLE* and *CONFRONT*) and in (1) and (5) corrected for 8 hypotheses (outcome variables *AvoidanceCONTRA, ΔIntendedVote, ΔReportedVote, ΔAgreementPRO*) for treatments *HIGH and LOW* each based on 10,000 replications.

In order to validate this interpretation, we test a potential alternative explanation for the effects of HIGH. This explanation hypothesizes that those exposed to the HIGH treatment regard the instigators of the Horncow Initiative as being driven by good intentions and try to reward them

by voting in favor of the initiative. The variables used to test this idea are *GoodIntent* measuring the degree to which participants agree with the claim that good intentions rather than good consequences of actions should be rewarded and its interaction with HIGH (*HIGH x GoodIntent*). We do not find evidence for the alternative explanation of our results (see (3) in Table 9).

Without controlling for *PriorAttitute* or any other control variables the share of *Yes* votes in the HIGH treatment is 40.2 percent relative to 36.6 percent in the NEUTRAL treatment. If only 36.6 percent of the 256 participants (i.e. 94 participants) exposed to the HIGH treatment that reported their voting decision had voted in favor of the initiative, then we would have seen nine fewer *Yes* votes. If we assume the same impact for all 719 participants exposed to the HIGH treatment, i.e. including those that did not complete wave 2 of the survey or did not report their vote, then the number of *Yes* votes has increased by 26 due to the experimental intervention. Using the coefficient from regression (6) in Table 9, i.e. 0.07 percent, the number of *Yes* votes increased by 18 in the sample reporting their vote and by 50 in the full sample exposed to the HIGH treatment. For comparison, in the ballot the number of No votes exceeded the number of *Yes* votes by 239,182 out of 2.6 million votes cast.

4. Further results

In this section we report a number of interesting correlations between variables elicited in the survey as detailed in the pre-registration. However, these relationships cannot be interpreted causally, and several variables were elicited after the treatment intervention (*GoodIntent, GoodEffect, FreqMeat, Intensive, Vegetarian, Vegan, NoEggsMilk, Overconfident*) and hence can correlate due to past exposure to these interventions. Table 10 shows that anticipated as well as reported votes in favor of the initiative decreases in the frequency of meat eating but not with other dietary habits related to animal products.

In addition, we test whether the self-reported level of prior informedness on the initiative correlated to specific ethical attitudes toward consequentialism. These attitudes are expressed by, first, the degree to which participants report to agree with a claim stating that rewards should be given to those whose actions result in good consequences regardless of his or her intentions (*GoodEffects*), and, second, the degree to which they agree with a claim stating that rewards should be given to those with good intentions regardless of the consequences of these actions (*GoodIntent*). If looked at separately, we find a negative correlation. In a joint analysis (regression (3) in Table 11), only *GoodIntent* remains significant. Hence, participants that reported to care for intentions behind actions also report to be less well informed. This is consistent as knowing the consequences of one's actions (and votes) is less relevant if one's focus is on intentions rather than consequences.

Table 10: Non-causal analysis: Voting

	(1)	(2)	(3)	(4)	(5)
	IntendedVote		*Int.V./wave 1*	*ReportedVote*	
FreqMeat	-0.096	-0.147	-0.174	-0.189	-0.206
	(0.018)	(0.075)	(0.001)	(0.001)	(0.038)
Intensive		0.063	0.103		-0.003
		(0.728)	(0.413)		(0.991)
Vegetarian		-0.436	-0.237		-0.188
		(0.390)	(0.500)		(0.782)
Vegan		-0.361	-0.407		-1.166
		(0.719)	(0.580)		(0.489)
NoEggsMilk		0.502	0.312		0.809
		(0.285)	(0.366)		(0.295)
_cons				0.469	0.545
				(0.107)	(0.245)
N	1021	1018	1949	772	770
chi^2	5.612	7.539	23.455	11.328	12.671
Pseudo R^2	0.002	0.002	0.004	0.011	0.012
Aic	3272.2	3269.2	6246.4	1021.5	1024.3
Bic	3296.9	3313.5	6296.6	1030.8	1052.2

Notes: Ordered Logit regression. Dep. Var.: *IntendedVote* and *ReportedVote*, reported are coefficients, *p*-values in parentheses

Table 11: Non-causal analysis: Prior information

	(1)	(2)	(3)
GoodEffects	-0.070		-0.038
	(0.048)		(0.303)
GoodIntent		-0.095	-0.073
		(0.005)	(0.043)
N	1015	1019	1001
chi^2	3.90	7.89	7.10
Pseudo R^2	0.001	0.002	0.002
Aic	3760.5	3764.3	3703.9
Bic	3795.0	3798.8	3743.2

Notes: Ordered Logit regression. Dependent variable: *Informed*; coefficients reported, *p*-values in parentheses

Furthermore, both variables measuring information selection that we used in the previous section are not significantly correlated with proxies of ethical schools of thought (Table 12). However, they are highly significantly correlated with both how emotionally touched participants are by

the initiative (*Emotions*) and how much their self-assessed prior informedness (*Informed*) with respect to the initiative differs from their performance in a quiz about the initiative and horned animals (*Overconfident*). The latter is a dummy that equals one if a participants is above the median with respect to self-reported informedness but below the median in terms of quiz performance. Emotional involvedness is associated with more and overconfidence with less information selection.

Table 12: Non-causal analysis: Information selection

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
		AvoidanceCONTRA				*ReadOpposingAttitute*		
GoodEffects	-0.088			-0.053	0.086			0.049
	(0.106)			(0.348)	(0.110)			(0.387)
GoodIntent	0.048			0.029	-0.101			-0.088
	(0.362)			(0.597)	(0.052)			(0.107)
Emotions		-0.150	-0.170	-0.168		0.149	0.147	0.171
		(0.001)	(0.000)	(0.001)		(0.001)	(0.000)	(0.001)
Overconfident		1.070	0.895	1.155		-1.127	-0.989	-1.189
		(0.000)	(0.000)	(0.000)		(0.000)	(0.000)	(0.000)
_cons	-1.500	-1.809	-1.539	-1.883	1.467	1.809	1.610	1.866
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	1002	1031	1964	979	1002	1031	1783	979
chi^2	2.84	50.40	80.65	55.66	4.95	55.11	79.34	60.70
Pseudo R^2	0.003	0.050	0.040	0.060	0.005	0.055	0.044	0.065
Aic	955.8	955.6	1961.4	882.7	962.6	959.6	1747.6	883.7
Bic	970.5	970.4	1978.2	907.1	977.3	974.4	1764.1	908.1

Notes: Logit regression. Dependent variable: *AvoidanceCONTRA* (regressions (1) – (4)) and *ReadOpposingAttitute* (regressions (5) – (8)). Regressions (3) and (5) based on all participants that completed wave 1 of the survey, all other regressions based on sample completing both waves. *p*-values in parentheses.

Information acquisition in wave 1 is consistent with participants suffering from confirmation bias. Those ex-ante supporting (opposing) the initiative are more likely to only read the arguments supporting (opposing) the initiative.[17]

[17] Both Chi^2-tests and univariate logit regressions yield $p<0.01$. Results available upon reported.

5. Concluding remarks

In the context of a randomized controlled trial around a popular vote on animal welfare in Switzerland, we find evidence of *moral self-image* concerns as a motivation behind voting behavior. We do not find evidence for social-insulation effects. An increase in the moral self-signaling value of voting in favor of the initiative did not affect information selection prior to voting but did instead bias information processing. Participants exposed to scientific evidence establishing a correlation between kindness towards animals and kindness towards fellow humans assigned significantly more importance to the arguments supporting the initiative than those not exposed to such evidence and were more likely to vote in favor of the initiative. This specifies a precise channel through which individuals, in particular voters, generate motivated biases as modelled in Bénabou and Tirole (2002, 2006, 2011) and the literature that builds on them.

The arguments that our subjects could choose to read after our informational intervention were carefully chosen such as not to present real news (they were largely identical to information provided to all voters ahead of the vote via official channels). Hence, if information selection rather than information processing was the prevalent way to keep up one's moral self-image, it is likely that our design would have had no or only a negligible effect on voting behavior. However, we could not preclude non-negligible effects for biased information processing. The biased weighting of the arguments by those exposed to the treatment that increased the salience of morality resulted in more anticipated and more reported actual votes in favor of the initiative. However, our experiment had no decisive effect on the outcome of the ballot: though the vote turned out quite close, the Horncow Initiative was refuted in the end.

References

Andreoni, J., Rao, J.M. & Trachtman, H. (2017). Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of Political Economy* 125(3): 625-53.

Bénabou, R., Falk, A., Henkel, L. & Tirole, J. (2020). Eliciting moral preferences. Mimeo.

Bénabou, R. & Tirole, J. (2002). Self-confidence and personal motivation. *Quarterly Journal of Economics 117*(3): 871-915.

Bénabou, R. & Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review* 96(5): 1652-78.

Bénabou, R. & Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *Quarterly Journal of Economics* 126(2): 805-55.

Brennan, G. & Hamlin, A. (1998). Expressive voting and electoral equilibrium. *Public Choice* 95(1-2): 149-75.

Brennan, G. & Lomasky, L. (1993). *Democracy and decision: The pure theory of electoral preference*. Cambridge University Press.

Clarke, D., Romano, J.P. & Wolf, M. (2019). The Romano-Wolf multiple hypothesis correction in Stata. IZA Discussion Paper No. 12845.

Dana, J., Weber, R. & Kuang, J.X. (2007). Exploiting moral wiggle room: Experiments demonstrating an illusory preference for fairness. *Economic Theory* 33: 67-80.

Della Vigna, S., List, J.A., Malmendier, U. & Rao, G. (2016). Voting to tell others. *Review of Economic Studies* 84(1): 143-181.

Ehrich, K.R. & Irwin, J.R. (2005). Willful ignorance in the request for product attribute information. *Journal of Marketing Research* 42(3): 266-77.

Enke, B. (2020): Moral values and voting. *Journal of Political Economy* 128(10): 3679-729.

Feddersen, T. & Sandroni, A. (2006). A theory of participation in elections. *American Economic Review* 96(4): 1271-82.

Feddersen, T., Gailmard, S. & Sandroni, A. (2009). Moral bias in large elections. Theory and experimental evidence. *American Political Science Review* 103(2): 175-92.

Freddi, E. (2017). Do people avoid morally relevant information? Evidence from the refugee crisis. CentER Discussion Paper Series No. 2017-034.

Garrett, K.N. & Bankert, A. (2020). The moral roots of partisan division: How moral conviction heightens affective polarization. *British Journal of Political Science* 50(2): 621-40.

Gerber, A.S. & Green, D.P. (2000). The effects of canvassing, telephone calls, and direct mail on voter turnout: A field experiment. *American Political Science Review* 94(3): 653-63.

Gerber A.S., Green, D.P. & Larimer, C.W. (2008). Social pressure and voter turnout: Evidence from a large-scale field experiment. *American Political Science Review* 102(1): 33-48.

Gerber, A.S. & Rogers, T. (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *Journal of Politics* 71(1): 178-91.

Grossman, Z. & Van der Weele, J.J. (2017). Self-image and willful ignorance in social decisions. *Journal of the European Economic Association* 15(1): 173-217.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.

Hestermann, N., Le Yaouanq, Y. & Treich, N. (2020). An economic model of the meat paradox. *European Economic Review* 129: 103569.

Hüning, H., Mechtenberg, L., & Wang, S. (2020). Tell me who you speak with and I tell you how you vote. An online-chat experiment on a ballot on rent control. Mimeo.

Kendall, C., Nannicini, T. & Trebbi, F. (2015). How do voters respond to information? Evidence from a randomized campaign. *American Economic Review* 105(1): 322-53.

Nickerson, D.W. & Rogers, T. (2010). Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making. *Psychological Science* 21(2): 194-9.

Niehaus, P. (2020). A theory of good intentions. Mimeo, UC San Diego.

Nyborg, K. (2011). I don't want to hear about it: Rational ignorance among duty-oriented consumers. *Journal of Economic Behavior & Organization* 79(3): 263-74.

Rabin, M. & Schrag, J. L. (1999). First impressions matter: A model of confirmatory bias. *Quarterly Journal of Economics* 114(1): 37-82.

Romano, J.P. & Wolf, M. (2005a). Exact and approximate stepdown methods for multiple hypothesis testing. *Journal of the American Statistical Association* 100(469): 94-108.

Romano, J.P. & Wolf, M. (2005b). Stepwise multiple testing as formalized data snooping. *Econometrica* 73(4): 1237-82.

Romano, J.P. & Wolf, M. (2016). Efficient computation of adjusted *p*-values for resampling-based stepdown multiple testing. *Statistics & Probability Letters* 113: 38-40.

Sandel, M. (2006): *Public Philosophy: Essays on Morality in Politics*. Harvard UP.

Serra-Garcia, M. & Szech, N. (2019). The (in)elasticity of moral ignorance. CESifo Working Paper No. 7555.

Shayo, M. & Harel, A. (2012). Non-consequentialist voting. *Journal of Economic Behavior and Organization* 81(1): 299-313.

Tyran, J.-R. (2004). Voting when Money and Morals Conflict. An Experimental Test of Expressive Voting. *Journal of Public Economics* 88(7): 1645-64.

Zaller, J. R. (1992). *The nature and origins of mass opinion*. Cambridge University Press.

Appendix

Table A.0: Information provided in treatments HIGH and LOW

The interventions HIGH and LOW had the following wording (translation from German):

Treatment	Text shown
HIGH	Did you know that according to a scientific study (Arluke and Madfis 2013, available on request) cruelty to animals and anti-social behaviour towards humans are correlated? The study reports that those being cruel to animals are more likely to conduct criminal acts against humans. Examples from the study: ● Someone torturing animals is much more likely to be violent against humans than someone who is kind towards animals. ● Someone torturing animals is much more likely to run amok than someone who is kind towards animals. ● Someone torturing animals is much more likely to disrespect property rights than someone who is kind towards animals. According to psychological research a common cause of anti-social behavior is a lack of compassion (empathy). Another study (Erlanger und Tsytsarev 2012, available on request) shows that: Compassionate people are much more likely to treat animals kindly than non-compassionate people. Compassionate people are much more opposed to cruelty to animals and animal testing than non-compassionate people. Being compassionate is a necessary condition for kind-hearted behavior. Overall this implies: Kind-hearted people who care about the wellbeing of others and the good rules of living together are also more caring towards animals!
LOW	Did you know that according to a scientific study (Levin, Arluke und Irvine 2017, available on request) care for animals and indifference towards humans can co-exist? The study reports that those helping animals might well ignore the suffering of other humans. Examples from the study: ● A call for donations to help a sickly dog motivated more people to donate than a call for donations of a sickly child.

	● A dog that had been knocked out induced an emotional response in more people than an adult that had been knocked out. What is the reason for some people to be more indifferent towards other people than towards animals? According to the researchers, a possible reason is that such people believe humans but not animals to be responsible ("at fault") for their own hardship. The following true event provides further evidence for the possibility that compassion towards animals and indifference towards humans can co-exist: In a western industrialized country many people actively protested that a police officer who shot a dog out of an unfounded feeling of threat gets punished. The same people did not care whether a police officer who shot a mentally ill woman out of an unfounded feeling of threat gets punished. Being compassionate is a necessary condition for kind-hearted behavior. Overall this implies: People that care about animals are not necessarily kind-hearted people that care about the wellbeing of others and the good rules of living together!

Table A.1: Variable descriptions

Variable name	**Description**
Explanatory variables	
HIGH	Dummy variable that equals 1 if participant in HIGH treatment
LOW	Dummy variable that equals 1 if participant in LOW treatment
BUBBLE	Dummy variable that equals 1 if participant in BUBBLE treatment
CONFRONT	Dummy variable that equals 1 if participant in CONFRONT treatment
Female	Dummy variable that equals 1 if participant reports to be female (rather than male or other).
Age categ.	Dummies for eight age categories. Lowest bracket: '18-24 years', then in steps of ten years up to 84. Highest bracket: 'above 84'.
Farmer	Dummy variable that equals 1 if participant reports to work as a farmer
FarmHorn	Dummy variable that equals 1 if participant reports to keep horned farm animals in particular horned cows or goats
Informed	Categorial variable centered around zero with seven categories. -3 indicated that the participant reports to be 'not at all informed' and 3 that the participant reports to be 'very well informed' about the Horncow Initiative and the upcoming ballot
PriorAttitude	Categorial variable on a seven point Likert scale measuring the attitude

	towards the Horncow Initiative. 1 represents 'Certainly against' and 7 'certainly in favor'.
Emotions	Categorial variable centered around zero with seven categories. -3 indicated that the participant reports that (s)he does 'not at all' and 3 that the participant reports to 'very much' respond emotionally to the Horncow Initiative.
FreqMeat	Categorial variable on an eight point scale capturing the self-reported estimate of the frequency of eating red or white meat or meat products such as sausages, ham and entrails. Categories: 1: never; 2: only as an exception; 3: once a month; 4: several times a month; 5: once a week; 6: several times a week; 7: once a day; 8: several times a day.
Intensive	Dummy variable that equals 1 if participant reports to eat meat at least once a day. Constructed from *FreqMeat.*
Vegetarian	Dummy variable that equals 1 if participant reports never to eat meat. Constructed from *FreqMeat.*
Vegan	Dummy variable that equals 1 if participant reports to adhere to a vegan diet.
NoEggsMilk	Dummy variable that equals 1 if participant reports not to eat eggs and milk.
GoodEffects	Categorial variable on a seven point Likert scale measuring the attitude towards the statement that consequences are more important than intentions of someone's actions. 1 represents 'Certainly against' and 7 'certainly in favor'.
GoodIntent	Categorial variable on a seven point Likert scale measuring the attitude towards the statement that intentions are more important than consequences of someone's actions. 1 represents 'Certainly against' and 7 'certainly in favor'.
HIGHxGoodIntent	Interaction between variables *HIGH* and *GoodIntent*
Overconfident	Dummy variable that equals 1 if participant's self-reported degree of informedness (based on variable *Informed*) is above the median response (=0) but at the same time the participant's performance in the quiz is below the median performance (8 out of 10 questions correctly answered).
Outcome variables	
AvoidanceCONTRA	Dummy variable that equals 1 if participant chooses not to read the arguments opposing the Horncow Initiative.
ReadOpposingAttitute	Dummy variable that equals 1 if participant chooses to read the arguments opposing his/her own *PriorAttitude* towards the Horncow Initiative.
ΔIntendedVote	Continuous variable bound to interval [-1,1] capturing the normalized difference between the self-reported anticipated voting at the end of the first wave and *PriorAttitude*. The variable is computed as follows: *ΔIntendedVote = [(AnticipatedVoting - 3)/2 - (PriorAttitude - 4)/3]/2* such that negative numbers indicate that the likelihood to vote in favor of the initiative has decreased relative to the attitude expressed before the exposition to the PRO and/or CONTRA arguments. Where *AnticipatedVoting* is a categorical variable that measures the participant's voting plan in the ballot: 1 'certainly vote against the initiative'; 2 'likely to vote against the initiative'; 3 'I have not yet formed an opinion on how to vote', 4 'likely to vote infavor of the initiative', 5 'certainly vote infavor of

	the initiative'.
ΔReportedVote	Continuous variable bound to interval [-1,1] capturing the normalized difference between the self-reported actual voting and *PriorAttitude*. The variable is computed as follows: *ΔReportedVote = ReportedVoting - PriorAttitude/7* such that negative numbers indicate that the likelihood to vote in favor of the initiative has decreased relative to the attitude expressed before the exposition to the PRO and/or CONTRA arguments. Where *ReportedVoting* is a dummy that equals 1 if the participant reports to have voted in favor of the initiative.
ΔAgreementPRO	Continuous variable bound to interval [-1,1] capturing the normalized difference between the self-reported agreement with arguments infavor of the initiative at the end of the first wave and *PriorAttitude*. The variable is computed as follows: *ΔAgreementPRO = [(AgreementPRO - 3)/2 - (PriorAttitude - 4)/3]/2* such that negative numbers indicate that the agreement with arguments in favor of the initiative has decreased relative to the attitude expressed before the exposition to the PRO and CONTRA arguments. Where *AgreementPRO* is a categorical variable capturing how much the participant the PRO arguments he/she has just read convince him/her: 1 'not at all convincing'; 2 'more unconvincing than convincing, 3 'neither convincing nor unconvincing', 4 'more convincing than unconvincing', 5 'fully convincing'.
ΔAgreementCON	Continuous variable bound to interval [-1,1] capturing the normalized difference between the self-reported agreement with arguments against of the initiative at the end of the first wave and *PriorAttitude*. The variable is computed as follows: *ΔAgreementCON = [(AgreementCONTRA - 3)/2 - (PriorAttitude - 4)/3]/2* such that negative numbers indicate that the agreement with arguments in favor of the initiative has decreased relative to the attitude expressed before the exposition to the PRO and CONTRA arguments. Where *AgreementCONTRA* is a categorical variable capturing how much the participant the CONTRA arguments he/she has just read convince him/her: 1 'not at all convincing'; 2 'more unconvincing than convincing, 3 'neither convincing nor unconvincing', 4 'more convincing than unconvincing', 5 'fully convincing'.

Table A.2: Attitudes and Voting in Sample vs. Ballot

	In favor	**Opposing**	**Neutral**	*N*
Participants completing wave 1				
PriorAttitude	40.4	42.8	16.8	1,825
AnticipatedVoting	36.6	40.6	22.8	1,954
Participants completing both waves				
PriorAttitude	45.3	37.2	17.5	1,057
AnticipatedVoting	43.9	36.7	19.4	1,021
ReportedVoting	38.5	61.5		772
Ballot Result				
all of Switzerland	45.3	52.9		2.62 million
German speaking cantons	43.8	53.5		1.93 million

Note: Source of ballot results: Bundesamt für Statistik, Statistik der eidg. Volksabstimmungen (Abst.-Nr. 6230)

Table A.3: Avoidance of CONTRA arguments (Logit)

	(1) Logit	(2) OLS	(3) Logit	(4) OLS	(5) OLS/wave 1
HIGH	0.048 (0.206)	0.049 (0.199)	0.050 (0.180)	0.051 (0.189)	0.034 (0.236)
LOW	0.000 (0.999)	0.000 (0.999)	0.009 (0.821)	0.008 (0.843)	-0.024 (0.406)
BUBBLE			-0.056 (0.197)	-0.049 (0.226)	0.029 (0.322)
CONFRONT			0.024 (0.513)	0.026 (0.499)	0.027 (0.363)
Informed			-0.027 (0.024)	-0.026 (0.029)	-0.002 (0.824)
Farmer			-1.749 (0.989)	-0.152 (0.573)	-0.218 (0.266)
FarmHorn			1.931 (0.988)	0.439 (0.251)	0.318 (0.252)
PriorAttitude			0.007 (0.639)	0.008 (0.639)	0.005 (0.647)
Age categ.	*No*	*Yes*	*No*	*Yes*	*Yes*
_cons		0.160 (0.000)		0.162 (0.109)	0.128 (0.099)
N	578	578	576	576	1084
R^2		0.004		0.033	0.015
F		1.064		1.386	1.108
Aic		529.3		526.7	1042.0
Bic		542.4		592.0	1121.8

Notes: Dependent variable: Dummy variable whether CONTRA arguments have been avoided (1 = avoided, 0 = read). For the logit regressions (1) and (3) marginal effects are presented. *p*-values in parentheses unadjusted for multiple hypothesis testing. *Romano-Wolf p*-values not reported as even unadjusted *p*-values do not allow to reject null hypothesis.

Table A.4: Reading arguments opposing own attitude

	(1) Logit	(2) OLS	(3) Logit	(4) OLS	(5) OLS/wave 1
BUBBLE	0.051	0.049	0.049	0.047	0.003
	(0.102)	(0.102)	(0.111)	(0.114)	(0.896)
Romano-Wolf				(0.366)	
CONFRONT	0.013	0.014	0.009	0.009	-0.013
	(0.652)	(0.645)	(0.770)	(0.775)	(0.585)
HIGH			-0.035	-0.035	-0.017
			(0.230)	(0.229)	(0.467)
LOW			-0.008	-0.008	0.002
			(0.793)	(0.786)	(0.922)
Informed			0.015	0.015	-0.000
			(0.073)	(0.076)	(0.962)
Farmer			0.090	0.080	0.154
			(0.587)	(0.569)	(0.160)
FarmHorn			-0.212	-0.248	-0.152
			(0.328)	(0.271)	(0.367)
PriorAttitude			0.006	0.006	0.012
			(0.352)	(0.333)	(0.009)
Age categ.	*No*	*Yes*	*No*	*Yes*	*Yes*
_cons		0.791		0.808	0.795
		(0.000)		(0.000)	(0.000)
N	1057	1057	1052	1055	1817
R^2		0.003		0.018	0.013
F		1.339		1.304	1.544
Aic		1041.4		1042.1	1892.2
Bic		1056.3		1121.5	1980.2

Notes: Dependent variable: Dummy variable whether arguments opposing ones' own prior attitude have been read (1 = read, 0 = avoided). For the logit regressions (1) and (3) marginal effects are presented. *p*-values in parentheses unadjusted for multiple hypothesis testing unless specified otherwise; *Romano-Wolf p*-values in (4) corrected for 20 hypotheses (outcome variables *ReadOpposingAttitute, AvoidanceCONTRA, ΔIntendedVote, ΔReportedVote, ΔAgreementPRO*) for treatments *HIGH, LOW, BUBBLE* and *CONFRONT*) based on 10,000 replications.

Table A.5: Using prior attitude as control

	AgreementPRO		*IntendedVote*		*ReportedVote*	
	(1)	(2)	(3)	(4)	(5)	(6)
HIGH	-0.130 (-1.54)	-0.183 (-2.75)	0.057 (0.56)	0.155 (2.83)	0.035 (0.81)	0.064 (2.20)
LOW	-0.058 (-0.68)	-0.025 (-0.37)	0.021 (0.19)	-0.014 (-0.25)	0.022 (0.51)	0.017 (0.56)
BUBBLE	0.170* (2.01)	0.098 (1.46)	-0.047 (-0.45)	0.071 (1.25)	-0.030 (-0.69)	0.020 (0.68)
CONFRONT	0.139 (1.60)	0.111 (1.62)	-0.014 (-0.14)	-0.001 (-0.02)	0.006 (0.14)	0.017 (0.56)
PriorAttitude		-0.317 (-22.24)		0.586 (50.55)		0.170 (29.15)
_cons	2.587 (37.60)	3.813 (49.26)	2.870 (34.63)	0.619 (9.86)	0.372 (10.88)	-0.271 (-8.40)
N	825	825	1021	1021	772	772
R^2	0.010	0.383	0.001	0.716	0.002	0.527
Aic	2345.8	1958.0	3542.9	2260.8	1087.1	512.8
Bic	2369.3	1986.3	3567.5	2290.4	1110.3	540.7

Notes: t statistics in parentheses