Self-Signaling in Voting

Lydia Mechtenberg, Grischa Perino, Nicolas Treich,

Jean-Robert Tyran, and Stephanie W. Wang[1]

May 20, 2023

Abstract

This paper presents a two-wave survey experiment to examine the impact of self-image concerns on voting behavior. We elicit votes on a ballot initiative on animal welfare in Switzerland that spurred campaigns involving widely shared normative values. We investigate how messages that change the self-signaling value of a vote in favor of the initiative affect selection and processing of information, as well as reported voting behavior. We find that a message enhancing the self-signaling value of a Yes vote is effective in several ways: voters agree more with arguments in favor of the initiative, are more likely to anticipate voting in favor, and do report having voted in favor of the initiative more often.

JEL-codes: C93, D72, D91

Keywords: voting, self-image, multi-wave field experiment, information processing, animal welfare

---

Mechtenberg: 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 Department of Economics, University of Economics in Bratislava, Jean-Robert.Tyran@univie.ac.at, Wang (corresponding author): 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).

This study was pre-registered with the AER RCT Registry as #AEARCTR-0003551 and received IRB-approval of the ethics committee of the University of Hamburg.

1 Introduction

People often like to think that they are good members of society. They may donate money to strangers (Adena and Huck, 2020), express politically correct opinions (Monin and Miller, 2001), dislike being paid for harming others (Fiorin, 2022), avoid information on potentially negative consequences of their actions (Dana et al., 2007), and manipulate information they receive about themselves (Köszegi, 2006). In this paper, we explore the idea that voting on morally relevant policies may also serve as a self-signaling device: By supporting a cause that is framed as the ethical choice, individuals can signal to themselves that they are good people.

To test this idea, we designed an experiment in which we change the extent to which voting for an actual, ethically framed policy can contribute to feeling good about oneself, i.e., we change the self-signaling value of a Yes vote. Our setting is a ballot in Switzerland on a popular initiative called "for the dignity of farm animals" that was framed as an animal-welfare policy. Whether or not the initiative would indeed improve animal welfare if successful was contested. The government and the media informed voters about both sides of the debate. In a large online two-wave survey experiment ($N > 1000$), we sent an evidence-based message to randomly selected voters. This message informed them of research on the correlation between being animal-friendly and being kind in general, i.e., also towards other humans ("Good-hearted people tend to be good to animals"). The purpose of this information intervention (Haaland et al., 2023) is to increase the self-signaling value of voting for the initiative: believing in the animal-friendliness of the initiative is more valuable for the treated than the untreated voters due to increased self-image gains.[2] We show that our intervention has an effect on voters' opinion in the debate about the initiative's animal-friendliness and their intended and reported voting decisions. This suggests that treated voters did use the vote as a means to self-signal that they are good people.

We also investigate the channel(s) through which self-signaling affects voting decisions. Building on the motivated beliefs literature (Bénabou and Tirole, 2002, 2011; Di Tella et al., 2015; Grossman and van der Weele, 2017), we hypothesize that subjects can improve their self-image through the selection and processing of pro vs. con arguments before voting, and

[2] Previous studies have shown that self-image (and related concepts such as self-esteem, self-view or self-awareness) can be momentarily altered (Heatherton and Polivy, 1991; Gao et al., 2009).

by voting according to their thus biased beliefs. To see this, note that both treated and untreated voters in our experiment must first of all believe in the animal-friendliness of the initiative to be able to use voting in favor of it as a signal of being a good-hearted person. Doing so pays off more in terms of potential self-image gains for the treated than for the untreated voters. Hence, our treated voters have a higher psychic incentive to select and overweigh arguments that make the animal-friendliness of the initiative more salient.

To investigate these potential channels, we let voters in our experiment choose whether they want to read arguments for or against the initiative and to rate how much they agree with these arguments. We find that our intervention, i.e. the increase in the self-signaling value of voting in favor of the initiative, did not affect selection of pro vs. con arguments prior to voting. However, the intervention did bias the processing of the arguments. That is, the participants who received our intervention were more inclined to accept information that supported voting for the animal-friendly initiative. Our interpretation of this last result is that treated voters, compared to untreated voters, did indeed see a greater potential gain in self-image from voting for the animal-friendly initiative after getting our message, and hence had a higher incentive to accept the pieces of information that support voting *yes*. This interpretation is consistent with a simple model of voting with motivated beliefs, in which voters can mis-encode an informative signal at some cost as shown in Appendix B (for a general analysis, see Le Yaouanq, 2021).[3] Our intervention increases the self-image value of agreeing with the animal-friendliness of the initiative, thus providing an incentive to mis-encode. Using a causal mediation analysis, we confirm statistically that biased processing of pro vs. con arguments drives the voting decision. In sum, our study provides evidence of self-signaling motives in an important, under-studied domain: voting; and it identifies a key channel for successful self-signaling: biased processing of arguments.

Self-image and social image concerns are both plausible determinants shaping voting behavior. For example, DellaVigna et al. (2016) provide evidence for social image concerns by showing that turnout is higher when potential voters are told that they will be asked to "tell others" about whether they voted. Bursztyn et al. (2022) find that people are more willing to express dissent after being given a rationale or "social cover." To study social image

[3] Thus, we contribute to the literature on how psychic benefits can lead to biased information selection and processing. There is a recent complementary literature that examines how material incentives can bias information selection and processing; see, e.g. Ambuehl (2021).

concerns, we tell some randomly chosen participants before the vote that they will discuss in an online "chat" with other participants after the vote has taken place. This variation is orthogonal to the main information intervention, meaning that whether a participant was informed that they would chat with a like-minded participant, with one who has a different opinion, or did not receive such a notice was independent of whether they got the message to increase self-image concerns ("Good-hearted people tend to be good to animals"). We expected social image concerns among those who were told the position of the future chat partner to shape which arguments were selected. However, we find no such effects, perhaps because the participants in our study were discussing with other voters independently of our study.

Our main result is thus to show that self-signaling shapes voting and that this effect operates through information processing. It is consistent with evidence of biased information processing on politically contentious, value-laden issues such as the death penalty (Lord et al., 1979; Fryer et al., 2019), abortion (Pomerantz et al., 1995), homosexuality (Munro and Ditto, 1997), war (Nyhan and Reifler, 2010), and climate change (McCright and Dunlap, 2011; Fryer et al., 2019). It suggests that a possible underlying motive for biased information processing in those political issues is self-signaling, and that this process is strong enough to impact political decisions such as voting.

We believe that the animal-welfare initiative provides an ideal setting to cleanly study self-signaling because of its strong ethical dimension, its simplicity, and its negligible material cost to voters. Our results imply more generally that self-signaling motives may play an important role in voting when the individual material stakes tied to chances to affect the outcome of the vote are small. Hence, self-signaling may potentially play an important role in large elections, and therefore in the fate of our societies regarding various ethical issues. Our results also suggest that playing the ethical card during political campaigns can effectively mobilize voters to support a specific cause. As a matter of fact, ethical claims about policies, candidates, and parties are prevalent in political campaigns and discourse (Sandel, 2005; Haidt, 2012; Enke, 2020). Political campaigners might thus be using these claims to trigger self-image concerns among voters, and gain their support. Their opponents' best response may be to engage in the same strategy, thereby escalating political polarization (Fryer et al., 2019; Garrett and Bankert, 2020).

2 Online experiment

2.1. The initiative "for the dignity of farm animals" and self-signaling

On November 25, 2018, the Swiss voted on the proposal of a grass-root initiative "for the dignity of farm animals" which was colloquially called "horncow Initiative". This proposal was framed as an initiative to improve animal welfare by incentivizing farmers to refrain from cauterizing their animals' horns. It demanded to pin down the dignity of horned animals in the Swiss constitution.[4] In addition, it asked for subsidizing farmers who do not cauterize their animals' horns, to thus limit this practice. The subsidy would be funded by cutting subsidies for less animal-friendly farming and would thus not burden taxpayers.[5] The initiative argued that cauterization is an act of violence against animals since horns are blooded organs. Both supporters and opponents of the initiative tended to agree that dehorning hurts the animals. However, they disagreed on whether the suggested policy – redistribution of subsidies toward farmers with horned cattle – would indeed improve animal welfare.[6] The initiative assumed that farmers who dehorned their cattle would stop doing so and would use the subsidies to invest in enlarging their stables instead, a necessary measure to prevent horned animals from hurting each other when they can move freely. The opponents of the initiative argued that farmers, instead of enlarging their stables, would switch from cauterizing to tethering (i.e., immobilizing) their cattle, with similar suffering inflicted on the animals as with de-horning. Whether the initiative is seen as increasing animal-welfare as framed, rather than an initiative to get subsidies, depends to a large extent on which side of the argument is right. The debate around this issue was salient and prevalent in the media in the weeks before the ballot and voters were generally well informed (Milic et al., 2019).

[4] The initiative was rejected (45.3% Yes votes, participation 48.3%). Previous Swiss initiatives to improve animal welfare concerned restrictions on animal testing (voted on 16.2.1992, 7.3.1993, 13.2.2022), on factory farming (4.6.1989, 25.9.2022), and on strengthening animal rights (7.3.2010). All have been rejected.

[5] The initiative is ideal to study self-signaling in voting not only because of its clear focus on animal welfare but also because a potential concern for animal welfare would not be superimposed by economic concerns. In fact, accepting the initiative would not have had any effects on agricultural prices, taxes or incomes for all voters except for farmers who breed (un-)horned animals. The subsidy for farmers who breed horned animals would have come from cutting subsidies for those who breed horned animals. While many ballots have some moral aspect, most also have economic consequences which may overlap or counteract the moral dimension to an unknown extent, making identification of self-signaling difficult.

[6] A representative survey (Milic et al., 2019) of the Swiss voting population commissioned by the Federal Chancellery and conducted right after the ballot found that virtually all (96%) Yes voters and 51% of No voters said that the initiative supports the dignity of farm animals (but 42% of No voters said to oppose the subsidy).

To understand how self-image concerns may influence voting in this setting, consider a voter who wants to improve animal welfare. If she votes No, i.e. against the initiative, the status quo remains untouched and she definitely does not contribute to improving animal welfare. If, by contrast, she votes Yes, i.e. in favor of the initiative, she contributes to implementing incentives to stop the practice of dehorning. This either improves animal welfare (if indeed farmers invest into enlarging their stables to hold their now horn-bearing cattle according to higher animal-welfare standards) or not (if farmers switch from dehorning to tethering). In the latter case, a Yes vote only contributes to some farmers getting subsidies, while other farmers lose an equivalent amount. How will the voter make up her mind on how to vote?

Suppose that the voter has self-image concerns: she cares for being a good person, i.e., she derives a self-image utility from believing that she has a good heart. Suppose further that every benevolent act she makes increases her belief in being a good person, as in the model of Bénabou and Tirole (2011). Since contributing to increasing animal welfare is a good act, the voter has an incentive to believe that the initiative, if successful, would indeed improve the situation of the animals, and to vote in favor of it. On the other hand, she does not want to merely help a group of farmers to get subsidies if this has no beneficial consequences for anyone else, least the animals. Imagine for a moment that her aversion against merely shifting subsidies outweighs the incentive to improve her self-image, so that she leans toward voting No. Suppose now that one day before the ballot, she learns a new piece of information, namely that good deeds toward animals are more indicative of the person's intrinsic goodness of character than she thought: she learns that according to science, there is a positive correlation between being good to animals and being good to other humans. Thus, a good act toward animals is a stronger signal than she thought before of being an overall good-hearted person. Now the incentive to believe in the initiative's framing and to vote Yes becomes stronger because after what she learned she can infer from her Yes vote a higher probability of being an overall good-hearted person. If this increased incentive outweighs her aversion against supporting a narrowly focused cause, she tilts toward voting Yes.

We hypothesize that a significant number of voters are like her in that their self-image concerns matter for what they want to believe about an ethically framed policy proposal, and whether they want to vote for it. To test this hypothesis, and to identify the channel of

belief formation, we conducted a two-wave survey experiment timed before and after the ballot. See also Appendix B for a simple theoretical framework and predictions (Le Yaouanq, 2021).[7]

2.2. Experimental design

Our online experiment has two waves. In the first wave, we elicit prior voting intentions and implement three self-image treatments with different informational interventions. In the main treatment, HIGH, we provide the piece of information mentioned above, i.e., we (truthfully) inform subjects about the correlation between being good to animals and being good to other humans, suggesting that being good to animals signals being an overall good-hearted person. Thus, treatment HIGH increases the self-signaling value of believing in the initiative's framing and voting for it, as discussed above.

In the control treatment (NEUTRAL) we do not give any such information. In treatment LOW, we attempt to decrease the self-signaling value of voting Yes, i.e., to reduce the incentive to believe in the initiative's framing and to vote for it. We do so for two reasons. The first is that we could not know how successful the initiative's framing initially was. If almost all voters fully believed in the animal-friendliness of the initiative before our intervention, and tended toward voting Yes, our intervention in HIGH would not have any significant effect on voting even if self-image concerns played a role. In such a situation, however, decreasing the self-signaling value of a Yes vote might well be effective in the presence of self-image concerns and make voters switch toward voting No. Second, we were ethically motivated to limit the overall effect of our experiment on voting just in case our sample turned out pivotal. To this end, it would of course have been most effective if LOW were the exact opposite to HIGH, i.e., if the treatment informed subjects that being good to animals tends to be negatively correlated with being good to humans. However, we could not find any scientific study that claims such a relation. Hence, the best we could do while not deceiving subjects was to tell them in LOW that people can be good to animals and indifferent toward other humans ("People who care about animals are not necessarily kind-hearted").

[7] Note we assume that the treatment HIGH affects self-image but we also (implicitly) assume that it does not make participants more altruistic towards animals in the sense of deriving direct utility from the animals' well-being. We believe, however, that this is not a strong assumption. Indeed, since participating in the experiment makes animal welfare more salient in all treatments, and the emphasis on animal welfare is comparable across treatments, we do not see any reason to suspect that the specific treatment which boosts self-image (HIGH) would generate more altruism towards animals than the other treatments.

To test how self-image concerns interact with social image concerns to shape belief formation and voting choices, we announced to randomly selected subjects that they would be matched with another subject in Wave 2 to chat about how they voted. One third of the subjects was told to be matched with a like-minded subject (BUBBLE), one third with an opposing subject (CONFRONT), and the rest got no such announcement (NOCHAT) but was given an unanticipated opportunity to chat in Wave 2. The social image treatment variation is orthogonal to the information intervention, i.e., whether there was an announcement of the chat and the chat partners' position on the initiative was independent of which self-image treatment the subject got in Wave 1.

In the second wave, we re-contacted all subjects who completed the first wave. We elicited how they actually voted and then matched them for the chat as announced in the first wave. The first wave was implemented in the two weeks prior to the ballot, and the second wave a few days after. Participants were recruited by the standing panel of the LINK Institute and provided written consent.[8] Only truthful information was given to them. Subjects were informed as part of their consent that the survey in wave one might vary across participants. We screened out voters who had voted already voted (by mail) before the start date of wave one, participants not eligible to vote and those not complying with LINK Institute guidelines for chat interactions.[9]

In total, we conducted nine randomized versions of one survey (three self-image times three social image treatments). All versions elicited relevant demographics, the attitude toward the initiative before exposure to any treatment (*PriorAttitude*), and how well-informed the participant was about the initiative (*Informed*).[10] The outcome variables are discussed below and concern the selection of arguments, the processing of arguments, and the intended and reported vote. See Table A.1 in the Appendix for a full list of variables.

---

[8] Switzerland has four national languages of which German is the most common (about 70% of the voting population). Subjects were recruited only from the German-speaking part of Switzerland to avoid difficult issues in (back-)translation of instructions. Differences in voting patterns across linguistic regions of Switzerland are not uncommon but not the focus of this study.

[9] The screening according to the guidelines of LINK Institute was pre-registered. For more details see section 3. LINK Institute conducts ex-post surveys of ballots on behalf of the Swiss government on a regular basis. Participants in the standing panel repeatedly participate in such surveys (e.g., Milic 2018). On the day of voting on the horncow initiative, two other issues were put before voters (one relating to international treaties, one relating to social insurance).

[10] We checked the validity of self-reported informedness in a quiz. The quiz also provides us with a control variable measuring overconfidence (when the subject's reported informedness is above the median but their quiz performance is below).

Table 1: Arguments provided to subjects (translated from German)

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, 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, some 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. However, the booklet sent to all Swiss voters mentioned a very similar argument. All other arguments are direct translations from the booklet.

Selection of arguments. We provide subjects with a list of three arguments to vote for the initiative (PRO) and a list of three arguments to vote against (CON), and let subjects choose which list(s) they wanted to read. The arguments were taken from a booklet that the Swiss government sent to all Swiss voters several weeks before our experiment started with the exception of one argument that was widely circulated in the media. We added this argument to create a balanced information menu. The three PRO arguments claimed that dignity and physical well-being of animals as well as justice among farmers would improve, should the initiative be approved. The three CON arguments addressed these same three goals and argued that none of them would be reached in case of the proposal's success (see Table 1). Subjects had to choose which list(s) of arguments to read: all, only PRO, only CON, or none. When making the choice, they did not know that the lists only contained arguments they

already were very likely to know from the official booklet or the media.[11] This procedure allows us to test whether those treated in HIGH were more likely to avoid CON arguments without altering the majority's information set.[12]

Processing of arguments. We also test whether self-image concerns operate through processing of arguments, i.e., how HIGH and LOW affect the extent to which our subjects agree with each set of arguments (from 'not at all' to 'fully'). In HIGH, we expect to find stronger agreement with PRO arguments than in LOW and NEUTRAL based on theoretical considerations (e.g., Bénabou and Tirole, 2011), and experimental evidence (Eil and Rao, 2011; Sharot et al., 2011; Sharot and Garrett, 2016; Kuzmanovic et al., 2018).

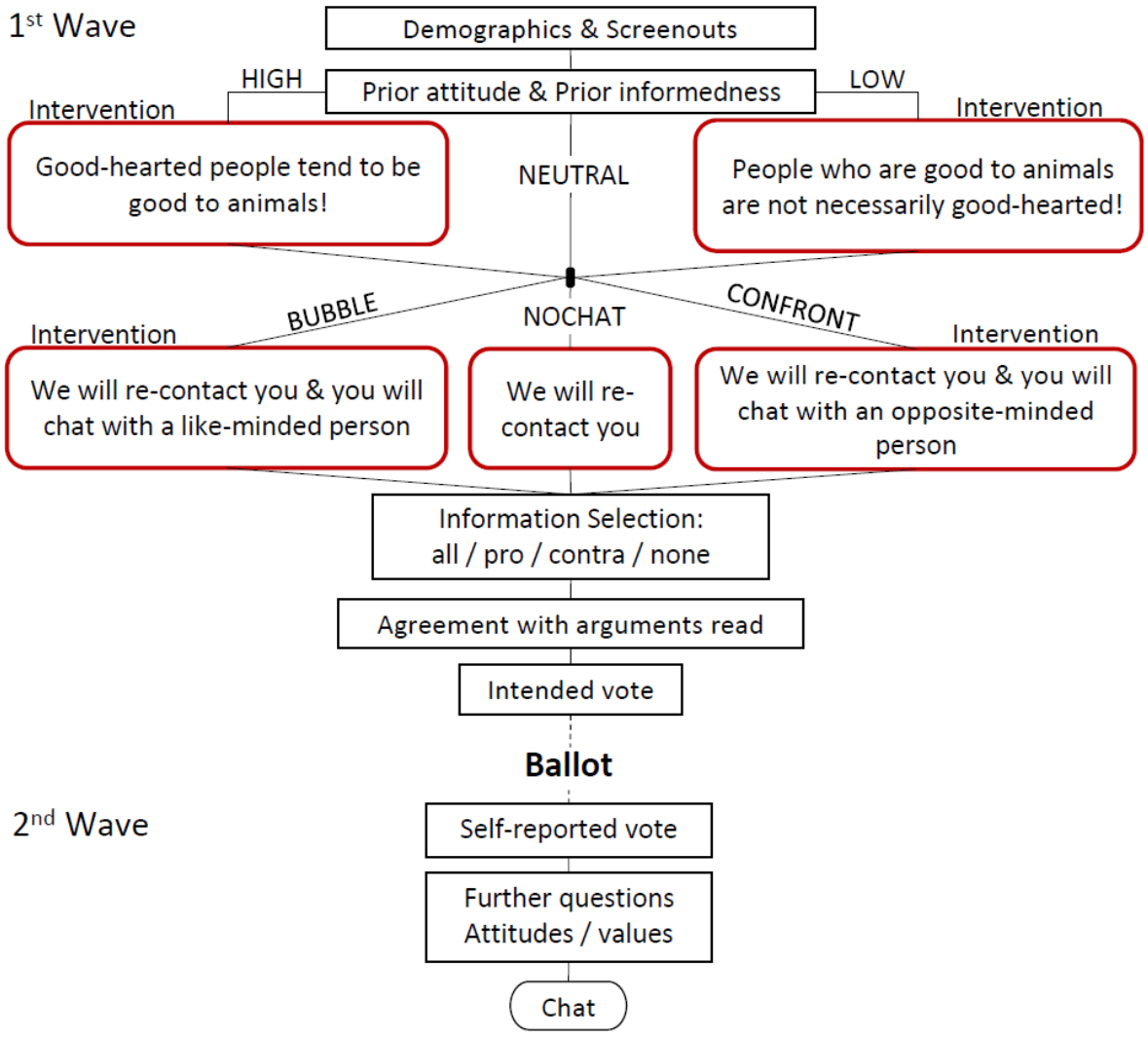
Intended and actual votes. Voting plans tend to operate as commitment devices (Nickerson and Rogers, 2010). We hence elicit the immediate effect of our informational treatments by asking our subjects whether they intend to turn out and, if they do, how they intend to vote at the end of wave 1. In combination with the variable *PriorAttitude*, this allows us to measure changes in how subjects evaluate the proposal after being treated. Actual votes were elicited in wave 2 by asking completers of wave 1 whether, and how, they voted.

Figure 1 summarizes the experimental design. It is important to note that our design measures agreement with the arguments *conditional* on having chosen to read them. This design choice means that we can test the effects on information processing (i.e., on how much our subjects agree with the arguments) only on those subjects who read them. We made this choice for the following reason. Based on the previous literature that had found effects of self-signaling on information selection in other contexts (Matthey and Regner, 2011; Nyborg, 2011; Feiler, 2014; Grossman, 2014; Serra-Garcia and Szech, 2021; Freddi, 2019), we prioritized studying information selection as an outcome variable (i.e., which arguments subjects choose to read vs. avoid).

Figure 1: Study Design

[11] According to ex-post survey (Milic et al., 2019), 88% of respondents had read the booklet.

[12] For an overview about the literature on information selection, in particular information avoidance, see Golman et al. (2017).



2.3. Hypotheses

Our main hypothesis is

Hypothesis H1 (Self-image concerns shape voting)

(a) HIGH increases voting for the initiative (relative to NEUTRAL),

(b) LOW reduces voting for the initiative (relative to NEUTRAL).

Hypotheses H2 and H3 concern the channels through which the informational intervention affects outcomes:

Hypothesis H2 (Self-image concerns operate through selection of arguments)

(a) HIGH increases avoidance of CON arguments (relative to NEUTRAL),

(b) LOW decreases avoidance of CON arguments (relative to NEUTRAL).

Hypothesis H3 (Self-Image concerns operate through processing of arguments)

(a) HIGH increases agreement with PRO arguments, and thereby voting Yes (relative to NEUTRAL).

(b) LOW decreases the agreement with PRO arguments, and thereby voting Yes (relative to NEUTRAL).

Hypothesis H4 (Social image concerns operate through selection of arguments) [13]

(a) BUBBLE increases avoidance of arguments opposing the participant's own prior attitude (relative to NOCHAT).

(b) CONFRONT decreases avoidance of arguments opposing the participant's own prior attitude (relative to NOCHAT).

3. Results

3.1. Data and descriptive statistics

The first wave of the survey was completed by 2,112 participants recruited from the standing LINK Institute panel that is representative for the Swiss adult voting population. Of those, 1,756 answered the questions for outcome variables *PriorAttitude* (pre-treatment) and either intended voting (*IntVote*) in wave 1 or reported voting (*RepVote*) in wave 2 (both post-treatment). Six participants were dropped because control variables (*Female* and *Farmer*) were missing. A further 214 participants with a very high emotional involvement with the initiative were not re-invited to the second wave in compliance with guidelines of the LINK Institute to avoid exposing subjects to stressful situations during the chat.

Table 2 reports descriptive statistics for the resulting sample of 1, 536 subjects (Wave 1) and the proper subsample of 1,032 subjects who also completed wave 2. The results reported below are based on these samples or proper subsamples (see Table A.6 for robustness checks when including the highly emotional subjects). Attrition was independent of assignment to treatments (see Appendix Tables A.3 and A.4 for details). The share of participants who supported the initiative in the final sample is close to the nationwide ballot (see Table A.5 for details).

[13] We thank a referee for suggesting this hypothesis.

Table 2: Summary statistics

		Wave 1		Wave 2	
Type	**Variable**	**Obs.**	**Mean** (std.dev.)	**Obs.**	**Mean** (std.dev.)
Treatment	HIGH	1,536	.342	1,032	.340
	LOW	1,536	.308	1,032	.303
	BUBBLE	1,536	.237	1,032	.242
	CONFRONT	1,536	.238	1,032	.236
Control	*Age 18-34*	1,536	.227	1,032	.232
	Age 35-64	1,536	.596	1,032	.583
	Age > 64	1,536	.177	1,032	.185
	Female	1,536	.557	1,032	.538
	Farmer	1,536	.012	1,032	.012
	FarmHorn	1,536	.006	1,032	.005
	Informed	1,536	.293 (1.63)	1,032	.369 (1.62)
	Emotions	1,536	2.91 (1.65)	1,008	2.91 (1.69)
Outcome	*PriorAttitude*	1,536	-.068 (.657)	1,032	-.086 (.657)
	ReadPROonly	1,536	.038	1,032	.041
	ReadCONonly	1,536	.019	1,032	.020
	ReadBoth	1,536	.786	1,032	.803
	AvoidCON	1,536	.195	1,032	.177
	ReadOpp	1,536	.802	1,032	.821
	AgreePRO	1,250	.220 (.502)	862	.213 (.505)
	IntVote	1,521	-.052 (.675)	1,017	-.061 (.683)
	RepVote	-	-	768	.381

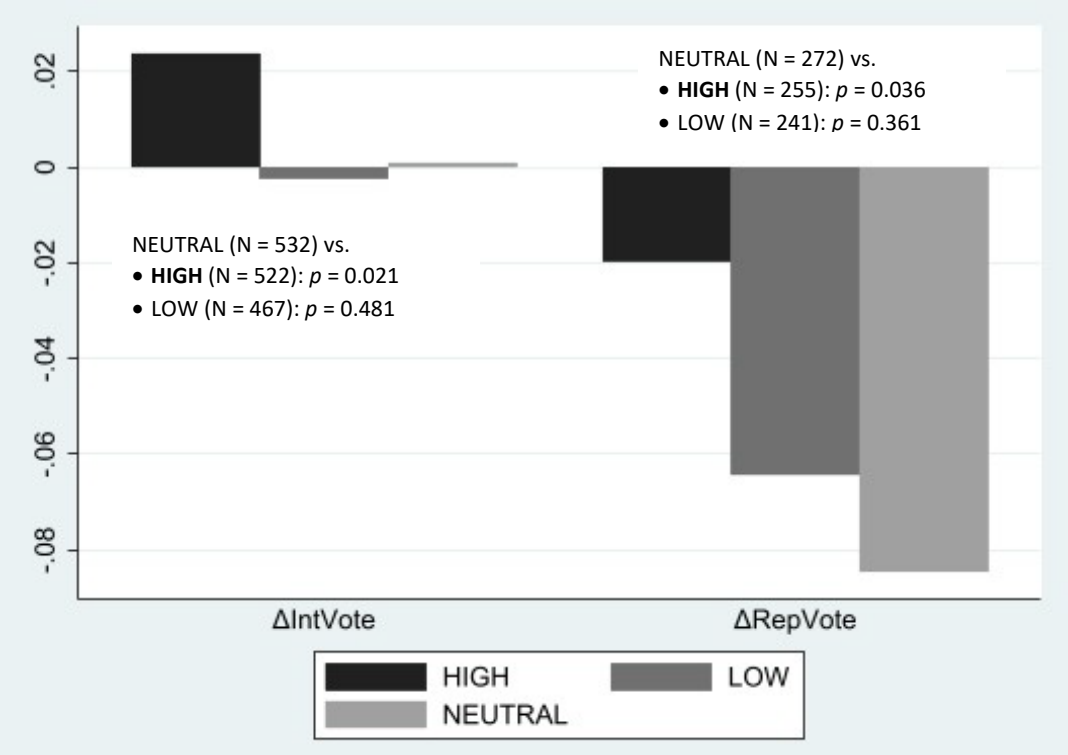
Notes: Means for the treatment variables indicate the share of subjects in the respective treatment, with NEUTRAL and NOCHAT as left-out categories. All variables except *RepVote* were elicited in wave 1. All treatment and *Age* variables as well as *Female, Farmer, FarmHorn*, the three *Read* and *RepVote* variables are dummies. No standard variation listed for dummies. *Informed* ranges from 'not at all' (-3) to 'very well' (3). *Emotions* codes expressed emotional response to the initiative from 'not at all' (0) to 'considerably' (5). *PriorAttitude* ranges from 'certainly against' (-1) to 'certainly in favor' (1). *ReadPROonly* and *ReadCONonly* identify respondents who only read PRO and CON arguments, respectively. *ReadBoth* identifies those who read both arguments. *AgreePRO* indicates how convincing a participant finds the PRO arguments conditional on having read them ranging from 'totally not convincing' (-1) to 'totally convincing' (1) in steps of 0.5. *IntVote* is the voting intention stated at the end of wave 1 ranging from 'certainly No' (-1) to 'certainly Yes' (1) in steps of .5.

Table 2 shows that participants were relatively well informed about and emotionally involved with the initiative. The share of subjects who reported to have voted for the initiative was below the respective outcome of the ballot in the German-speaking Cantons (38.1% vs. 43.8%) but reported voting is well in line with prior attitudes (37.4% in favor) and voting intentions (36.7% in favor, see Appendix Table A.5 for details).

3.2. Self-signaling shapes voting

Figure 2 shows that our intervention to increase the self-signaling value of voting Yes (in HIGH) increased both voting intentions and reported voting, compared to NEUTRAL. In contrast, the (weaker) intervention LOW had no significant effect. Figure 2 reports the normalized effects on these key outcome variables relative to *PriorAttitude* (see Table A.1 for definitions): *ΔIntVote* is the deviation of the intended vote, and *ΔRepVote* the deviation of the reported vote from a participant's pre-treatment prior attitude. The immediate effect of our HIGH intervention is significant (p = 0.02, MW-test), and the longer-run effect on voting is quantitatively even stronger (recall from Figure 1 that *IntVote* is elicited immediately after the interventions, *RepVote* is elicited about 2 weeks after the intervention). Bars below zero for *ΔRepVote* suggest that absent our informational intervention, participants tended to grow more critical towards the initiative over time but the HIGH intervention is significant according to Mann-Whitney tests (p = 0.036) despite this general trend.

Figure 2: Effect of self-signaling interventions on voting



Notes: Figure shows effect of treatment on intended / reported voting relative to *PriorAttitude* (-1 is 'certainly against' to 1 'certainly in favor'). Δ*IntVote* is the normalized [-1,1] difference between intended voting reported at the end of wave 1 and *PriorAttitude*: Δ*IntVote* = [(*IntVoting* - 3) / 2 - *PriorAttitude*] / 2. Δ*RepVote* is the normalized [-1,1] difference between self-reported actual voting reported in wave 2 and *PriorAttitude*: Δ*RepVote* = *RepVoting* - (*PriorAttitude* - 1) / 6. Figure 2 is based on samples who reported the respective outcome variable (Δ*IntVote*: 1,521; Δ*RepVote*: 768, respectively, see Table 2). *p*-values for Mann-Whitney tests against NEUTRAL.

The effects of HIGH on self-reported voting are remarkably strong. For example, the share of Yes votes is 40.0% in HIGH vs. 36.4% in NEUTRAL, corresponding to an increase in Yes votes of almost 10% relative to NEUTRAL. This pronounced effect is surprising given that many voters already had firm voting intentions before our intervention. In fact, 58% of respondents in an ex-post survey (Milic et al., 2019) say "it was clear from the beginning how I would vote". Accordingly, we find that prior intentions are strong predictors of voting (see also Table 3).

Table 3 confirms the results from the non-parametric tests shown in Figure 2 in regression analysis: Columns (2) and (4) show that HIGH is significant relative to NEUTRAL which is the left-out category in all specifications, while LOW is not. In addition, in (1) and (3) we also report regressions on the non-differenced variables *IntVote* and *RepVote*, i.e., voting variables that are not defined relative to *PriorAttitude*, but include *PriorAttitude* as a regressor. Doing so makes no difference for the conclusions. Column (3) shows that exposure to the HIGH treatment increased the reported voting for the initiative on average by 6.8 percentage points. Overall, we find strong support for H1a but no support for H1b.

Table 3: Average treatment effects on voting (with controls)

	(1) ***IntVote***	(2) ***ΔIntVote***	(3) ***RepVote***	(4) ***ΔRepVote***
HIGH	0.049	0.026	0.068	0.069
	(0.031)	(0.029)	(0.016)	(0.019)
LOW	-0.001	-0.001	0.017	0.025
	(0.980)	(0.919)	(0.555)	(0.397)
BUBBLE	0.030	0.018	0.029	0.023
	(0.205)	(0.133)	(0.301)	(0.441)
CONFRONT	-0.008	-0.006	0.014	0.020
	(0.717)	(0.611)	(0.629)	(0.502)
PriorAttitude	0.854		0.383	
	(0.000)		(0.000)	
Controls	*Yes*	*Yes*	*Yes*	*Yes*
N	1,521	1,521	768	768
R^2 / Pseudo R^2	0.706	0.016	0.497	0.033
F / LR Chi^2	328.7	2.4	507.8	2.6

Notes: Column (4) shows marginal effects of probit a regression, all other columns show coefficients from OLS regressions. The specifications shown here include controls (regressions without controls lead to the same conclusion, see Table A.6). *p*-values in parentheses. Estimates are based on samples in Table 2. Running regressions (2) and (4) with samples of those having read a balanced set of arguments, i.e. either both types or none, to exclude potential effects driven by biased information selection, yields very similar results: HIGH in (2) becomes 0.028 (0.021) and in (4) 0.069 (0.019)).

The last row shows that *PriorAttitude* is a very strong predictor of both intended and actual voting. The fact that the coefficient of *PriorAttitude* is weaker for actual than for intended voting indicates that opinions became generally less favorable for the initiative in the days immediately preceding the ballot (as is also indicated by the negative values for Δ*RepVote* in Figure 2).

Rows 3 and 4 show that CONFRONT and BUBBLE had no effect on voting overall. Although this is in line with our expectations, the near-absence of effects of social signaling in our experiment should not be taken as indicating that social image concerns are irrelevant in general.[14] Participants may have discussed the initiative with their friends and neighbors, and such discussions before the ballot may well have affected voting through the channel of (verbal) social signaling and commitment to what one communicated.

3.3. Information selection

We find no robust effects of our treatment variations on information selection, i.e., which arguments participants choose to read. The main effects of HIGH on voting documented in Figure 2 and Table 3 therefore do not operate through this channel. This absence of treatment effects on information selection is not surprising ex post since we find that very few (5.7%, see Table 2) subjects engage in one-sided information selection. If subjects avoid information, they avoid reading both types of arguments (15.7% do, see Table 2). The vast majority (79%) chose to read both sets of arguments.

Table 4: Information selection

Outcome Variable	**Treatments**		**Mann-Whitney** (*p*-values)
AvoidCON	HIGH (*N* = 526) LOW (*N* = 473)	vs. NEUTRAL (*N* = 537)	0.634 0.557
ReadOpp	BUBBLE (*N* = 364) CONFRONT (*N* = 366)	vs. NOCHAT (*N* = 806)	0.721 0.800

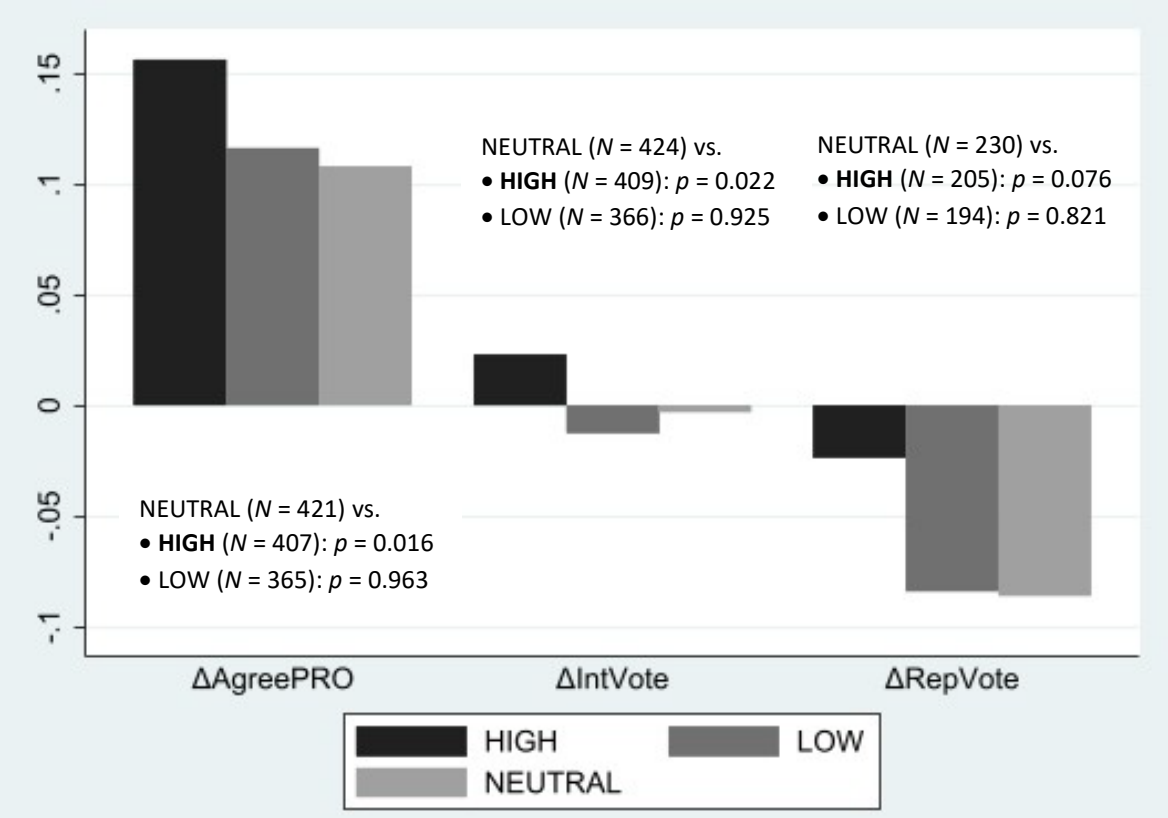
Note: *AvoidCON* is a dummy that equals 1 if the participant does not read CON arguments, *ReadOpp* is a dummy that equals 1 if the participant reads the arguments opposing his/her own prior attitude. Based on participants who completed wave 1 of the survey (*N* = 1,536). Of these, 78.6% read both PRO and CON arguments, 15.7% read neither, 3.8% read only PRO, and 1.9% only CON arguments, respectively.

[14] We expected social image concerns to affect argument selection (see H4). However, in our specific setting, argument selection should not be expected to affect voting since the arguments which our subjects could choose to read were already widely known.

Table 4 presents non-parametric tests showing that neither self-image nor social image concerns had an effect on information selection. The first two rows show that HIGH had no effect on avoiding CON arguments (*AvoidCON*), when tested against NEUTRAL (the same holds for testing against NEUTRAL and LOW jointly). The next two rows show that announcing the chat had no significant effects on reading arguments counter to one's own initial position, independent of who they would chat with. Probit regressions in Table A.8 confirm these conclusions. Hence, Hypotheses H2 and H4 concerning information selection are not supported. We also note that the results on voting (main results on H1) are robust to restricting the sample to those reading either all or none of the arguments (see note to Table 3), confirming that the treatment effect of HIGH is not due to biased information selection.

3.4. Information processing

Figure 3: Effect of self-signaling interventions on agreement with PRO arguments



Notes: Based on participants who read both PRO and CON arguments (1,208 out of 1,536), indicated *AgreePro* (1,193), *IntVote* (1,199) or *RepVote* (629), respectively. *ΔAgreePRO* is the normalized [-1,1] difference between the self-reported agreement with arguments in favor of the initiative at the end of wave 1 and *PriorAttitude*: *ΔAgreePRO* = [*AgreePRO* - *PriorAttitude*]/2. Positive values indicate increased agreement relative to the attitude expressed before reading PRO and CONTRA arguments. *AgreePRO* indicates how convincing PRO arguments are from -1 ('not at all convincing') to 1 ('fully convincing'). *p*-values are for Mann-Whitney tests. Results are robust to comparing HIGH against NEUTRAL & LOW which reduces *p*-values to 0.008, 0.011, and 0.059, respectively.

Figure 3 shows that the intervention to increase the self-signaling value of a Yes vote (HIGH) had a strong effect on agreement with PRO arguments relative to the pre-treatment attitude toward the initiative for those who have read PRO & CON arguments (see leftmost bars). In contrast, LOW had no significant effect on agreement with PRO arguments. This result suggests that the main effect of HIGH on voting demonstrated in Figure 2 and Table 3 operated through information processing, as stated in hypothesis H3a. For purposes of comparison to Figure 2, we also report the effects of the self-signaling intervention on intended (Δ*IntVote*) and reported voting (Δ*RepVote*) in the reduced sample of those who have read PRO & CON arguments (80.3% did, see Table 2). Again, we find that HIGH has significant effects.

3.5. Mediation analysis and robustness tests

Table 5 reports results from causal mediation analysis (Imai et al., 2010) and serves to estimate the effect of biased information processing for the treatment's impact on voting.[15] The total effect and the average causal mediation effect (ACME) of HIGH on both measures of voting are significant in all specifications. The direct effect is not significant in any of them. In the absence of a direct effect, the ACME is equal to the total effect. The estimated percentage of the total effect mediated through information processing ranges from 15 to 63 percent. The percentages tend to be lower when controls are included and for reported vs. intended voting. With controls, the effect of self-signaling is in the range of 16 to 32 percent. These percentages are broadly in line with the shares found in other studies assessing transmission channels on voting and public opinion (Tomz and Weeks, 2013). In summary, Table 5 shows that hypothesis H3a is supported, i.e. that the effect of HIGH on voting can be causally attributed to the effect of our intervention on agreement with PRO arguments.

[15] Table A.9 presents results from an IV regression where we instrument *(Δ)AgreePRO* with treatments *HIGH* and *LOW*. They confirm that the treatments affect voting via information processing.

Table 5: Mediation analysis: Effect sizes of *HIGH* on voting via information processing

Mediator	***AgreePRO***				***ΔAgreePRO***			
Outcome	*IntVote*		*RepVote*		*ΔIntVote*		*ΔRepVote*	
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Total	0.054*	0.056*	0.065*	0.071*	0.028*	0.029*	0.065*	0.071*
ACME	0.034***	0.031***	0.025***	0.023***	0.016**	0.015*	0.014**	0.011*
Direct	0.020	0.024	0.040	0.048	0.012	0.014	0.051	0.060
% mediated	63.4	55.4	38.5	32.2	56.1	52.0	20.0	15.7

Notes: Causal mediation analysis based on Imai et al. (2010) and samples described in note to Figure 3. 'Total' is the total effect of treatment on the outcome variable in an OLS regression which is then disaggregated into the 'Direct' effect of treatment on outcome and the Average Causal Mediation Effect (ACME). Outcome variables are two measures of voting. Voting intentions prior to the ballot (*IntVote*) and reported vote after the ballot (*RepVote*). The mediating variable is agreement with pro arguments (*AgreePRO*) given both pro and con arguments have been read. Variables starting with *Δ* are changes relative to the pre-treatment position toward the initiative (*PriorAttitude*). Since the *medeff* command in Stata does not compute precise *p*-values, we report ranges: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 6 tests whether the impact of HIGH on voting and its transmission via biased information processing remains significant if we take into account that we are testing multiple hypotheses simultaneously. To do so, we use the *mhtreg* package for Stata based on List et al. (2019) and explained in detail in Barsbai et al. (2020). The first row considers only the two main hypotheses and tests them simultaneously using Δ*IntVote* (left part of the table) and Δ*RepVote*. We find that the effect of HIGH on Δ*AgreePro* (Column 3) and of Δ*AgreePro* on Δ*IntVote* (Column 2) remains significant when considering multiple hypothesis testing (i.e., H3a holds when simultaneously testing for H1a). This also holds for the direct effect of HIGH on Δ*IntVote* (i.e., H1a holds when simultaneously testing for H3a, see Col. 1). The same findings prevail for reported voting Δ*RepVote* (see right half of Table 6).

The second row of Table 6 simultaneously tests for all eight hypotheses presented in section 2. The effect of treatment on voting (H1a) now ceases to be significant but the effect mediated via biased information processing on voting (H3a) remains (weakly) significant. The third row additionally includes seven exploratory hypotheses that were included in the pre-registration (see Appendix C) but that are not based on randomized treatments and hence are mere correlations. In summary, correcting for multiple hypotheses testing confirms that the treatment effect on voting via the biased information processing channel is robust.

Table 6: Multiple Hypothesis Testing

Hypotheses #	Labels	Depen.: / Indepen.:	(1) ΔIntVote HIGH	(2) ΔIntVote ΔAgreePRO	(3) ΔAgreePRO HIGH	(4) ΔRepVote HIGH	(5) ΔRepVote ΔAgreePRO	(6) ΔAgreePRO HIGH
3	H1a, H3a		.027	.000	.007	.035	.001	.013
9	H1, H2, H3, H4		.186	.000	.046	.197	.002	.058
16	H1, H2, H3, H4 + 7 exploratory		.278	.000	.069	.292	.006	.102

Notes: Adjusted *p*-values based on Theorem 3.1 in List et al. (2019) using implementation by Barsbai et al. (2020) which asymptotically controls familywise error rates and is asymptotically balanced. Number of bootstrap simulation samples: 5.000. All regressions based on OLS with controls (*Female, Informed, Farmer, FarmHorn*, age categories). Results from exploratory analysis are presented in Appendix C. Based on samples in Table 2. The first column shows the number of tests performed, column 2 names the hypotheses tested.

Next, we consider the possibility of an experimenter demand effect. We believe it to be unlikely that our findings are confounded by such an effect for three reasons. First, to create an experimenter demand effect, participants would need to update beliefs about which answers to specific survey questions the experimenters prefer. However, none of the interventions made any reference to the initiative, horned animals, or the acquisition or processing of information. HIGH and LOW directly target the self-signaling value of being good to animals but do so by reporting descriptive scientific evidence rather than making normative statements. The entire survey was framed in a neutral way. Second, De Quidt et al. (2018) show that experimenter demand effects are generally modest even if the experimental hypothesis of the impact of the treatment is revealed to participants. Third, the interventions used in HIGH and LOW are similar in intent but in opposite directions. However, our findings show that HIGH has a substantial impact on information processing and voting but LOW does not.

Our findings on voting behavior concern intended and reported votes. Hence, the question arises whether consistency bias in survey responses could have influenced our results.[16] Consistency bias could motivate participants to cast the same vote as the intended vote or at least report having cast the same vote in the experiment. However, we would not expect the prevalence of consistency bias to differ across treatments, thus the differences we find across treatments are not attributable to this source.

[16] We thank an anonymous referee for pointing this out to us.

4. Conclusion

This paper presents experimental evidence that self-image motives affect voting in a controversial ballot in Switzerland. This ballot, which concerned the "dignity of horned animals", is ideal to study self-signaling in voting because it has a clear ethical dimension, was easy to understand, and had no material consequences for almost all voters (except for redistribution of subsidies between farmers). Thus, our experimental design ensured that the ethical dimension of the ballot did not overlap with narrow material economic concerns, such as increased taxes or meat prices, making it possible to isolate self-signaling effects.

We send a message to voters about scientific evidence supporting the claim that "good-hearted people tend to be good to animals" to increase the self-signaling value of voting for the initiative. We show that this message results in strong effects. In particular, the share of Yes votes increases by about 10 percent.

We study the mechanism by which the message affected self-signaling and, ultimately, voting. We do not find evidence supporting the claim that participants avoid reading arguments against the initiative (emphasizing subsidies), or choose to read only arguments in favor (emphasizing animal welfare). Instead, we do find support for the claim that self-signaling operates through information processing. We find that voters treated with the message that "good-hearted people tend to be good to animals" agree more with arguments emphasizing the positive effects of the initiative on animal welfare. Mediation analysis shows that about thirty percent (between 16 and 39 percent, depending on the specification) of the total effect on reported voting operates through information processing. Our results are in line with and provide support for the formal paradigm developed by Bénabou and Tirole (2002, 2011) and the subsequent literature building on this work.

While we demonstrate the effects of self-signaling using an actual initiative in a particular ethical context (animal welfare), we believe that studying the effects of self-signaling as a potentially important driver of voting behavior is worthwhile in many other politically contentious, value-laden issues such as the death penalty, abortion, and preventing climate change.

References

Adena, M., & Huck, S. (2020). Online fundraising, self-image, and the long-term impact of ask avoidance. Management Science, 66(2), 722-743.

Ambuehl, S. (2021). Can Incentives cause harm? Tests of undue inducement. Working paper.

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

Bénabou, R., & Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. Quarterly Journal of Economics, 126(2), 805-855.

Barsbai, T., Licuanan, V., Steinmayr, A., Tiongson, E., & Yang, D. (2020). Information and the acquisition of social network connections. NBER working paper 27346.

Bursztyn, L., Egorov, G., Haaland, I. K., Rao, A., and C. Roth (2022). Justifying dissent. NBER working paper 29730.

Dana, J., Weber, R. A., & Kuang, J. X. (2007). Exploiting moral wiggle room: experiments demonstrating an illusory preference for fairness. Economic Theory, 33(1), 67-80.

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

De Quidt, J., Haushofer, J. and Roth, C. (2018). Measuring and bounding experimenter demand. American Economic Review, 108(11), 3266-3302.

Di Tella, R., Perez-Truglia, R., Babino, A., & Sigman, M. (2015). Conveniently upset: avoiding altruism by distorting beliefs about others' altruism. American Economic Review 105 (11), 3416-3442.

Eil, D., & Rao, J. M. (2011). The good news-bad news effect: asymmetric processing of objective information about yourself. American Economic Journal: Microeconomics, 3(2), 114-38.

Enke, B. (2020). Moral values and voting. Journal of Political Economy, 128(10), 3679-3729.

Feiler, L. (2014). Testing models of information avoidance with binary choice dictator games. Journal of Economic Psychology, 45, 253-267.

Fiorin, S. (2022). Reporting peers' wrongdoing: Experimental evidence on the effect of financial incentives on morally controversial behavior, mimeo.

Freddi, E. (2019). Do people avoid morally relevant information? Evidence from the refugee crisis. Review of Economics and Statistics, 1-45.

Fryer Jr, R. G., Harms, P., & Jackson, M. O. (2019). Updating beliefs when evidence is open to interpretation: Implications for bias and polarization. Journal of the European Economic Association, 17(5), 1470-1501.

Gao, L., Wheeler, S. C., & Shiv, B. (2009). The "shaken self": Product choices as a means of restoring self-view confidence. Journal of Consumer Research, 36(1), 29-38.

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), 621-640.

Ghanem, D., Hirshleifer, S., & Ortiz-Becerra, K. (2022). Testing Attrition Bias in Field Experiments. CEGA Working Paper No. WPS-113. University of California, Berkeley.

Golman, R., Hagmann, D., & Loewenstein, G. (2017). Information avoidance. Journal of Economic Literature, 55(1), 96-135.

Grossman, Z. (2014). Strategic ignorance and the robustness of social preferences. Management Science, 60(11), 2659-2665.

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.

Haaland, I., Roth, C., & Wohlfahrt, J. (2023). Designing information provision experiments. Journal of Economic Literature, 61(1), 3-40.

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

Heatherton, T. F., & Polivy, J. (1991). Development and validation of a scale for measuring state self-esteem. Journal of Personality and Social Psychology, 60(6), 895-910.

Imai, K., Keele, L., & Tingley, D. (2010). A general approach to causal mediation analysis. Psychological Methods, 15(4), 309.

Köszegi, B. (2006). Ego utility, overconfidence, and task choice, Journal of the European Economic Association, 4(1), 673-707.

Kuzmanovic, B., Rigoux, L., & Tittgemeyer, M. (2018). Influence of vmPFC on dmPFC predicts valence-guided belief formation. Journal of Neuroscience, 38(37), 7996-8010.

Le Yaouanq, Y. (2021). Motivated cognition on a model of voting. Mimeo.

List, J. A., Shaikh, A. M., & Xu, Y. (2019): Multiple hypothesis testing in experimental economics. Experimental Economics 22(4), 773-793.

Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of Personality and Social Psychology, 37(11), 2098.

Matthey, A., & Regner, T. (2011). Do I really want to know? A cognitive dissonance-based explanation of other-regarding behavior. Games, 2(1), 114-135.

McCright, A. M., & Dunlap, R. E. (2011). The politicization of climate change and polarization in the American public's views of global warming, 2001–2010. Sociological Quarterly, 52(2), 155-194.

Milic, T., Feller, A., Kübler, D. (2019) VOTO-Studie zur eidgenössischen Volksabstimmung vom 25. November 2018, URL: https://doi.org/10.5167/uzh-162717

Monin, B., & Miller, D. T. (2001). Moral credentials and the expression of prejudice. Journal of Personality and Social Psychology, 81(1), 33.

Munro, G. D., & Ditto, P. H. (1997). Biased assimilation, attitude polarization, and affect in reactions to stereotype-relevant scientific information. Personality and Social Psychology Bulletin, 23(6), 636-653.

Nickerson, D. W., & Rogers, T. (2010). Do you have a voting plan? Implementation intentions, voter turnout, and organic plan making. Psychological Science, 21(2), 194-199.

Nyborg, K. (2011). I don't want to hear about it: Rational ignorance among duty-oriented consumers. Journal of Economic Behavior & Organization, 79(3), 263-274.

Nyhan, B., & Reifler, J. (2010). When corrections fail: The persistence of political misperceptions. Political Behavior, 32(2), 303-330.

Pomerantz, E. M., Chaiken, S., & Tordesillas, R. S. (1995). Attitude strength and resistance processes. Journal of Personality and Social Psychology, 69(3), 408-419.

Sandel, M. J. (2005). Public philosophy: Essays on morality in politics. Harvard University Press.

Serra-Garcia, M., & Szech, N. (2021). The (in) elasticity of moral ignorance. Management Science 68(7), 4815-4834.

Sharot, T., & Garrett, N. (2016). Forming beliefs: Why valence matters. Trends in Cognitive Sciences, 20(1), 25-33.

Sharot, T., Korn, C. W., & Dolan, R. J. (2011). How unrealistic optimism is maintained in the face of reality. Nature Neuroscience, 14(11), 1475-1479.

Tomz, M. R., &Weeks, J. L. P. (2013). Public opinion and the democratic peace. American Political Science Review, 107(4), 849-865.

Appendix A: Additional tables

***Table A.1*:** Information provided in treatments HIGH and LOW (translated from German)

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 and 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 who care about animals are not necessarily kind-hearted people who care about the wellbeing of others and the good rules of living together!

References to Table A.1

Arluke, A. and Madfis, E. (2013): Animal abuse as a warning sign of school massacres. Homicide Studies 18(1), 7-22.

Eckardt Erlanger, A. C. & Tsytsarev, S. V. (2012): The relationship between empathy and personality in undergraduate students' attitudes toward nonhuman animals. Society and Animals 20 (1), 21-38.

Levin, J., Arluke, A., & Irvine, L. (2017): Are people more disturbed by dog or human suffering? Influence of victim's species and age. Society & Animals: Journal of Human-Animal Studies 25(1), 1-16.

Table A.2: Variable descriptions

Treatment variables	
HIGH	Dummy variable that equals 1 if participant is in HIGH treatment
LOW	Dummy variable that equals 1 if participant is in LOW treatment
BUBBLE	Dummy variable that equals 1 if participant is in BUBBLE treatment
CONFRONT	Dummy variable that equals 1 if participant is in CONFRONT treatment
Control variables	
Female	Dummy variable that equals 1 if participant reports to be female (rather than male or other).
Age categ.	Dummies for three age categories: age_1: 'below 35 years', age_2: '35-64 years', age_3: 'above 64'.
Emotions	Categorial variable with seven categories. 0 indicates that the participant reports that (s)he does 'not at all' and 6 that the participant reports to 'very much' respond emotionally to the Horncow Initiative. 233 respondents stating a very high emotional involvement (*Emotions* = 6) were excluded from the survey due to the guidelines of the panel provider LINK that aim to protect participants from stressful exposure during the study. This screening was pre-registered. Hence, in the sample the variable takes values 0 to 5.
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 indicates 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.
FreqMeat	Categorial variable on an eight-point scale capturing the self-reported 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, equals 1 if participant reports to adhere to a vegan diet.

NoEggsMilk Dummy variable, equals 1 if participant reports not to eat eggs and milk.

GoodEffects Categorial variable on a seven-point Likert scale measuring the agreement with the statement that consequences are more important than intentions of someone's actions. 1 represents 'certainly disagree' and 7 'certainly agree'.

GoodIntent Categorial variable on a seven-point Likert scale measuring the agreement with the statement that intentions are more important than consequences of someone's actions. 1 represents 'certainly disagree' and 7 'certainly agree'.

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

PriorAttitude Categorial variable elicited in wave 1 before the interventions on a seven-point Likert scale measuring the attitude towards the Horncow Initiative. 1 represents 'Certainly against' and 7 'certainly in favor'. Normalized to range from -1 to 1 in steps of 0.333. *PriorAttitude = (Elicited response - 4)/3*

AvoidCON Dummy variable that equals 1 if participant chooses not to read the arguments opposing the Horncow Initiative.

ReadOpp Dummy variable that equals 1 if participant chooses to read the arguments opposing his/her own *PriorAttitude* towards the Horncow Initiative.

IntVote Based on 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 in favor of the initiative', 5 'certainly vote in favor of the initiative'. *IntVote* = (*elicited response* - 3)/2 is a normalized version ranging from 'certainly vote against the initiative' (-1) to 'certainly vote in favor of the initiative' (1).

ΔIntVote 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: Δ*IntVote =* [*IntVote - PriorAttitude*]/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.
RepVote	Dummy that equals 1 if the participant self-reports to have voted in favor of the initiative and 0 against. Abstainers are treated as missing.
ΔRepVote	Variable bound to interval [-1,1] capturing the normalized difference between the self-reported actual voting and *PriorAttitude*. The variable is computed as follows: Δ*RepVote* = *RepVoting* – (*PriorAttitude*-1)/6 such that negative numbers indicate that the likelihood to vote in favor of the initiative has decreased relative to the attitude expressed prior to the interventions.
AgreePRO	Based on a categorical variable capturing how convincing PRO arguments just read are in the opinion of the participant: 5 'not at all convincing'; 4 'more unconvincing than convincing, 3 'neither convincing nor unconvincing', 2 'more convincing than unconvincing', 1 'fully convincing'. *AgreePRO* = (3 - *elicited response*)/2 is a normalized to range from 'not at all convincing' (-1) to 'fully convincing' (1).
ΔAgreePRO	Variable bound to interval [-1,1] capturing the normalized difference between the self-reported agreement with arguments in favor of the initiative at the end of the first wave and *PriorAttitude*. The variable is computed as follows: Δ*AgreePRO* = [*AgreePRO* - *PriorAttitude*]/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.

Note: ΔIntVote, ΔRepVote, and *ΔAgreePro* are constructed variables not included in the pre-registration.

Sample attrition

Between inviting participants and completion of the second wave, we lost participants for three reasons: non-response to key variables in the first wave of the survey (362 participants), deliberate screening based on the pre-registered panel guidelines of the LINK Institute (214 participants), and self-selection by participants (504 participants).

For the first type, we cannot report descriptive statistics as the data is missing. We find that incomplete responses do not systematically correlate with treatment assignment (Chi²-tests: HIGH $p = .79$, LOW $p = .14$, BUBBLE $p = .73$, CONFRONT $p = .98$).

The second type of attrition occurs by design. Screening out participants with extreme emotional involvement with the Initiative changes the composition of the sample. The first two columns in Table A.4 show that the treatment effects measured in wave 2 are not representative of the original sample recruited, i.e. before screening, according to the attrition tests proposed in Ghanem et al. (2022). However, the treatment effects measured in wave 2 are representative of the "core sample" (i.e., the sample excluding those with extreme emotional responses). Screening was not correlated with treatment assignment (Chi²-tests: HIGH $p = .87$, LOW $p = .99$, BUBBLE $p = .12$, CONFRONT $p = .13$).

The third type of attrition is based on self-selection of participants into the second wave of the survey. Systematic self-selection across treatments could undermine the internal validity of our findings. However, we find that completion of the second wave of the survey was not

treatment-specific (Chi²-tests: HIGH p = .70, LOW p = .57, BUBBLE p = .49, CONFRONT p = .81).

Table A.3 reports descriptive and test statistics for the treatments and the samples used (see Table 2) in the analysis reported in the main part of the paper. Table A.3 shows that in NEUTRAL older participants (their share increases from wave 1 to 2 by 2.4 percentage points) and better-informed participants tended to drop out less, and in HIGH those initially opposed tended to drop out less (significant at the 5% level). The values of all other variables are not significantly different across waves (see footnote to the table for description of the test used).

Table A.4 (columns 3 and 4) confirms that the treatment effects are internally consistent both for the sample of those completing both waves of the survey and for those that were invited to complete the second wave.

Table A.3: Attrition by treatment

	LOW			HIGH			NEUTRAL			Attrition		
Wave	1	2	*p*	1	2	*p*	1	2	*p*	LOW	HIGH	NEUT
BUBBLE	.268	.278	.52	.209	.211	.85	.236	.241	.70	.009	.002	.005
CONFRONT	.241	.243	.90	.230	.223	.58	.244	.244	.99	.002	-.007	-.000
FEMALE	.562	.546	.33	.572	.546	.08	.536	.523	.36	-.016	-.027	-.013
age_1	.249	.256	.67	.241	.266	.07	.194	.179	.20	.006	.024	-.015
age_2	.571	.569	.90	.586	.560	.09	.628	.618	.49	-.002	-.026	-.010
age_3	.180	.176	.75	.173	.174	.91	.179	.203	.03	-.004	.001	.024
FARMER	.004	.006	.31	.010	.009	.76	.020	.019	.71	.002	-.001	-.002
FarmHorn	.002	.003	.47	.004	.003	.62	.011	.008	.32	.001	-.001	-.003
Informed	.277	.304	.45	.304	.389	.14	.296	.407	.03	.027	.084	.110
PriorAttitude	-.050	-.055	.79	-.082	-.129	.02	-.071	-.072	.93	-.005	-.046	-.002
Post-treatment variables												
IntVote	-.055	-.065	.61	-.034	-.046	.54	-.068	-.073	.76	-.011	-.013	-.005
RepVote		.386			.400			.364				
AgreePRO	.238	.208		.256	.252		.171	.177		-.030	-.004	.007
ReadCONonly	.019	.022		.019	.017		.019	.024		.003	-.002	.006
ReadPROonly	.038	.042		.040	.04		.035	.041		.003	.000	.005
ReadBoth	.780	.799		.783	.794		.795	.808		.019	.011	.012
AvoidCON	.201	.179		.198	.189		.186	.168		-.022	-.009	-.018
ReadOpp	.801	.821		.795	.803		.810	.829		.020	.008	.019

Notes: Variable means for treatments HIGH, LOW and NEUTRAL for wave 1 (N = 1,536) and 2 (N =1,032) and differences between waves (columns 'Attrition'). For variables elicited before treatment, p-values for treatment columns are from two-sided tests of proportions for dummy variables and Mann-Whitney for categorial variables each testing the difference between dropouts and retainers. p-values are not adjusted for multiple hypotheses testing.

***Table A.4*:** Regression-based attrition tests of internal validity

	W1 largest sample -> W2 final sample		W1 core sample -> W2 final sample	
	HIGH/LOW	BUBBLE/CONFRONT	HIGH/LOW	BUBBLE/CONFRONT
internal validity for the:				
respondent subpopulation	.819	.529	.605	.443
study population	.000	.000	.344	.266
N	1,750	1,750	1,536	1,536

Note: Reports *p*-values of regression-based attrition tests proposed in Ghanem et al. (2022) and implemented in Stata command *attregtest* of internal validity of treatment effects for the respondent and the study population. Tests are based on the pre-treatment outcome variable *PriorAttitude* elicited in wave 1. 'W1 largest sample' refers to the largest sample in wave 1 for which the variables necessary to conduct the tests are available. 'W1 core sample' is the sample presented in Table 2 that excludes participants with the highest emotional involvement and those for which key outcome and control variables are missing. 'W2 final sample' is the subsample of participants that completed wave 2 of the survey (see also Table 2).

Table A.5 compares pre-treatment attitudes (*PriorAttitude*), post-treatment intended voting (*IntVote*) and post-ballot reported voting (*RepVote*) in our sample with the official results in the ballot for all of Switzerland and for the German-speaking cantons from which our survey sample was drawn. The next-to-last row shows that the initiative was rejected with 54.7% in all of Switzerland (54.9% in the German-speaking cantons). Intended voting of those completing both waves was similar to the final result (53.7% if "Neutral" voters are assumed to evenly split to those favoring and opposing). However, the reported voting was clearly more negative than the official result (61.7% vs 54.9%).

Table A.5: Attitudes and Voting in Sample vs. Ballot

	Attitude towards the initiative			
	In Favor	**Opposing**	**Neutral**	*N*
Participants completing wave 1				
PriorAttitude	38.4%	44.1%	17.4%	1,536
IntVote	36.9%	42.9%	20.1%	1,521
Participants completing both waves				
PriorAttitude	37.4%	45.4%	17.2%	1,032
IntVote	36.7%	44.0%	19.4%	1,017
RepVote	38.3%	61.7%		768
Ballot Result				
all of Switzerland	45.3%	54.7%		2.53 million
German-speaking cantons	45.1%	54.9%		1.93 million

Note: Source of ballot results: Bundesamt für Statistik, Statistik der eidg. Volksabstimmungen (Abst.-Nr. 6230). *N* = 2.53 Mio. refers to valid votes. *PriorAttitude* was elicited before any treatment intervention. "In favor" groups the three response categories that indicate attitudes supportive of the initiative, "Opposing" groups the three response categories that indicate attitudes opposing the initiative and "Neutral" is the middle category. *IntVote* was elicited after treatment interventions at the end of wave 1 on a five-point Likert scale. 3 = "Neutral" 1 &2 = "Opposing", 4 & 5 "In Favor". *RepVote* is a dummy representing whether participants reported to have voted in favor of the initiative (1) or against the initiative (0).

Table A.6 reports a robustness check for the main finding reported in Table 2 that treatment HIGH (but not LOW, BUBBLE or CONFRONT) affects voting. Table 2 used sample sizes of 1,356 subjects (Wave 1) and the proper subsample of 1,032 while Table A.6 uses the largest possible samples for which the respective analysis can be performed (1,741 and 771 subjects, respectively). These larger samples include subjects who did not answer questions for control variables as well as – for *IntVote* – those who were screened out due to their extreme emotional involvement with the initiative. Table A.6 shows that results reported in Table 2 are robust to such inclusion.

Table A.6: Treatment effects on voting for largest possible sample

	(1)	(2)	(3)	(4)
	IntVote	Δ*IntVote*	*RepVote*	Δ*RepVote*
HIGH	0.044	0.023	0.059	0.064
	(0.040)	(0.038)	(0.036)	(0.029)
LOW	0.005	0.001	0.006	0.017
	(0.817)	(0.907)	(0.825)	(0.573)
BUBBLE	0.020	0.012	0.028	0.019
	(0.368)	(0.284)	(0.329)	(0.516)
CONFRONT	-0.030	-0.016	0.009	0.017
	(0.182)	(0.162)	(0.763)	(0.578)
PriorAttitude	0.864		0.388	
	(0.000)		(0.000)	
_cons	0.001	-0.000		-0.091
	(0.948)	(0.974)		(0.000)
N	1,741	1,741	772	772
R^2	0.719	0.006	0.486	0.007
F	886.3	2.4	500.5	1.4

Notes: Columns (1), (2) and (4) show coefficients from OLS regressions, column (3) shows marginal effects of probit a regression. The specifications shown here do not include further controls. Estimates are based on all participants that answered questions on the respective outcome variables. No further restrictions were applied. In particular, regressions (1) and (2) include those participants that were not invited to wave 2 of the survey based on their high emotional involvement to the initiative. *p*-values in parentheses.

Information selection and processing

Table A.7 reports post-treatment information selection by treatment. Most subjects (94.3%) choose to read both PRO and CON arguments (78.6%) or none (15.7%). A remarkable 80.2% of subjects read arguments opposing their initial position (e.g., read CON arguments when their *PriorAttitude* was in favor of the initiative). There are no significant differences across treatments (see section 3.3).

Table A.7: Descriptive statistics on information selection across treatments

Arguments read	HIGH	LOW	NEUTRAL	BUBBLE	CONFRONT	NOCHAT	ALL	*N*
Both sides	78.3%	78.0%	79.5%	79.9%	78.4%	78.2%	78.6%	1,208
None	15.8%	16.3%	15.1%	15.7%	16.1%	15.5%	15.7%	241
Only PRO	4.0%	3.8%	3.5%	2.7%	4.1%	4.1%	3.8%	58
Only CONTRA	1.9%	1.9%	1.9%	1.6%	1.4%	2.2%	1.9%	29
Opposing	79.5%	80.1%	81.0%	81.0%	79.5%	80.1%	80.2%	1,232
N	526	473	537	364	251	806		1,536

Note: Table shows percentage of subjects selecting to read a particular combination of arguments. 'Opposing' refers to the list of arguments that support the opposite position towards the initiative compared to that 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. The fifth row overlaps with rows 1, 3 and 4.

Table A.8 reports regression results testing H2 (Self-image concerns operate through selection of arguments) and H4 (Social image concerns operate through selection of arguments). None of the four treatments has a systematic impact on information acquisition by survey participants. H2 and H4 are therefore not confirmed.

Table A.8: Regression analysis of information selection

	(1) *AvoidCON*	(2) *AvoidCON*	(3) *ReadOpp*	(4) *ReadOpp*
HIGH	0.012	0.013		-0.015
	(0.634)	(0.604)		(0.532)
LOW	0.015	0.016		-0.009
	(0.557)	(0.529)		(0.714)
BUBBLE		-0.012	0.009	0.009
		(0.633)	(0.721)	(0.720)
CONFRONT		0.006	-0.006	-0.006
		(0.800)	(0.800)	(0.795)
Controls	*No*	*Yes*	*No*	*Yes*
N	1,536	1,536	1,536	1,536
r^2	0.00	0.01	0.00	0.01
LR Chi²	0.82	15.5	0.3	18.4

Notes: Marginal effects of probit regressions on information selection. Regressions (1) and (2) test H2 where *AvoidCON* is a dummy that indicates whether participants avoid reading CON arguments (1) or not (0). Regressions (3) and (4) test H4 where *ReadOpp* is a dummy that indicates whether participants read arguments of the side opposing their own prior attitude. Recall that we did not hypothesize an impact of *CONFRONT* on *AvoidCON*. *p*-values in parentheses

Table A.9: IV regression of biased processing

	(1) *IntVote*	(2) *ΔIntVote*	(3) *RepVote*	(4) *ΔRepVote*
AgreePRO	0.559		2.574	
	(0.029)		(0.000)	
PriorAttitude	0.592		0.440	
	(0.000)		(0.533)	
ΔAgreePRO		0.625		0.900
		(0.013)		(0.056)
_cons	-0.130	-0.077		-0.188
	(0.030)	(0.018)		(0.005)
N	1,184	1,184	627	627
R^2	0.726	0.019		
(Wald) Chi^2	3015.3	6.2	382.1	3.7

Notes: IV regressions where (*Δ*)*AgreePRO* is instrumented with treatments HIGH and LOW. Regression (3) reports marginal effects of an IV probit regression. Regressions (1), (2) and (4) report coefficients of two-stage least square regressions. *p*-values in parentheses.

Appendix B: Theory background

We present here the simplest model we can think of to illustrate the impact of our experimental manipulation on information processing and voting.

There are two states of nature, $x = 0$ and $x = 1$. If $x = 1$, the initiative, if accepted, would improve animal welfare. If $x = 0$, animal welfare would remain unchanged even if the initiative is accepted. If the initiative is rejected, animal welfare remains unchanged.

Individual i observes the state of the world x. But he can pay a cost c_i to bias his belief so that he believes the opposite state of the world to obtain. In particular, if $x = 0$, he can pay c_i to obtain belief $x' = 1$. If he does not pay c_i, then $x' = x$.

His action taken after belief formation is $v = 1$ (Yes vote) or $v = 0$ (No vote). The agent derives utility from voting according to his subjective preferences: $u(v|x' = 1) = \{\mu \text{ if } v = 1, 0 \text{ else}\}$ and $u(v|x' = 0) = 0$. The parameter $\mu > 0$ is discussed below.

It is easy to see from this utility function that there is no incentive to bias one's belief when $x = 1$, but there is an incentive to do so when $x = 0$. The agent biases his belief, moving from true $x = 0$ to false $x'=1$ (and naively forgetting that x' was forged) if and only if $\mu > c_i$.

How to interpret μ? It can be conceived as $\mu = m + \pi$, where $m \geq 0$ is the hedonic utility derived by the individual from the real consequence of the initiative on animals (possibly accounting for the probability of being pivotal), and π the "self-image utility" if the voter believes that if he improves animal welfare then he is a good person.

In the experiment, we manipulate this self-image utility π. Since voters are indexed by c_i, by increasing π in HIGH, we thus increase the share of Yes voters by increasing the share of those who bias themselves.

Le Yaouanq (2021) provides a more general model of voting with biased beliefs. Building on the memory management model of Bénabou and Tirole (2002, 2011), Le Yaouanq shows in his Proposition 1 that, in any equilibrium, an increase in μ increases the probability to bias beliefs, consistent with the prediction above.

Appendix C: Exploratory analysis

In this section, we report a number of 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 C.1 shows that intended as well as reported votes in favor of the initiative decrease in the frequency of meat eating (see first row), but not with other dietary habits related to animal products.

Table C.1: Correlation of eating habits and voting

	(1) *IntVote*	(2) *IntVote*	(3) *RepVote*	(4) *RepVote*
Freq_Meat	-0.040	-0.069	-0.113	-0.117
	(0.002)	(0.004)	(0.001)	(0.059)
Vegetarian		-0.149		-0.057
		(0.355)		(0.893)
Intens_Meat		0.076		-0.010
		(0.163)		(0.944)
Vegan		-0.314		-0.726
		(0.386)		(0.490)
NoEggMilk		0.193		0.503
		(0.240)		(0.290)
Constant	0.151	0.271	0.263	0.280
	(0.026)	(0.015)	(0.145)	(0.340)
N	1,520	1,516	768	766
F /Chi²	9.6	2.7	10.5	11.7
r^2	.006	.009	.01	.01

Notes: Regressions (1) and (2) report coefficients from OLS regressions, (3) and (4) marginal effects of probit regressions. *Freq_Meat* is an eight-point categorial variables measuring the frequency of meat eating ranging from 1: 'never' to 8: 'several times a day'. All other variables are dummies indicating whether a participant is a vegetarian, an intensive meat eater, a vegan or does not eat eggs and milk. *p*-values in parentheses.

Table C.2 tests whether prior informedness on the initiative correlates to 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 C.2), both are no longer significant.

Table C.2: Prior information (Dep. Var. *Informed*)

	(1)	(2)	(3)
GoodEffects	0.064		0.040
	(0.050)		(0.237)
GoodIntent		0.078	0.055
		(0.012)	(0.095)
N	995	999	983
Chi²	3.9	6.3	2.9
Pseudo R^2	0.004	0.006	0.006

Notes: coefficients from OLS regressions, *GoodEffects* measures agreement with the statement that consequences are more important than intentions of someone's actions. 1 'Certainly against', 7 'certainly in favor'. *GoodIntent* uses the same scale but asks for agreement with the opposite. *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 C.3). 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 participant 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 C.3: Information selection

	(1)	(2)	(3)	(4)	(5)	(6)
		AvoidCON			***ReadOpp***	
GoodEffects	0.091		0.057	-0.089		-0.053
	(0.103)		(0.327)	(0.108)		(0.362)
GoodIntent	-0.037		-0.017	0.095		0.080
	(0.490)		(0.771)	(0.078)		(0.156)
Emotions		-0.168	-0.146		0.161	0.150
		(0.000)	(0.004)		(0.000)	(0.003)
Overconfident		0.984	1.183		-1.059	-1.217
		(0.000)	(0.000)		(0.000)	(0.000)
Constant	-1.784	-1.239	-1.679	1.512	1.264	1.375
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
N	983	1,499	960	983	1499	960
Chi²	2.7	66.4	51.8	4.4	73.0	56.5
Pseudo R^2	0.003	0.045	0.059	0.005	0.049	0.064

Notes: Logit regressions. Dependent variable: *AvoidCON* (regressions (1) – (3)) and *ReadOpp* (regressions (4) – (6)). Regressions (2) 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.