Lasting Political Attitude Change Induced by False Feedback About Own Survey Responses

David Sivén (david.siven@gmail.com)
Thomas Strandberg (thomas.strandberg@lucs.lu.se)
Lars Hall (lars.hall@lucs.lu.se)
Lund University Cognitive Science, Lund University, Box 192, S-221 00, Lund, Sweden.
Petter Johansson (petter.johansson@lucs.lu.se)
Lund University Cognitive Science, Lund University, Box 192, S-221 00, Lund, Sweden.
Swedish Collegium for Advanced Study, Linneanum, Thunbergsvägen 2, S-752 38 Uppsala, Sweden.
Philip Pärnamets (philip.parnamets@ki.se)
Lund University Cognitive Science, Lund University, Box 192, S-221 00, Lund, Sweden.
Division of Psychology, Karolinska Institutet, Nobels väg 9, S-171 77, Solna, Sweden

Abstract

False feedback on choices has been documented to induce lasting preference change. Here we extend such effects to the political domainand investigatethetemporal persistence of induced preferences, as well as, the possible role the length of confabulatory justifications may play. We conducted a two-day choice blindness experiment using political statements, with sessions being roughly one week apart. Changes in political preferences remained one week after initial responses, and were most prominent in participants who were allowed to confabulate freely. These findings, being the first to demonstrate lasting preference change using choice blindness, are discussed in light of constructivist approaches to attitude formation through a process of self-perception.

Keywords:political attitudes; attitude change; choice blindness; persuasion; confabulation

Central to human social interaction is not only to understand the attitudes and preferences of your interlocutors, but also how to influence and shape them.In democratic societies this can be the difference between, successfully running for office and affecting the future course of one's society, or facing electoral defeat and slowly recede into the margins of history.

Attitudes are most commonly defined as "psychological tendencies that are expressed by evaluating a particular entity with some degree of favor or disfavor" (Eagly & Chaiken, 1993), parts of larger interrelated networks of attitudes and values (Jost, Federico & Napier, 2009). Apartfrom evidence indicating that attitudes are constructed within the domain of a specific task in a specific context, there is also evidence that preferences sometimes arise not before an actual decision is made, but rather as a result of the choice made or action performed (Ariely & Norton, 2008).

In the domain of public opinion, political attitudes are often seen as deliberate mental states, central in societal life, that are motivating important behaviors such as voting, or donating money to charity etc. However, there is also a perspective that political attitudes are highly flexible, often constructed on the fly, and prone to contextual influence (Azjen et al., 2014; Bishop, 2005). So, if attitudes are guiding action and behavior, then understanding the functions of attitude change and contextual influence are crucial in deeply understanding the cognition involved in constructing political preferences and persuasive messages.

Choice Blindness and Preference Change

Choice blindness is the finding that participants are at times blind to false feedback about the outcome of previously made choices. In the classical experiment participants make binary preferential choices between pairs of faces, and, following a covert manipulation, are subsequently presented with the opposite face to their original choice (Johansson, Hall, Sikström & Olsson, 2005). The key findings are that participants only detected about 25% of the manipulations, and that they often constructed coherent arguments supporting the opposite of the original choice.

Choice blindness has recently been proposed as a viable alternative to the *free-choice-paradigm* (Brehm, 1956) as a method for studying the effect of choices on later preferences. If participants are asked to make a second round of choices, they are more likely to change their preference following false feedback about their choice compared to when they receive veridical feedback (Johansson, Hall, Tärning, Sikström & Chater, 2014). This has been interpreted as supporting self-perception view of the mechanisms underlying preferential change (Johansson et al., 2014). Similarly, following a choice blindness manipulation participants' source memory becomes distorted leading them to believe they have made choices in line with the false feedback (Pärnamets, Hall & Johansson, 2015).

However, the longevity of choice-induced preferences is uncertain. Using a choice blindness manipulation Taya,Gupta, Farber and Mullette-Gillman, (2014), investigated the temporal extent of choice induced preferences over a two day experiment using photographs of faces as stimuli. They found an indication of preference change as a result of the false feedback in the short-term, but no lasting effect. This could, however, be due to the fact that the participants were made aware of the manipulation

prior to the second round of choices.By contrast, using an alternative paradigm known as blind-choice, where participants are led to believe they have made subliminally guided ("blind") choices, Sharot and colleagues found an effect lasting 2-3 years for ratings concerning preferred holiday destinations (Sharot, Fleming, Yu, Koster & Dolan, 2012).

Here we are concerned with investigating whether false feedback about responses in a political survey can be used to induce both immediate as well as lasting changes in participants' political attitudes. Previous work has demonstrated the applicability of the choice blindness paradigm to both moral (Hall, Johansson & Strandberg 2012), as well as, political attitudes (Hall et al., 2013) using covert manipulations of ratings on paper-based scales. In Hall et al. (2013) participants' responses on a range of salient political issues were manipulated weeks before a general national election. The false feedback was issued so as to go consistently against the grain of participants' left/right-wing orientation. Correction rates were low: 22% of manipulated statements, 48% of participants failed to make any corrections. More strikingly, participants were asked pre- and post-test to rate their voting intention on a left- to right-wing scale, with findings that almost half the participants changed their voting intentions in the direction of the manipulation of their underlying attitudes. However, any effect of the false feedback on the individual political issues was not tested.

Attitudes and Reasoning

An important aspect of the choice blindness paradigm is that it not only involves false feedback about a previous choice or rating, but also an element of confabulation. When confabulating, a person is unconsciously expressing fabricated aspects about oneself, the world, or the reasons behind a choice, without any deceptive intentions (Fotopoulou, Conway, & Solms, 2007). If the false feedback in a choice blindness task is accepted, we know that whatever reasons are given are confabulations.

The literature on confabulation in everyday discourse is scarce, and most research has traditionally focused on confabulation in the clinical domain (Hirstein, 2010). Inspired by research on attitude strength and elaboration, we wanted to investigate the effect of confabulation as a moderator of the attitude change induced by the false feedback. There are studies on persuasion suggesting that elaboration is a causal mechanism responsible for attitude strength (Barden and Tormala, 2014). Following this view, more elaboration produces stronger judgments held with greater confidence. In a study on the effects of elaboration on attitude strength, Barden and Tormala (2014) found that attitude strength largely depended on people's perception of their own elaboration.

In the context of attitudes, according to the *elaboration likelihood model* of persuasion (ELM) (Petty & Cacioppo, 1986; Petty, Haugtvedt, & Smith, 1995), the strength of an attitude can be predicted by the amount of issue-relevant thinking a person has spent on the object (Petty, Haugvedt, & Smith, 1995). The ELM builds on literature concerning attitude persistence and suggests that attitude change can be seen as resulting from two different kinds of persuasion. The first is persuasion by perceptual cues, not involving any careful deliberation as to the merits of a specific argument. The second is careful and issue-relevant consideration of a specific argument (Petty & Cacioppo, 1986).Clarkson, Tormala, and Leone (2011) found that if participants get to think about some object for up to 300 s, their confidence regarding their own attitudes directed at this object will increase, and their attitudes will become more polarized (extreme). Shorter deliberation times (60 s) were shown to lead to lower confidence and attitude depolarization (Clarkson, Tormala, & Leone, 2011).

In this study we were interested if changing the amount of confabulatory response participants were asked to give in response to the false feedback would affect the amount of change in their ratings following the initial manipulation. We expected more elaborate confabulationsto enhance the self-perception mechanisms hypothesized to underlie choice induced preference change and increase both the size of the attitude change as well as its longevity.

Method

Participants

We recruited 140 participants (88 female, 52 male) from the student population at Lund University, with an average age of 22.7 (*SD* = 3.0). Participants received two cinema vouchers in exchange for their participation in two experimental sessions, roughly one week apart.

Materials

For registering answers and manipulating the participants' ratings, we used a Self-Transforming Suvey (STS) specifically developed for providing false feedback about attitude responses (Strandberg et al., in prep.). The STS presented political statements one at a time in a randomized order. In addition to the STS, two follow-up paper-based surveys were used for measuring any attitude change following the manipulation.

All three surveys consisted of 12 political statements.The political statements were very specific, and divided into the three subcategories of healthcare, school politics, and environmental issues. Of the statements in the STS, six were carried forward to the two follow-up surveys. Of these six statements, four were always the same target statements. For each participant two of the target statements would be manipulated. The four target statements concerned salient political issues in Sweden at the time of the experiment [spring 2015]. The issues were: higher gasoline taxes for urban denizens, introducing subsidies for energy-efficient household appliances, making the government rather than local municipalities responsible for public schooling, and, introducing free after-school homework programs. All statements were constructed so as to state a proposed policy

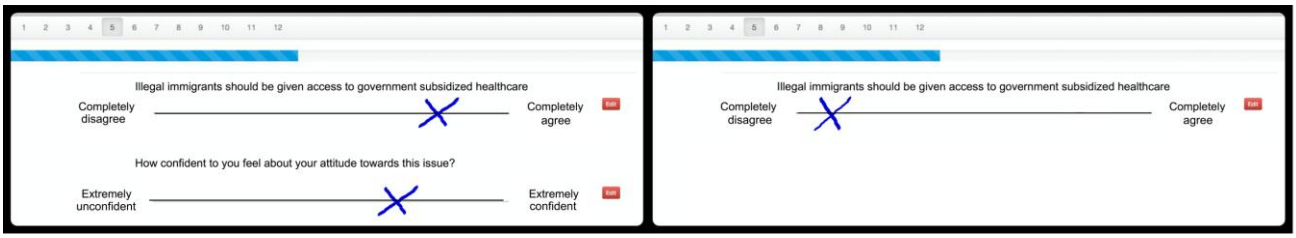

Figure 1 The STS. **Left:** Initial rating of a target statement. The X on the upper axis corresponds with agreeing with the statement. The X on the lower axis corresponds to neutral-to-high confidence**. Right:** False feedback following the completion of the survey. The application has now moved the X to correspond with disagreeing with the statement.

and give a brief explanation of that policy. One example full statement from the set of targets is:

The Swedish elementary school should be re-nationalized. Apart from the fact that local municipalities would lose much, albeit not all, influence, a re-nationalization would mean that the state becomes head of the school and assumes the responsibility for resource allocation and quality assurance.

The remaining items were considered fillers and had the same structure as the target statements. The filler items in each follow-up survey were unique to that survey.

Finally, an audio recorder was used to record the explanations of responses.

Choice Blindness and Confabulation Conditions

There were two independent variables in the experiment, one being the false feedback about responses, i.e. the choice blindness manipulation, the second being what confabulatory condition participants were assigned to.

The false feedback manipulation moved the participants rating, an 'X' they drew on the tablet across the midline of the bidirectionalvisual-analoguescale (see Fig. 1).The X ismoved by the STS to the other side of the axis andplaced either somewhere between 15% and 35% or between65% and 85%. The reason for not moving them closer to themiddle than this is to manipulate the participants intobelieving that they have a somewhat clear attitude for oragainst the statement.

In addition, participants were randomly assigned to one of two confabulatory conditions: Short or Long. In the Short condition, when participants were confronted with their ratings, they were asked to read aloud each statement as they appeared on the tablet, state where on the axis they had written their X, and if this meant that they agreed or disagreed with the statementas well as to what extent (e.g. whether the rating meant that they strongly agreed with the statement). In the Long condition, the participant was asked to, in addition to the above; also as thoroughly as possible account for the reasons behind their answer.

Procedure

The experiment consisted of two sessions, separated by roughly one week.

First Session (T1 & T2) The first session consisted of three subtasks: initial rating, interaction with manipulated answers, and follow-up rating.

First the participant was asked to answer 12 statements (T1) on a tablet, by marking an X along an axis in response to each statement. The endpoints were anchored as "Completely disagree" and "Completely agree". The experiment instructions emphasised that they were intended to represent extremes on a possible spectrum. Following each statement participants were asked to state their confidence in their attitude using the same scale anchored with "Extremely uncertain" and "Extremely certain". The participant was left to respond to the survey at her own pace.

Once the participant completed, the experimenter returned to the room and explained the interaction task. During this subtask, unbeknownst to the participants, the choice blindness manipulation took place. Each participant was told that the tablet would now randomly present four of the 12 statements and also the participants' responses to each of these statements. Instead the presented statements presented were the four target statements. Participants' responses to two of the presented statements had now been manipulated by the survey application running on the tablet. Participants were asked to read each of the presented statements and their rating, according to which condition, Short or Long, they had been assigned to (see above).

While performing the second subtask, all participants were knowingly and willingly recorded by an audio recorder. In the event of a participant in any way indicated that the response they saw did not correspond with their view, they were told that they could change their response if they wanted to, after which they could base their explanation on the position of the X in the now corrected position.

The third subtask (T2) consisted of the participant responding to a follow-up survey very similar to the one they had previously responded to on the tablet; only it was in the form of a traditional pen-and-paper survey. Six of the 12 statements were new to the participants, but the four possible target statements as well as two statements concerning healthcare, remained in the exact same form as in the STS. The participants were told that the motivation behind this additional survey was to measure if there was

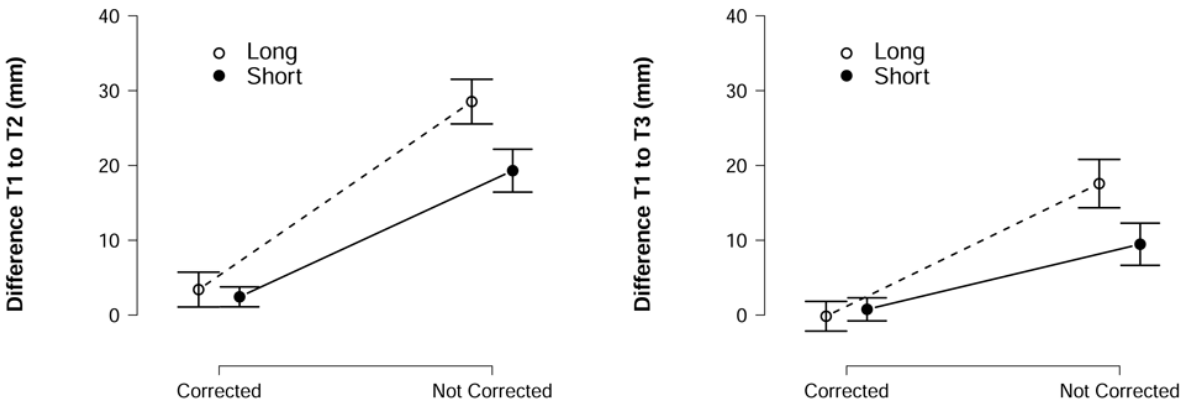

Figure 2 Differences in ratings during manipulated trials divided by correction and confabulatory condition. **Left:** Differences between original rating (T1) and first follow-up (T2). **Right:** Differences between original rating (T1) and second follow-up one week later (T3). Error bars are 95% confidence intervals.

any difference in responses if a survey was conducted electronically or physically. Further, the participants were told that it was likely that some of the statements that they had responded to on the tablet also would be included in the paper survey, since they were all randomly selected from the same bank of statements. Upon completion, the experimenter and the participant agreed on a time and date for the second session.

Second Session (T3) The participants were asked to return a week following the first session. The second session took place on average 6.3 days (*SD* = 1.8) later.During this second session (T3) participants were asked to fill out a survey of the political attitudes, similar to the one they had filled out the previous week. Once again, six statements were carried forward from the original and first follow-up survey, and six new statements were presented. Upon completion participants were debriefed in full, signed informed consent and data release statements and paid for their participation.

Analysis

All ratings were converted to a 0-100mm scale to facilitate comparisons between mediums (i.e. STS and paper-pen). We analyzed our data using linear mixed-effects models from the `lme4` package in R. Random-effects were modelled as per participant intercepts and slopes mirroring the full fixed-effects structure, or the maximally permitted structure that would converge. Significance of fixed-effects was assessed with likelihood ratio tests using the `car` package.

Results

Correction Rates

Of the 277 manipulated (M) trials, 134 (48.4%) were corrected by participants. Average by participant correction rate was 0.96 (*SD* = 0.78). 45 (32%) participants made no corrections, 56 (40%) one correction and 39 (28%) two corrections.

Overall Effect of Manipulation on Ratings

Rating differences comparing by manipulation We first analyzed differences in ratings for T2 and T3 compared to T1 (i.e. difference scores) based on manipulation. Time and manipulation were dummy-coded taking T2 and non-manipulated (NM) trials as the reference levels.

We found significant main effect of manipulation ($\chi^2_{(1)}$ = 39.23, $p = 3.7*10^{-10}$) as well as a significant interaction between time and manipulation ($\chi^2_{(1)} = 31.64$, $p = 1.9*10^{-8}$), but no main effect of time ($\chi^2_{(1)} = 2.41$, $p = .12$), with model conditional R^2 = .35. Interpreting the coefficients, participants were somewhat accurate in restating their original attitude in T2 during NM-trials ($b_{intercept}$ = 7.5mm, *SE* = 0.7) and this changed little from T2 to T3 (b_{T3} = 2.7mm, *SE* = 0.9). There was a large increase in distance from original rating for T2 manipulated (M) trials (b_M = 10.9mm, *SE* = 1.3) which decreased during T3 ($b_{T3\_X\_M}$ = -7.6mm, *SE* = 1.4).

Rating differences comparing by correction To compare the effect of correcting or failing to correct the manipulation, we performed the same analysis as above, subsetting the data on M-trials only. Differences scores were here transformed to be directional, meaning that a positive change is to be interpreted as a change in attitude in the direction of the manipulation (compared to T1). Time and correction were dummy-coded with T2 and corrected (C) trials as reference levels.

We found significant main effect of correction ($\chi^2_{(1)}$ = 96.87, $p = 2.2*10^{-16}$) and of time ($\chi^2_{(1)} = 30.85$, $p = 2.8*10^{-8}$), as well as a significant interaction between time and correction ($\chi^2_{(1)}$ =10.31, $p = .0013$), with model conditional R^2 = .47. Interpreting the coefficients, participants were displayed virtually no directional change in attitudes in T2 during C-trials ($b_{intercept}$ = 2.3mm, *SE* = 1.3) and this changed little from T2 to T3 (b_{T3} = -2.7mm, *SE* = 1.8). There was a large directional change in attitudes for T2 not

corrected (NC) trials (b_{NC} = 21.3mm, *SE* = 2.2) which decreased, but far from entirely, during T3 (b_{T3*NC} = -7.9mm, *SE* = 2.5).

Qualitative shifts by correction Given the shifts in attitudes post manipulation, we examined the proportion of these that crossed the mid-line of the attitude spectrum – hence implying a qualitative shift. In T2, 73% of responses represented such a shift for NC-trials, compared to 10% for C-trials. In T3, where the attitudinal effects of the manipulation were weakened, 41% of responses were still qualitatively shifted for NC-trials compared to 10% for C-trials.

Effect of Confabulation Condition

Rating differences comparing by manipulation To capture the effect of confabulation condition, we added confabulation as a fixed and random effect to the model. We analyzed all trials, with time, condition and manipulation being dummy-coded taking T2, short and non-manipulated (NM) trials as the reference levels.We found significant main effects of manipulation ($\chi^2_{(1)}$ = 44.14, p = 3.1*10^{-11}),and of condition ($\chi^2_{(1)}$ = 3.92, p = .048), but not of time ($\chi^2_{(1)}$ = 2.91, p = .088). These were qualified by significant interactions between manipulation and time ($\chi^2_{(1)}$ = 32.97, p = 9.4*10^{-9}), as well as between condition and time ($\chi^2_{(1)}$ = 4.15, p = .041). The interactions between manipulation and condition ($\chi^2_{(1)}$ = 0.040, p = .84) and the three-way interaction were not significant ($\chi^2_{(1)}$ = 0.21, p = .65). Model conditional R^2 = .37.

Interpreting the coefficients, participants were somewhat accurate in restating their original attitude in T2 during NM-trials in the Short condition ($b_{intercept}$ = 5.6mm, *SE* = 0.9) and the difference increased from T2 to T3 (b_{T3} = 4.3mm, *SE* = 1.3). There was a large increase in distance from original rating for T2, Short, manipulated (M) trials (b_M = 11.0mm, *SE* = 1.8) which decreased during T3 (b_{T3*M} = -8.2mm, *SE* = 1.8). The Long condition increased differences (b_{LONG} = 3.7mm, *SE* = 1.5), though this effect disappeared for T3 ($b_{LONG*T3}$ = -3.3mm, *SE* = 1.8). There was no specific effect of confabulation length for M-trials at T2 (b_{LONG*M} = -0.2mm, *SE* = 2.5) or at T3 ($b_{LONG*M*T3}$ = 1.2mm, *SE* = 2.7).

Rating differences comparing by correction To assess the effect of condition on directional change in ratings, we performed the same analysis as above, subsetting the data on M-trials only. Differences scores were here, again, transformed to be directional, meaning that a positive change is to be interpreted as a change in attitude in the direction of the manipulation (compared to T1). Time, condition and correction were dummy-coded with T2, Short condition and corrected (C) trials as reference levels. We found significant main effects of detection ($\chi^2_{(1)}$ = 120.51, p = 2.2*10^{-16}), condition ($\chi^2_{(1)}$ = 4.30, p = .038), and of time ($\chi^2_{(1)}$ = 28.88, p = 7.7*10^{-8}). These were qualified by significant interactions between detection and condition ($\chi^2_{(1)}$ = 7.32, p = .0068), as well as between detection and time ($\chi^2_{(1)}$ = 9.80, p = .0017). The interactions between condition and time ($\chi^2_{(1)}$ = 0.047, p = .49) and the three-way interaction were not significant ($\chi^2_{(1)}$ = 0.0078, p = .93). Model conditional R^2 = .42.

Interpreting the coefficients (see also Fig 2), participants displayed no directional attitude change in T2, C-trials for either Short ($b_{intercept}$ = 2.7mm, *SE* = 2.3) or Long (b_{LONG} = 0.3mm, *SE* = 3.2) conditions, with no changes for T3 (b_{T3} = -1.9mm, *SE* = 2.8). Importantly, there was a large directional attitude change for NC-trials (b_{NC} = 16.9mm, *SE* = 2.9), which was enhanced for Long condition ($b_{NC*LONG}$ = 8.7mm, *SE* = 4.1). Both directional effects due to NC-trials and Long condition diminished during T3 (b_{NC*T3} = -7.8mm, *SE* = 3.6; $b_{LONG*T3}$ = -1.5mm, *SE* = 3.7; $b_{NC*TALK*T3}$ = -0.5mm, *SE* = 5.1), though not sufficiently to remove the lingering directional effect (see Fig. 2).

Discussion

We investigated whether false feedback concerning specific ratings to political statements would influence later attitudes to the same statements. We found that participants' ratings, both immediately following false feedback and one week later, were shifted in the direction of the manipulation. These effects were large, representing considerable shifts in the direction away from the original attitude and towards the opposite. Notably there were no concomitant shifts for corrected trials, where participants had noticed the manipulation or shift for non-manipulated trials. A large portion of the shifts indicated qualitative shifts in attitudes. That there are shift in ratings immediately following the false feedback is in line with previous findings on choices between faces (Johansson et al., 2014; Taya et al., 2014). In those previous studies, however, the false feedback concerned preferential binary choices between pairs of faces. Here, in contrast, we influence political attitudes, a domain where we would expect higher resilience of participants underlying attitudes (cf. Druckman, 2004).

Importantly, the attitude shifts were not only present shortly following false feedback, but also when the same statements and rating task was administered one week later. This is the first demonstration of lasting preference shifts using a choice blindness paradigm (*contra* Taya et al., 2014), and indicate that the effects of false feedback might become integrated into the participants' set of attitudes. That choices and feedback about choices influence future preferences is in line with findings from other choice paradigms (Sharot et al., 2010). In the case of political attitudes, as studied here, the attitude shifts were obtained absent any reinforcement of the altered position. The participant only viewed the manipulation once, and was given one chance to state her opinion and [in the Long condition] give an explanation for it. After the experiment, the participants immersed themselves in their ordinary life for a full week, with their usual sources of information and political biases. It is therefore remarkable to find any lingering effect at all.

As such these findings provide support to inferential and constructivist accounts of preference and attitude formation, whereby the act of choosing has a determining effect on a

persons' preference set (Ariely & Norton, 2008). In previous work we have suggested that this is consistent with a self-perception account of preferences – i.e. that we infer our own preferences much like how we infer others' preferences; by observing and interpreting overt behavior (Johansson et al., 2014).

The change in ratings most marked in participants in the Long conditions, who had been asked to give longer explanatory statements concerning their manipulated responses. This is important, as it suggests that confabulation serves to consolidate choice induced preference change in the same way as veridical argumentation. The effect of the Long condition was especially salient in the one week follow-up.This supports theories of attitude construction involving the effects of elaboration and information processing on attitude change, such as the ELM (Petty, Haugtvedt, and Smith, 1995). For example, Barden and Tormala (2014) suggested that more elaboration produced better judgments, and that these in turn could be held with greater confidence. It has also been demonstrated that a person's attitude can strengthen if she perceives herself successfully defending it from persuasive attempts (Tormala and Petty, 2002; Knowles and Linn, 2004). These theories suggest that the perceived adjustment to contextual factors generate inferences about the metacognitive evaluation of the attitude. Following a similar logic, we could argue that observing yourself gives coherent arguments for a manipulated response, actually strengthens the induced attitude in the manipulated direction. And the more arguments supporting the induced attitude you hear yourself expressing, the more evidence you have to feel confident that this is what you truly believe.

In summary, our findings support constructivist accounts of attitude formation through a process of self-perception. Further, our findings support models predicting elaboration as underpinning strong attitude shifts. Together these results demonstrate a powerful influence of feedback in the moment with self-persuasion on attitudes in a domain of central importance to the functioning of democratic public life.

Acknowledgments

Many thanks to Anders Lindén at Andvision for implementing the design of the STS. PJ was supported by Bank of Sweden Tercentenary Foundation and The Swedish Research Council (2014-1371). LH was supported by Bank of Sweden Tercentenary Foundation (P13-1059:1) and The Swedish Research Council (2011-1795).

References

Ajzen, I., & Gilbert Cote, N. (2014). Attitudes and the prediction of behavior. In Crano, W. D., and Prislin, R. (Eds.), *Attitudes and attitude change* (pp. 289-311). New York, NY: Psychology press.

Ariely, D., & Norton, M. I. (2008). How actions create – not just reveal – preferences. *Trends in Cognitive Sciences*, *12*(1), 13–16.

Barden, J., & Tormala, Z. L. (2014). Elaboration and attitude strength: The new meta-cognitive perspective. *Social and personality psychology compass*, 8/1, (pp. 17-29).

Bem, D. J. (1967). Self-perception: An alternative interpretation of cognitive dissonance phenomena. *Psychological Review*, *74*(3), 183–200.

Bishop, G. F. (2005). *The illusion of public opinion: Fact and artifact in American public opinion polls*. Lanham; MD: Rowman and Littlefield publishers inc.

Brehm, J. W. (1956). Postdecision changes in the desirability of alternatives. *The Journal of Abnormal and Social Psychology*, *52*(3), 384–389.

Clarkson, J. J., Tormala, Z. L., & Leone, C. (2011). A self-validation perspective on the mere thought effect. *Journal of Experimental Social Psychology*, *47*(2), 449–454.

Druckman, J. N. (2004). Political Preference Formation: Competition, Deliberation, and the (Ir)relevance of Framing Effects. *American Political Science Review*, *98*(04), 671-686.

Eagly, A. H., & Chaiken, S. (1993). *The psychology of attitudes*. Harcourt Brace Jovanovich College Publishers.

Fotopoulou, A., Conway, M. A., & Solms, M. (2007). Confabulation: Motivated reality monitoring. *Neuropsychologia*, *45*(10), 2180–2190.

Hall, L., Johansson, P., & Strandberg, T. (2012) Lifting the Veil of Morality: Choice Blindness and Attitude Reversals on a Self-Transforming Survey. *PLoS One*, *7*(9), e45457.

Hall, L., Strandberg, T., Pärnamets, P., Lind, A., Tärning, B., & Johansson, P. (2013). How the polls can be both spot on and dead wrong: Using choice blindness to shift political attitudes and voter intentions. *PLoS One*, *8*(4), e60554.

Hirstein, W. (2010). The misidentification syndromes as mindreading disorders. *Cognitive neuropsychiatry*, *15*(1-3), 233-260.

Jost, J. T., Federico, C. M., & Napier, J. L. (2009). Political ideology: Its Structure, Functions, and Elective Affinities. *Annual review of psychology*, *60*, 307-337.

Johansson, P., Hall, L., Sikström, S., & Olsson, A. (2005). Failure to detect mismatches between intention and outcome in a simple decision task. *Science*, *310*(5745), 116–119.

Johansson, P., Hall, L., Tärning, B., Sikström, S., & Chater, N. (2014). Choice blindness and preference change. *Journal of Behavioral Decision Making*, *27*(3), 281–289.

Petty, R. E., & Cacioppo, J. T. (1986). The elaboration likelihood model of persuasion. *Advances in Experimental Social Psychology*, *19*, 123–205.

Petty, R. E., Haugtvedt, C. P., & Smith, S. M. (1995). Elaboration as a determinant of attitude strength: *Attitude strength*, *4*, 93–130.

Pärnamets, P., Hall, L., & Johansson, P. (2015). Memory distortions resulting from a choice blindness task. In *Proceedings of the 37th Annual Conference of the Cognitive Science Society*

Sharot, T., Fleming, SM., Yu, X., Koster, R., & Dolan, R. J. (2012). Is choice-induced preference change long lasting? *Psychological Science 23*(10), 1123–1129.

Taya, F., Gupta, S., Farber, I., & Mullette-Gillman, O. A. (2014). Manipulation Detection and Preference Alterations in a Choice Blindness Paradigm. *PLoS One 9*(9), e108515.

Tormala, Z. L., & Petty, R. E. (2002). What doesn't kill me only makes me stronger: The effects of resisting persuasion on attitude certainty. *Journal of personality and social psychology*, vol. 83, no. 6, 1298-1313.