Online Appendix

Chapter 2

The effect of boosting in-group identities on tolerance of false facts

Christine Stedtnitz

University of Essex

cmsted@essex.ac.uk

September 30, 2019

I. Appendix I – Study 1: further results

The post-treatment questionnaire included three items to measure to what extent players noticed, and believed in the false claims. The results reported in the main section of the paper focus on one of them: belief that the author of a feedback was 'a good representative of my team'. This appendix includes results pertaining to the other two indicators of overlooking false claims: general agreement ('Generally speaking, do you agree with the author of this feedback?', where 1='Strongly Disagree', 2='Disagree', 3='Slightly Disagree', 4='Slightly Agree', 5='Agree', 6='Strongly Agree'), and perceived accuracy ('The points this person makes are factually accurate', measured on a scale from 0 (not at all) to 100 (very)).

i. Hypothesis 1

Hypothesis 1 held that in-group members overlook false false in the feedback that sides with their in-group: H1a is about the disadvantaged team:

- H1a: Among the disadvantaged team, there is no significant difference between *representation ratings* of a false 'unfair' feedback and a correct 'unfair' feedback.

- H1a-2: Among the disadvantaged team, there is no significant difference between *general agreement* with a false 'unfair' feedback and agreement with a correct 'unfair' feedback.
- H1a-3: Among the disadvantaged team, there is no significant difference between *accuracy ratings* of a false 'unfair' feedback and a correct 'unfair'feedback.

As hypothesized, both the disadvantaged and the advantaged teams rated the person who sided with their team, respectively, as a good representative – regardless of whether or not that person made a false claim (see figure 2a). These null findings extend to the other indicators of overlooking false claims: The disadvantaged were just as agreed with the correct version of the 'unfair' feedback as with the incorrect version of it: $M_{\text{false 'unfair'}}$=4.6/6, $M_{\text{correct 'unfair'}}$=4.7/6, t(130)=-0.53, p = 0.59. They rated both as 'factually accurate': ($M_{\text{false 'unfair'}}$=65/100, $M_{\text{correct 'unfair'}}$=68/100), t(126)=-0.91, p = 0.36).

Hypotheses for the advantaged team are analogous, and also confirmed (see graphs on page 2b):

- H1b: Among the advantaged team, there is no significant difference between *representation ratings* of a false 'fair play' feedback and a correct 'fair play'feedback.
- H1b-2: Among the advantaged team, there is no significant difference between *general agreement* with a false 'fair play' feedback and a correct 'fair play' feedback.
- H1b-3: Among the disadvantaged team, there is no significant difference between *accuracy ratings* of a false 'fair play' feedback and a correct 'fair play' feedback.

As hypothesized, the advantaged team turned a blind eye to false facts in the 'fair play' feedback. Accurate or not; they were 'slightly agreed' ($M_{\text{false 'fair play'}}$=4.1/6, $M_{\text{correct 'fair play'}}$=4.2/6, t(126)=-0.47, p = 0.63) and rated it around 60/100 on factual accuracy ($M_{\text{false 'fair play'}}$=60/100, $M_{\text{correct 'fair play'}}$=61/100), t(129)=-0.19, p = 0.84).

ii. Hypothesis 2

Hypothesis 2 held that in-group members notice false false in the feedback that sides with the out-group. H2a concerns the disadvantaged team rating the 'fair play' feedback:

- H2a: Among the disadvantaged team, the author of the false 'fair play' feedback is seen as an (even) worse team representative than the author of the correct 'fair play' feedback.

- H2a-2: Among the disadvantaged team, average agreement with the false 'fair play' feedback is (even) lower than average agreement with the correct 'fair play' feedback.
- H2a-3: Among the disadvantaged team, the author of the false 'fair play' feedback is seen as (even) less accurate than the author of the false 'fair play' feedback.

As shown in figure 3a and in the main section of this paper our data does not confirm H2a. The disadvantaged team rated the author of the the 'fair play' feedback that contained false claims as a slightly, but not significantly worse team representative than the author of the accurate version of the 'fair play' feedback. Yet in this case, the other two dependent variables show stronger evidence of noticing false claims in the other camp: Among disadvantaged players who saw false facts in their 'fair play' feedback, general agreement with this 'fair play' feedback averaged around 3.1, that is closest to 3='Slightly Disagree'. This is significantly lower than general agreement among those who saw the false fact-free copy of the 'fair play' feedback: Their ratings averaged around 3.7, that is, closest to 4='Slightly Agree' ($t(132)=-2.32$, $p = 0.02$). H2a-2 is therefore confirmed. Accuracy ratings of the false 'fair play' feedback were 7 points lower, on average (and much closer to the mid-point of the scale) than accuracy ratings of the correct 'fair play' feedback: $M_{\text{false 'fair play'}}=52/100$, $M_{\text{correct 'fair play'}}=59/100$, $t(133)=-1.82$, $p = 0.07$. As this is just short of significance H2a-b cannot be confirmed. Nonetheless, it ought to be seen as suggestive evidence that the disadvantaged team players were more likely to notice the false claims if they came from the other camp.

H2b is about the advantaged team rating the 'unfair' feedback:

- H2b: Among the advantaged team, the author of the false 'unfair' feedback is seen as an (even) worse team representative than the author of the correct 'unfair' feedback.
- H2b-2: Among the advantaged team, average agreement with the false 'unfair' feedback is (even) lower than average agreement with the correct 'unfair' feedback.
- H2b-3 Among the advantaged team, the author of the false 'unfair' feedback is seen as (even) less accurate than the author of the false 'unfair' feedback.

This study failed to find evidence for all three hypotheses: There was no difference whatsoever in how much the advantaged team supported the accurate and the inaccurate versions of the 'unfair' feedback. As discussed in the main text, both authors are seen as equally good team representatives. The same is true for the other two indicators: The advantaged team agreed 'sightly'

with the accurate and the inaccurate 'unfair' feedback ($M_{\text{false 'unfair'}}$=4.0, $M_{\text{correct 'unfair'}}$=4.0). When it came to rating how 'factually accurate' the 'unfair' feedback was, the advantaged team rated it as 60/100 – regardless of whether or not it contained any false claims ($M_{\text{false 'unfair'}}$=61/100, $M_{\text{correct 'unfair'}}$=60/100)).

A few exploratory OLS analyses reflect these findings: Models 1-3 in table 1 show players' opinions about the author of the 'unfair' feedback; models 4-6 shows opinions about the author of the 'fair play' feedback. The models show three dependent variables: general agreement (models 1 and 4), perceived accuracy (2 and 5), and perceived suitability as a team representative (3 and 6). Status – that is, team membership (advantaged / disadvantaged) – had a significant effect on all ratings except for accuracy ratings for the 'fair play' person. Factual accuracy – whether or not the feedback contained false claims – had a significant effect in model 3, where the dependent variable was how players rated the author of the 'unfair' feedback as a potential team representative.)

iii. Hypothesis 3

Hypothesis 3 held that in-group members who see false claims in the feedback that sided with their team overlook those false claims.

H3a-c are about members of the disadvantaged team who see a false 'unfair' feedback and a correct 'fair play' feedback (group 4):

- H3a: Members of the disadvantaged team who see a false 'unfair' feedback and a correct 'fair play' feedback rate the author of the (false) 'unfair' feedback as a better team representative than the author of the (correct) 'fair play' feedback.
- H3b: Members of the disadvantaged team who see a false 'unfair' feedback and a correct 'fair play' feedback are more agreed with the (false) 'unfair' feedback than the (correct) 'fair play' feedback.
- H3c: Members of the disadvantaged team who see a false 'unfair' feedback and a correct 'fair play' feedback rate the (false) 'unfair' feedback as more accurate than the (correct) 'fair play' feedback.

Indeed, the groups that were nudged to engage in motivated reasoning overlooked false facts in the feedback that sided with their team: The disadvantaged rated the author of the 'unfair' feedback that contained false claims as a much better team representative than the author of the

'fair play' feedback that did not contain any false claims. They 'agreed' with the former, on average, but agreed only 'slightly' with latter ($M_{\text{false 'unfair'}}$=4.6/6, $M_{\text{correct 'fair play'}}$=3.5/6, t(112)=4.43, p = 0.00). Hence, H3b is confirmed. However, when considering how accurate the two people's feedback was, group differences did not reach statistical significance: On a scale from 0 to 100 where 0 meant 'not accurate' and 100 meant 'accurate' the disadvantaged team gave the false 'unfair' feedback an average rating of 66/100 – only seven points higher than the rating they gave the correct 'fair play' feedback ($M_{\text{false 'unfair'}}$=59/100, $M_{\text{correct 'fair play'}}$=66/100, t(120)=-1.48, p = 0.14).

H3d-f looked at advantaged players who were nudged to be motivated reasoners, i.e. who saw false claims in the 'fair play' feedback, but not in the 'unfair' feedback (group 1):

- H3d: Members of the advantaged team who saw false claims in the 'fair play' feedback rate the author of the (false) 'fair play' feedback as a better team representative than the author of the (correct) 'unfair' feedback.
- H3e: Members of the advantaged team who saw false claims in the 'fair play' feedback are more agreed with the author of the (false) 'fair play' feedback than the author of the (correct) 'unfair' feedback.
- H3f: Members of the advantaged team who saw false claims in the 'fair play' feedback rate the (false) 'fair play' feedback as more accurate than the (correct) 'unfair' feedback.

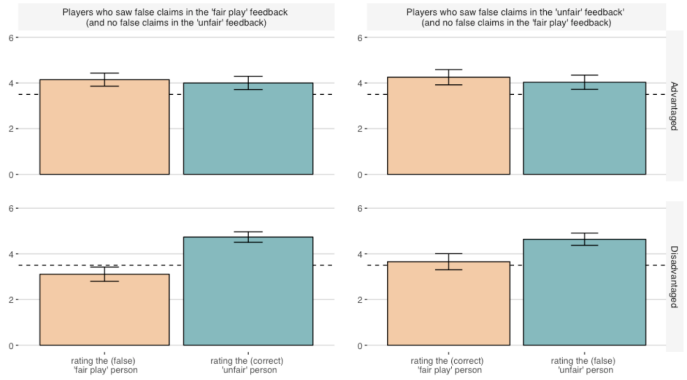
This study found evidence for H3d. H3e and f, however, are rejected: The advantaged were 'slightly agreed' with both the false 'fair play' feedback and the incorrect 'unfair' feedback ($M_{\text{false 'fair play'}}$=4.1/6, $M_{\text{correct 'unfair'}}$=4.0/6, t(136)=0.71, p = 0.48). Accuracy ratings did not differ at all: $M_{\text{false 'fair play'}}$=59/100, $M_{\text{correct 'unfair'}}$=60/100, t(135)=-1.20, p = 0.84). However, as discussed above, findings for the advantaged team are to be interpreted with caution.

II. Graphs – Full Results

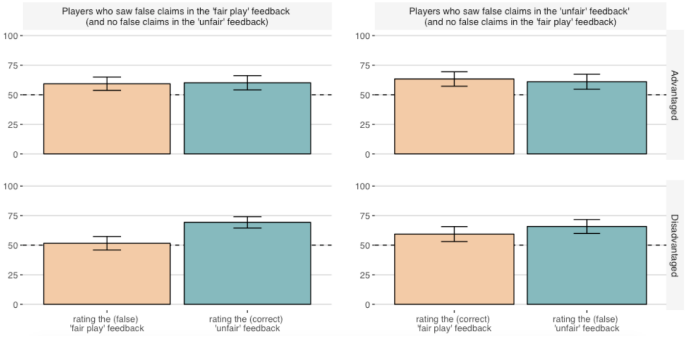
The graphs in this online appendix are divided into two sections: Figures 1a, 1b, 1c show the full results, i.e. average agreement, accuracy and representation ratings for all players. To facilitate comparisons the next sections splits up ratings by group. (It tests the three hypotheses outlined in the paper as well as two additional hypotheses that were left out of the main section.)

Figures 1a, 1b, 1c show how both teams rated the two feedback givers, depending on which of the two made false claims. The top row shows the advantaged team (Team A); the bottom row shows the disadvantaged team (Team B). The left-hand column shows respondents who rated a 'fair play' feedback containing false claims (and an 'unfair' feedback that did not). The right-hand column shows respondents who rated an 'unfair' feedback containing false claims (and a 'fair play' feedback that did not). Beige bars show how the 'fair play' person was rated; teal-coloured bars show how the 'unfair' person was rated. Opinions were measured on a scale from 0 (not at all) to 100 (very). Hence, higher scores indicate a better team representative.

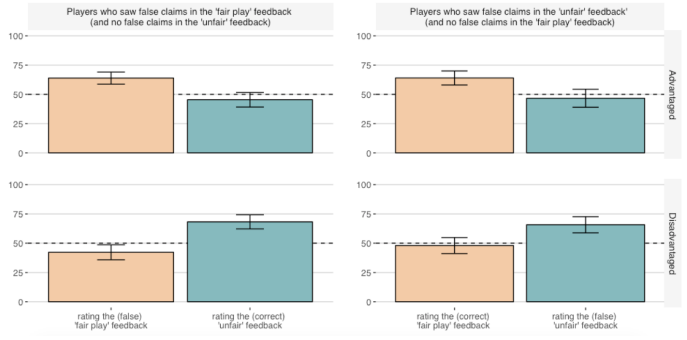
III. Graphs – Group Comparisons



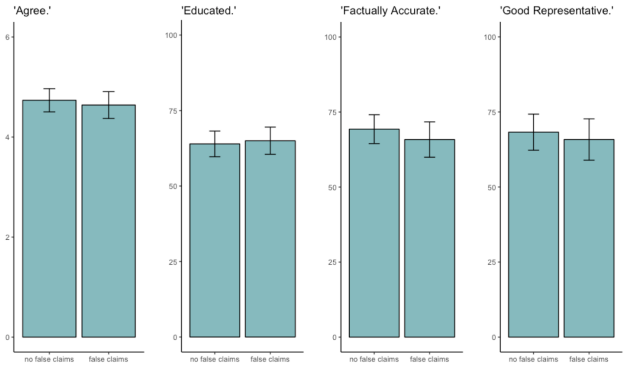
(a) *'Generally speaking, do you agree with the author of this feedback?' (1-strongly disagree to 6-strongly agree'*



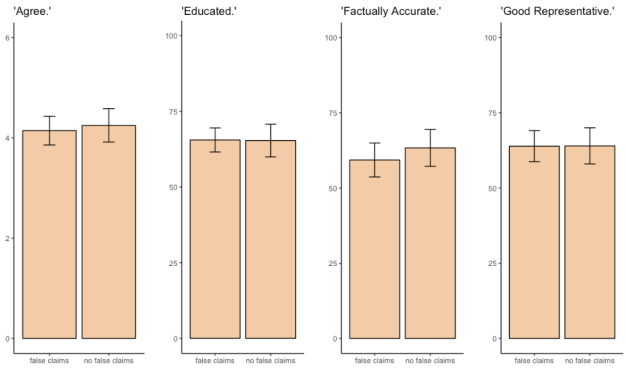
(b) *'The points this person makes are factually accurate.'*



(c) *'This person is a good representative of my team.'*

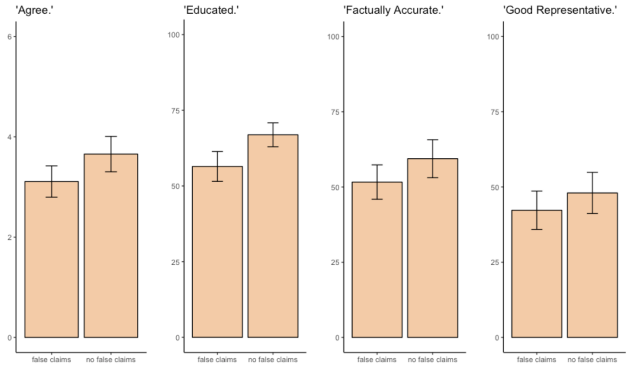


(a) *Disadvantaged team rating the 'unfair' feedback (bars on left-hand side: no false claims; right: false claims) H1a: Disadvantaged team members who see a factually inaccurate 'unfair' feedback rate the author of the 'unfair' feedback as just as just as accurate // just as good a team representative // are just as agreed as their peers who see a factually accurate 'unfair' feedback.*

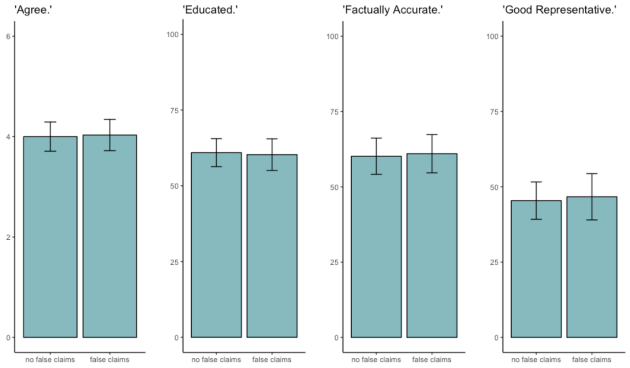


(b) *Advantaged team rating the 'fair play' feedback H1b: Advantaged team members who see a factually inaccurate 'fair play' feedback rate the author of the 'fair play' feedback as just as just as accurate // just as good a team representative // are just as agreed as their peers who see a factually accurate 'fair play' feedback.*

Figure 2: *Hypothesis 1 – There is no difference in how players rate the accurate and the inaccurate version of the feedback that sides with their team*

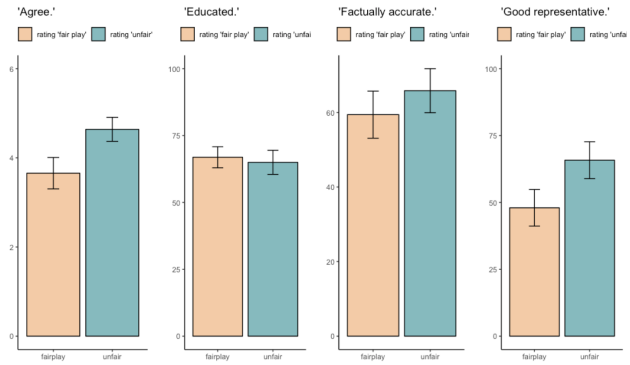


(a) *Disadvantaged team rating the 'fair play' feedback*
H2a: Disadvantaged team members who see a 'fair play' feedback that contains false claims rate the author of that feedback as (even) less accurate // an (even) worse team representative // and are (even) less agreed with it than their peers who see a 'fair play' feedback that does not contain any false claims.

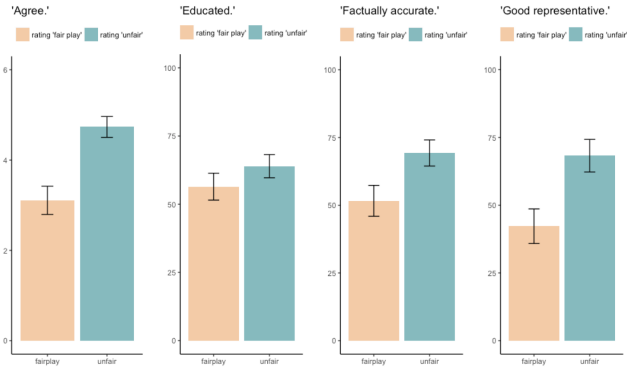


(b) *Advantaged team rating the 'unfair' feedback*
H2b: Advantaged team members who see an 'unfair' feedback that contains false claims rate the author of that feedback as (even) less accurate // an (even) worse team representative // and are (even) less agreed with it than their peers who see an 'unfair' feedback that does not contain any false claims.

Figure 3: *Hypothesis 2 – Players notice and punish false claims in the feedback that sides with the other team*

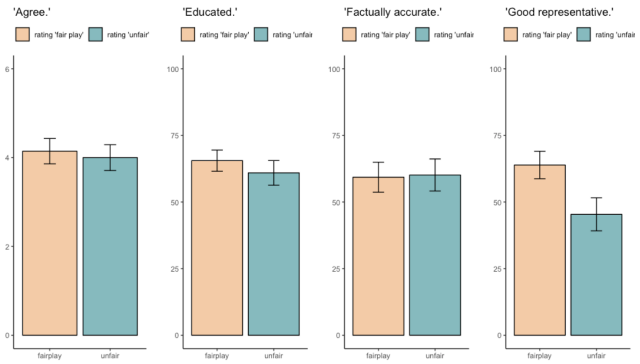


(a) *Disadvantaged players who saw false claims in the 'unfair' feedback (group 4) rating the (false) 'unfair' feedback and the (correct) 'fair play' feedback*
H3a: Group 4 is more agreed with the (false) 'unfair' feedback // rates its author as more accurate // and a better team representative than the author of the (correct) 'fair play' feedback.

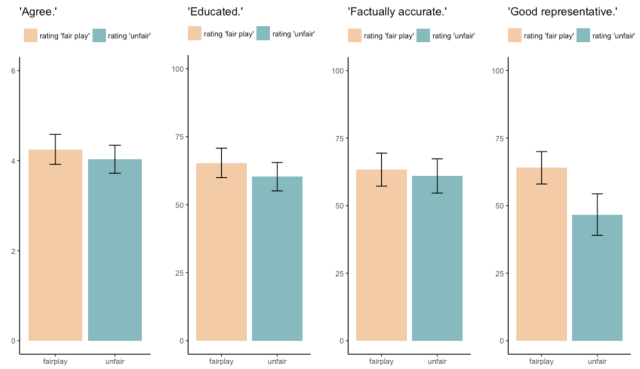


(b) *For comparison: Disadvantaged players who saw false claims in the 'fair play' feedback (group 3) rating the (correct) 'unfair' feedback and the (false) 'fair play' feedback*

Figure 4: *Hypothesis 3 – When exposed to a false feedback on their side and an accurate feedback on the other side players overlook the false claims and rate the former better than the latter (here: disadvantaged players)*

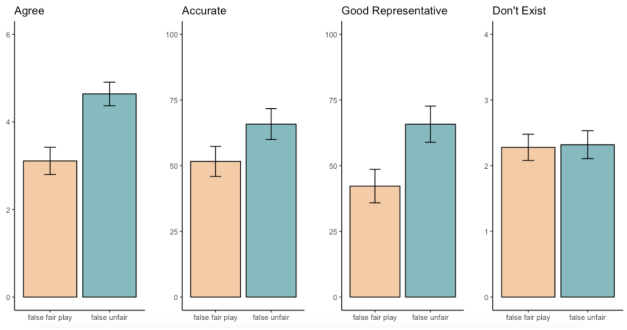


(a) *Advantaged players who saw false claims in the 'fair play' feedback (group 1) rating the (false) 'fair play' feedback and the (correct) 'unfair' feedback*
H3b Group 1 is more agreed with the (false) 'fair play' feedback // rate its author as more accurate // and a better team representative than the author of the (correct) 'unfair' feedback.

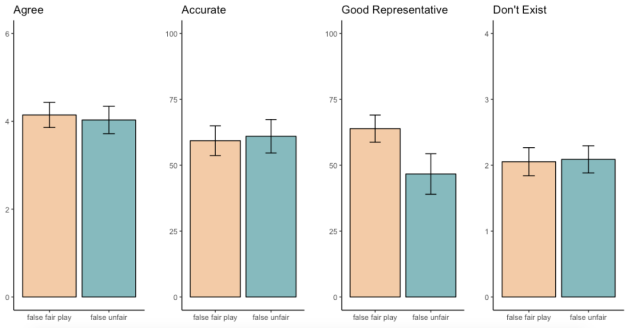


(b) *Advantaged players who saw false claims in the 'unfair' feedback (group 2) rating the (correct) 'fair play' feedback and the (false) 'unfair' feedback*

Figure 5: *Hypothesis 3 – When exposed to a false feedback on their side and an accurate feedback on the other side players overlook the false claims and rate the former better than the latter (here: advantaged players)*

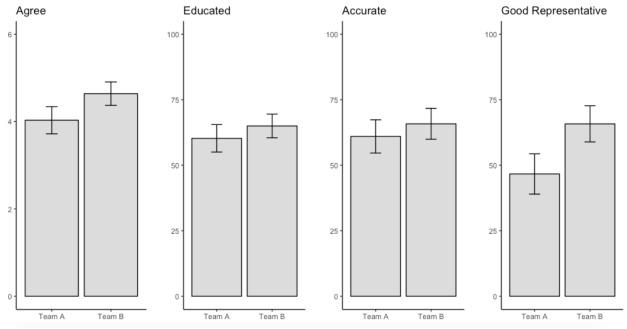


(a) *Team B rating the 2 false feedback pages*
– H4a: Team B overlooks false claims in the 'unfair' feedback more than they overlook false claims in the 'fair play' feedback –

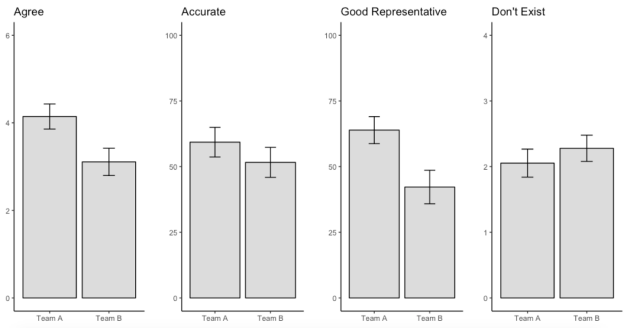


(b) *Team A rating the 2 false feedback pages*
– H4b: Team A overlooks false claims in the 'fair play' feedback more than they overlook false claims in the 'unfair' feedback –

Figure 6: *Hypothesis 4 – Players rate the false feedback on their side better than they rate the false feedback on the other side.*



(a) *Both teams rating the incorrect 'unfair' feedback – H5a: False facts in the 'unfair' treatment are more likely to go unnoticed among the disadvantaged than among the advantaged –*



(b) *Teams A and B rating the incorrect 'fair play' feedback – H5b: False facts in the 'fair play' treatment are more likely to go unnoticed among the advantaged than among the disadvantaged. –*

Figure 7: *Hypothesis 5 – False feedback are likely to go unnoticed among the team the feedback sides with than among the other team.*

Screenshots of the laboratory experiment

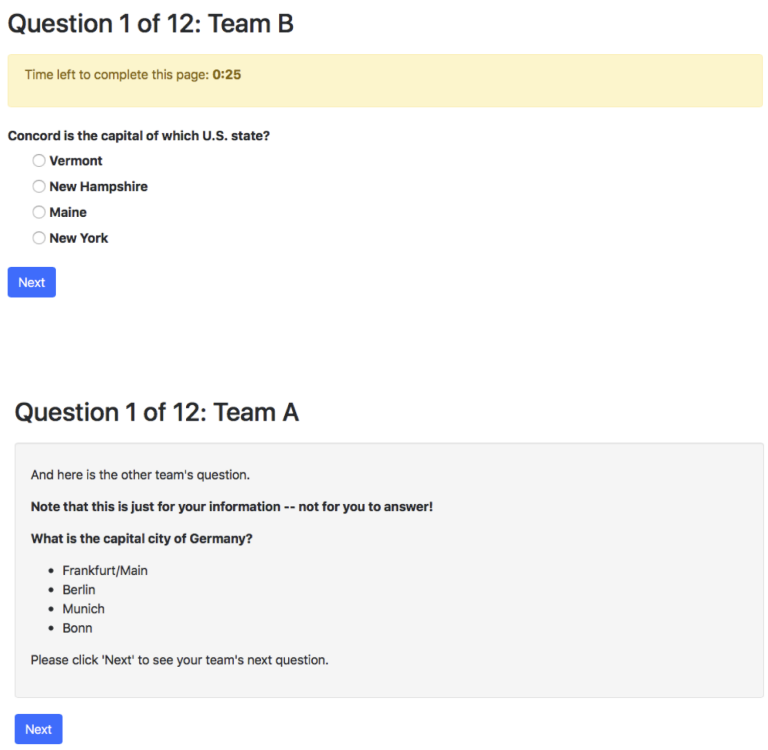


Figure 8: *Example quiz questions (as shown to a Team B player)*

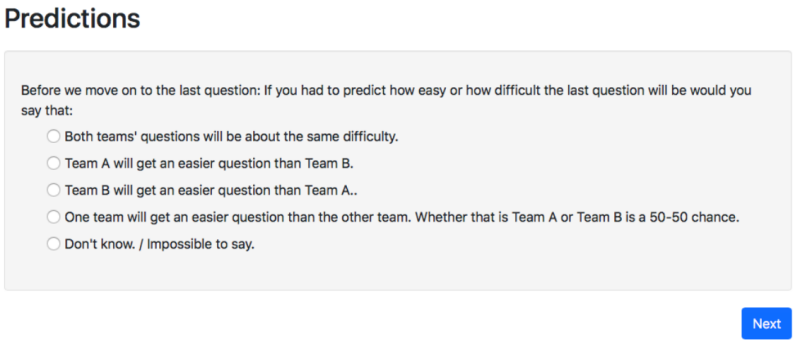
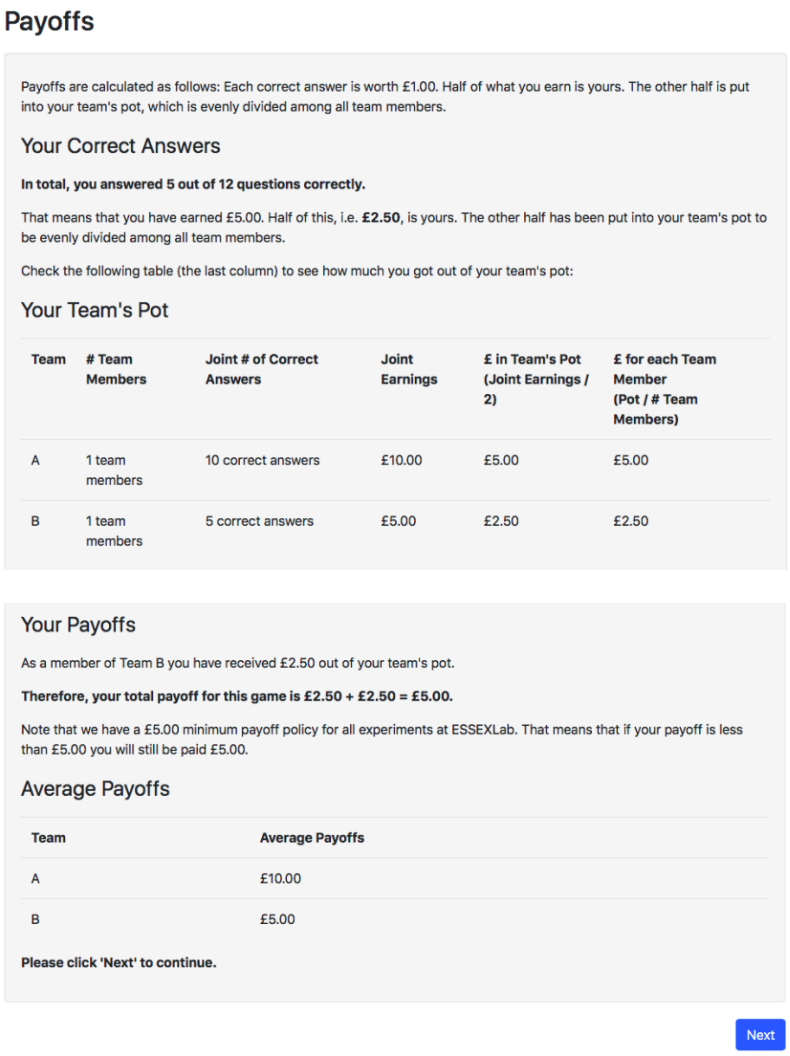


Figure 9: *Perceived bias (shown before the last question)*

IV. Regression Results

Table 1: *OLS Results: 'This person is a good representative of my team. // The points this person makes are factually accurate.'*

	Dependent variable:					
	'unfair': agreed	'unfair': accurate	'unfair': good rep	'fair play': agreed	'fair play': accurate	'fai
	(1)	(2)	(3)	(4)	(5)	
Disadvantaged	0.599***	9.813**	25.468***	−0.614**	−2.254	
	(0.203)	(4.661)	(5.162)	(0.252)	(4.748)	
False unfair	0.074	2.532	8.461	0.277	7.092	
	(0.208)	(4.767)	(5.280)	(0.258)	(4.856)	
Difference in payoffs	0.111**	2.015**	0.304	−0.081	1.344	
	(0.044)	(1.010)	(1.118)	(0.055)	(1.029)	
Luck of the draw	0.006**	0.028	0.222***	−0.010***	−0.048	
	(0.002)	(0.053)	(0.059)	(0.003)	(0.054)	
Payoffs not legitimate	0.522***	8.878**	8.094**	−0.720***	−14.662***	
	(0.150)	(3.450)	(3.821)	(0.186)	(3.514)	
Failed attention check	0.343**	4.851	7.897**	0.024	0.315	
	(0.155)	(3.559)	(3.942)	(0.192)	(3.626)	
Female	−0.124	−5.065	−5.222	0.310*	9.861***	
	(0.149)	(3.422)	(3.790)	(0.185)	(3.486)	
Income	−0.043*	−0.434	−0.744	0.002	0.211	
	(0.024)	(0.553)	(0.613)	(0.030)	(0.564)	
Disadvantaged * False unfair	−0.226	−8.829	−13.621*	0.178	−0.971	
	(0.286)	(6.572)	(7.280)	(0.355)	(6.696)	
Constant	3.195***	50.856***	28.287***	5.009***	53.547***	
	(0.286)	(6.578)	(7.285)	(0.355)	(6.701)	
Observations	216	216	216	216	216	
R^2	0.251	0.121	0.265	0.261	0.151	
Adjusted R^2	0.218	0.083	0.233	0.228	0.114	
Note:						*p<0.1; **p



Payoffs

Payoffs are calculated as follows: Each correct answer is worth £1.00. Half of what you earn is yours. The other half is put into your team's pot, which is evenly divided among all team members.

Your Correct Answers

In total, you answered 5 out of 12 questions correctly.

That means that you have earned £5.00. Half of this, i.e. **£2.50**, is yours. The other half has been put into your team's pot to be evenly divided among all team members.

Check the following table (the last column) to see how much you got out of your team's pot:

Your Team's Pot

Team	# Team Members	Joint # of Correct Answers	Joint Earnings	£ in Team's Pot (Joint Earnings / 2)	£ for each Team Member (Pot / # Team Members)
A	1 team members	10 correct answers	£10.00	£5.00	£5.00
B	1 team members	5 correct answers	£5.00	£2.50	£2.50

Your Payoffs

As a member of Team B you have received £2.50 out of your team's pot.

Therefore, your total payoff for this game is £2.50 + £2.50 = £5.00.

Note that we have a £5.00 minimum payoff policy for all experiments at ESSEXLab. That means that if your payoff is less than £5.00 you will still be paid £5.00.

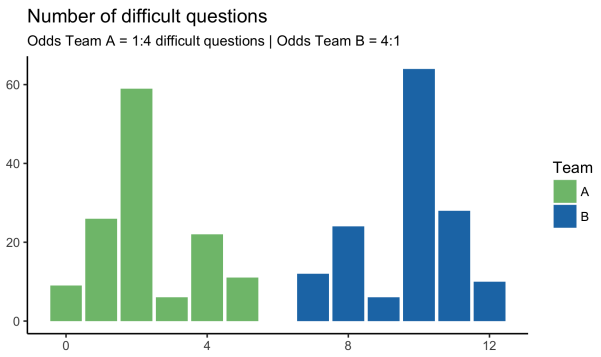
Average Payoffs

Team	Average Payoffs
A	£10.00
B	£5.00

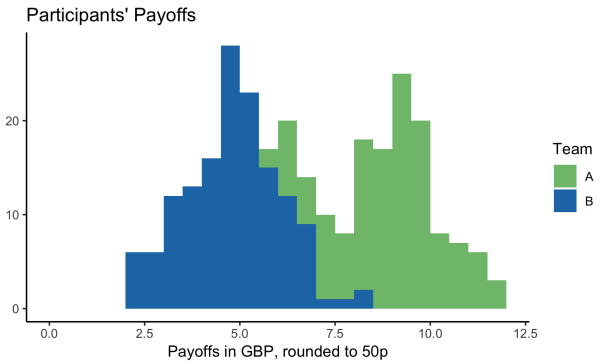
Please click 'Next' to continue.

Next

Figure 10: *Payoffs*



(a) *Number of difficult questions by team*



(b) *Payoffs by team*

Figure 11: *Number of difficult questions and payoffs for both teams*

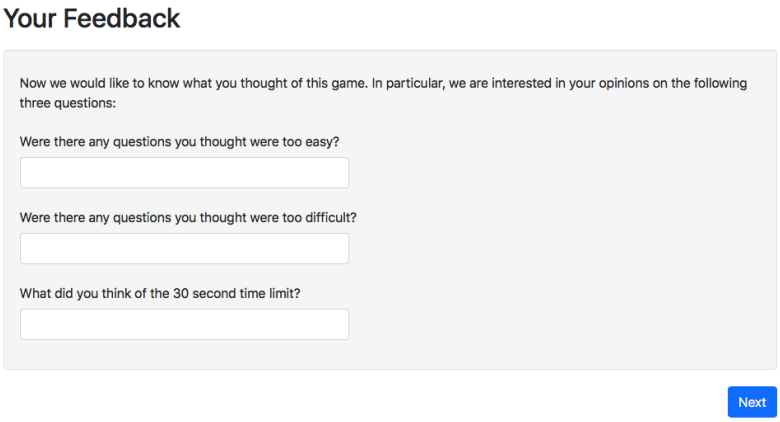
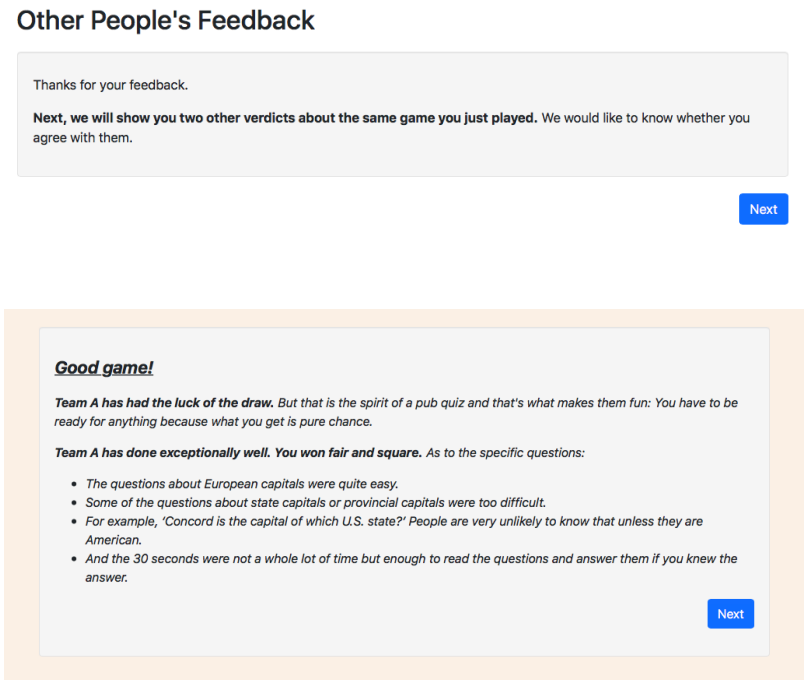
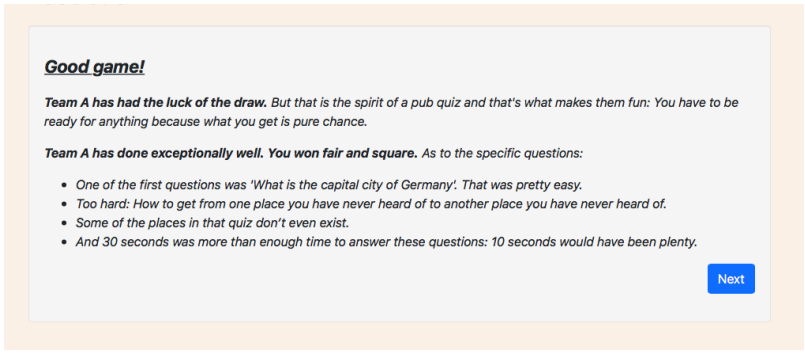


Figure 12: *Respondents' feedback*

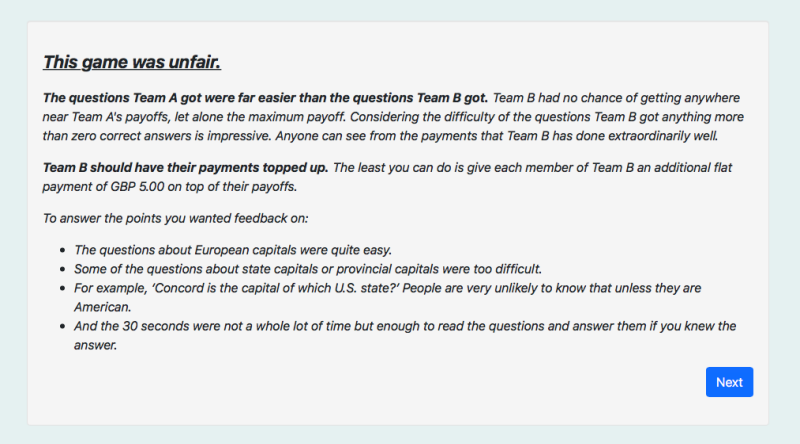


(a) *Not including any false claims*

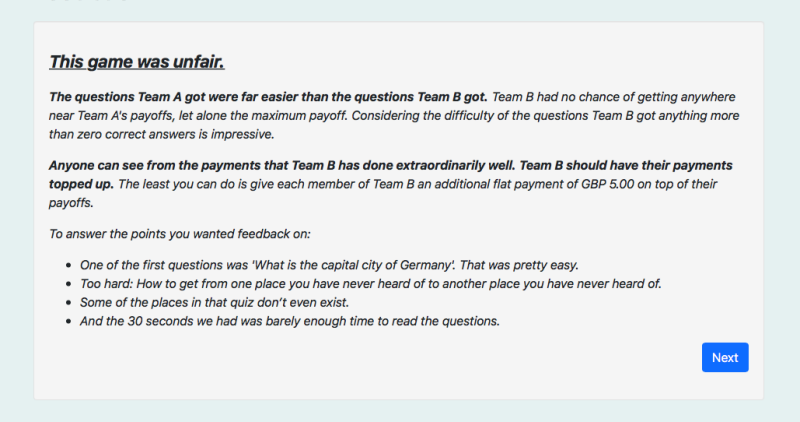


(b) *Including false claims*

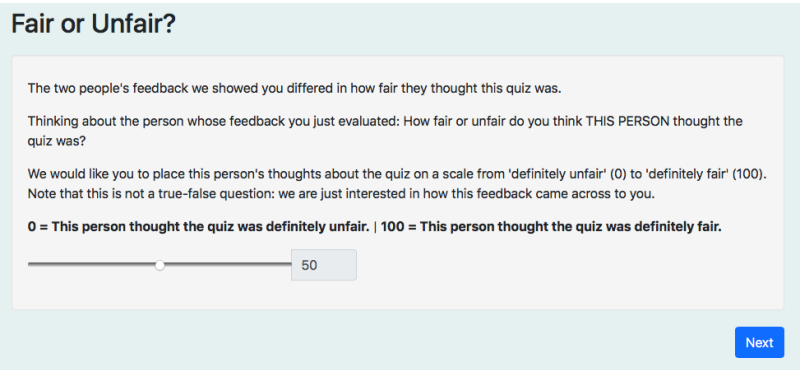
Figure 13: *'Fair Play' Feedback*



(a) *Not including any false claims*



(b) *Including false claims*



(c) *Attention Check*

Figure 14: *'Unfair' Feedback*

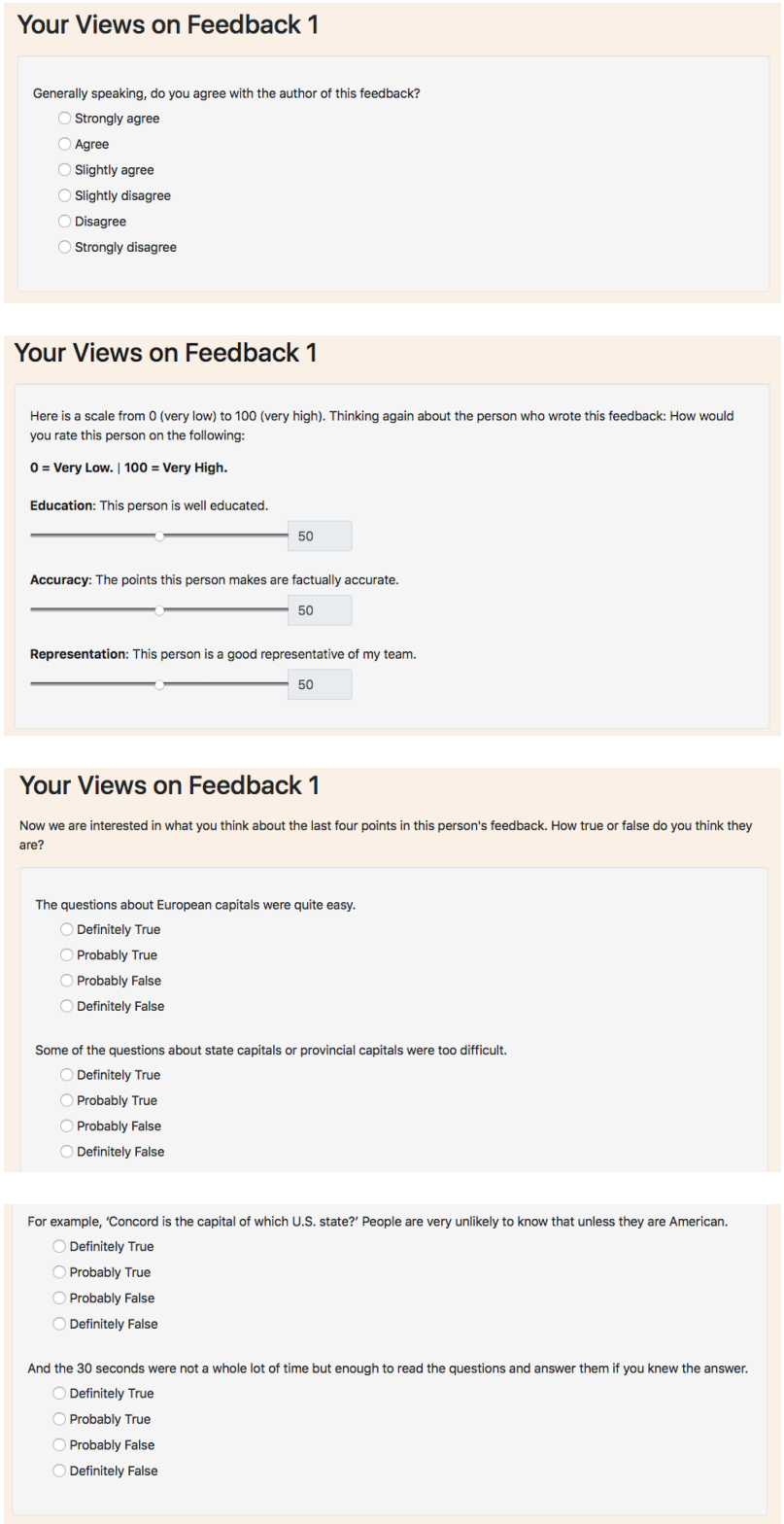


Figure 15: *DVs – 'Fair Play' Feedback*

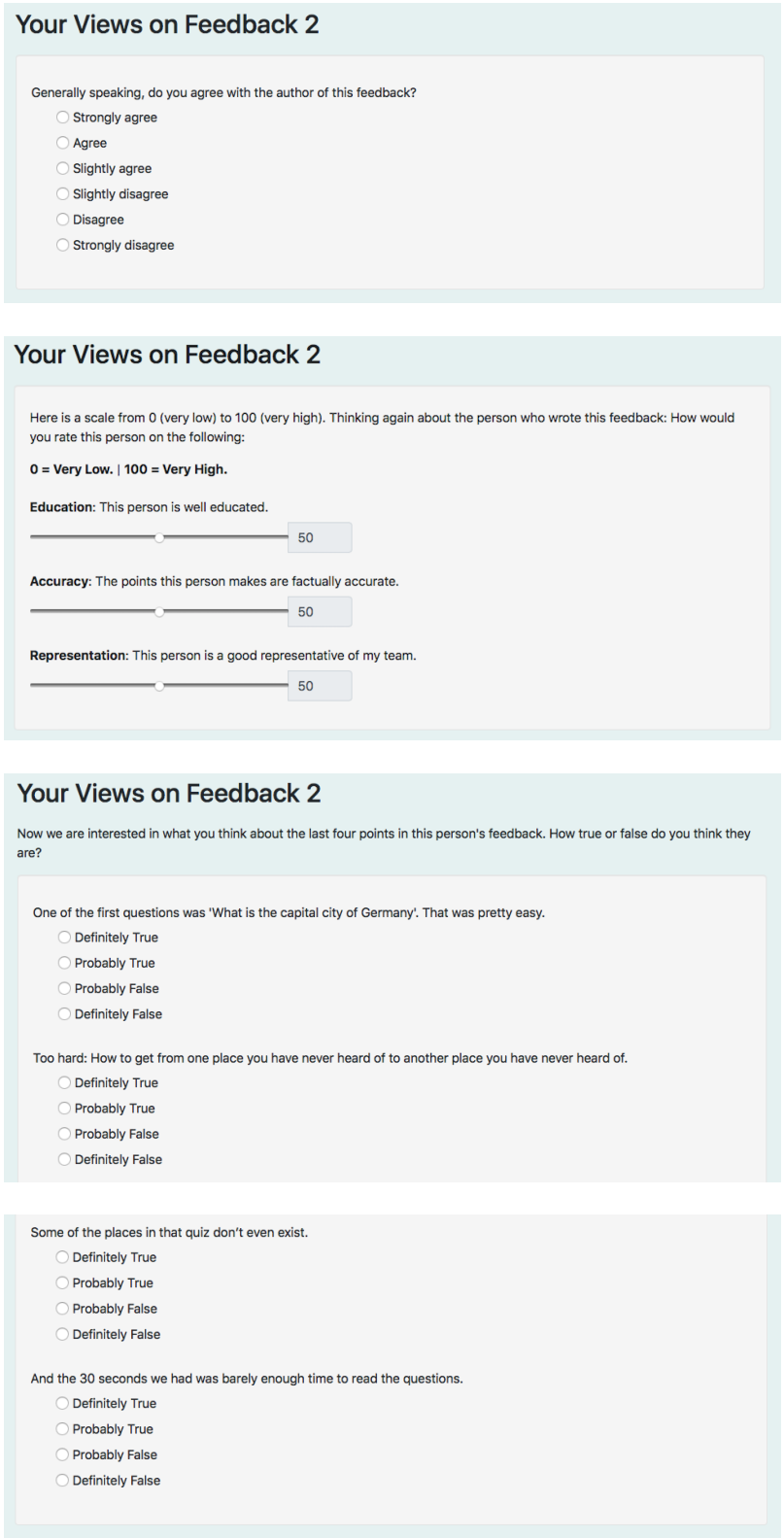


Figure 16: *DVs – 'Unfair' Feedback*