DEPARTMENT OF GOVERNMENT

UNIVERSITY OF ESSEX

Doctoral Thesis

How motivated reasoning leads to tolerance of false claims: Three experimental tests of mechanisms

---

Author:

Christine STEDTNITZ

Supervisors:

Rob JOHNS

Patrick LOWN

A thesis submitted for the degree of Doctor of Philosophy

September 30, 2019

Abstract

Christine STEDTNITZ

How motivated reasoning leads to tolerance of false claims: Three experimental tests of mechanisms

Many recent electoral events have been characterised by false claims which, despite abundant fact-checking, were often widely believed. This led to much talk about a 'post-truth' politics. Meanwhile, an extensive literature confirms that political misperceptions are often highly resistant to correction. The drivers of that resistance are well-known: citizens are prone to accept as fact those claims that confirm their prior beliefs and attitudes, and dismiss claims that challenge them.

This dissertation investigates reactions to challenging political information. Chapters 1 and 2 explore mechanisms that might affect reactions to false facts that affirm prior attitudes: a stressful environment (chapter 1) and a low group standing (chapter 2). Chapter 3 explores a mechanism that might affect reactions to fact-checks that disconfirm prior opinions: post-truth surroundings.

Chapter 1 is a survey experiment that (unsuccessfully) attempts to manipulate stress to examine its effect on tolerance of false facts in the Brexit

debate.

Chapter 2 is a laboratory experiment that uses a rigged pub quiz to create status differences. It investigates the effect of these status differences on reactions to false claims coming from a feedback giver who favours players' own team.

Chapter 3 is a survey experiment that identifies false beliefs about immigration (liberal and conservative) and fact-checks them. It explores if subsequent exposure to post-truth comments encouraging respondents to disregard expert advice affects increases the likelihood that respondents reject expert advice.

Contents

Abstract **iii**

1 Introduction **1**

2 The Effect of Stress on Belief in False Facts in the Brexit Debate **15**

2.1 Abstract 15

2.2 Introduction 16

2.3 Method 21

2.4 Results 26

2.5 Discussion and Conclusion 34

3 The effect of boosting in-group identities on tolerance of false facts **37**

3.1 Abstract 37

3.2 Introduction 38

3.3 Theoretical Background 40

3.3.1 Global Status Threat 44

3.3.2 Ethnic status threat 45

3.3.3 Status Threat and Vote Choice 47

3.3.4 Proposed mechanism 49

3.3.5 Method 52

3.3.6 Hypotheses 55
3.3.7 Results 58
3.4 Discussion 65
3.5 Conclusion 70
3.6 Acknowledgements 73

4 A Post-Truth Public? Investigating the Mechanisms of Resistance to Factual Correction 75

4.1 Abstract 75
4.2 Introduction 76
4.3 Introduction 78
4.4 Reactions to uncongenial expert information 80
4.5 Method 85
4.6 Hypotheses 88
Effect of the fact-check on belief in false facts 89
Effect of post-truth comment on reactions to the fact-check 89
Spillover effects 90
Evidence of post-truth reasoning 92
4.7 Results 93
Effect of the correction on belief in false claims . . . 96
Effect of post-truth comments 98
Spillover effects 101
Evidence of post-truth reasoning 104
Limitations 106
4.8 Discussion and Conclusion 108

5 Conclusion **115**

A Chapter 2 **125**

A.1 Screenshots of the laboratory experiment 125

B Chapter 3 **131**

B.1 Effect of Post-Truth Comment on Belief in False Claims . . 131

B.2 Evidence of Post-Truth Reasoning in the Control Group . . 132

B.3 Screenshots Survey Experiment 133

Chapter 1

Introduction

In the aftermath of the Brexit and the Trump election, two long-standing liberal democracies find themselves deeply divided, with half the electorate puzzled at how the other half voted. As in any election, remain voters had their reasons to vote remain; leave voters had their reasons to vote leave; Trump voters had their reasons to vote for Trump and Hillary voters had their reasons to vote for Hillary. What made these elections stand apart, however, was the level of factual inaccuracies. On both sides of the Atlantic, campaigners on both sides – but in particular the Leave and the Trump campaign – were widely accused of lying. Leave and Trump were undeterred. And they got away with it. This dissertation is about the citizen consumers of false facts. It asks: Why do voters vote for politicians who lie? What makes them overlook blatantly false facts?

Oxford Dictionary defines 'post-truth' as shorthand for "circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief.'"

The Leave campaign was a case in point. On his blog, Dominik Cummings, campaign director of Vote Leave in 2015-16, described how Leave

won the referendum. "Our story rested on five simple foundations", he wrote. The first foundation was "'Let's take back control'. The overall theme." Why 'back?' Had the UK lost control? No: Cummings explained, "When I researched opinion on the euro the best slogan we could come up with was 'keep control'. I therefore played with variations of this. A lot of people have given me a lot of credit for coming up with it but all I really did was listen. (NB: 'back' plays into a strong evolved instinct – we hate losing things, especially control.").

The second foundation was that, "The official bill of EU membership is £350 million per week – let's spend our money on our priorities like the NHS instead.' " It seems clear that the Leave campaign knew this number was misleading. Cummings wrote, "the Treasury gross figure is slightly more than £350m of which we get back roughly half..." But factual accuracy did not matter. What mattered was success at the polls: "Sometimes we said 'we send the EU £350m' to provoke people into argument. This worked much better than I thought it would." Cummings commented, "It was clearly the most effective argument not only with the crucial swing fifth but with almost every demographic. Even with UKIP voters it was level-pegging with immigration.." The campaign put the slogan on the side of their red campaign bus in which they toured the country. They also ran in targeted internet ads aimed at swing voters.

The third foundation was no less misleading: "If we stay there will be more new countries like Turkey joining and you won't get a vote." It was not a complete falsehood: Turkey had been negotiating accession to the EU since 2005. But progress had been slow and, in 2016, Turkey was nowhere near joining. What is more, any accession requires a unanimous

decision – the UK *had* a say: They could have vetoed it. In 2016, Turkey's accession to the EU was not an issue. Commings writes, "I was surprised at what a shock it was to IN when we hit them with Turkey." The fourth foundation was a gross exaggeration: "The euro is a nightmare, the EU is failing, unemployment is a disaster, their debts and pensions are a disaster, if we stay YOU will be paying the bills." The 2008 financial crisis had hit many EU countries – including the UK – hard. Unemployment is a disaster. But the implication that voters were choosing between leaving a sinking boat and financial ruin was wrong. Finally, the Leave campaign was anti-establishment: "We aligned ourselves with the public who had been let down by the system."

How did Leave win? Cummings admitted, "Anybody who says 'I always knew X would win' is fooling themselves." Nonetheless, he attributed much of the success to three of the five claims – the most misleading ones: "Would we have won without immigration? No. Would we have won without £350m/NHS? All our research and the close result strongly suggests No. Would we have won by spending our time talking about trade and the Single Market? No way." ... "If Boris, Gove, and Gisela had not supported us and picked up the baseball bat marked 'Turkey/NHS/£350 million' with five weeks to go, then 650,000 votes might have been lost."

Five months after the UK's EU referendum, Donald Trump was voted into office. If his opinion about factual accuracy was not clear by then it became clear on day one of his presidency. Trump's presidency began with a controversy over the number of people attending his inauguration ceremony. When photos of his crowd v. Obama's clearly larger crowd came out Trump resorted to creative evidence-making: The U.S.

President had a government photographer edit the official pictures of the inauguration (Swaine 2018). White House Secretary Sean Spicer called a special press conference on the matter. He insisted that Trump's inauguration had drawn "the largest audience to ever witness an inauguration, period, both in person and around the globe". The ungrateful task of explaining why Sean Spicer had lied in his very first press conference fell to Kellyanne Conway, senior aid to the president. The next day, on NBC's Meet the Press, she coined what would become the catchphrase of the year: "Don't be so overly dramatic about it, Chuck. You're saying it's a falsehood. (...) Sean Spicer, our press secretary, gave alternative facts to that" (NBC News 2017).

One of the early cases in which Donald Trump himself had to explain one of his false statements was in February 2017, when NBC's Peter Alexander enquired after his statement that he had achieved the biggest electoral college win since Ronald Reagan. Alexander was about to start quoting statistics about the number of votes Obama had received in 2008. Trump interrupted him, saying he was talking about Republicans. Next, Alexander started giving statistics about the number of votes George H.W. Bush had won in 1988 Trump interrupted him again, saying: "I was given that information. I was just given it. We had a very, very big margin." Alexander proceeded to ask: "... why should the American people trust you when you accuse the information they receive as being fake when you're providing information that's not accurate?", upon which Trump replied: "I was given that information. I actually, I've seen that information around. But it was a very substantial victory, do you agree with that?" The story was widely cited. It reveals two things: First,

the United States President used false numbers to exaggerate his victory. More harshly, one might say, he lied. Second, and more importantly, Trump did not seem to care whether his inititial claim was true or false – he used the numbers to make a point. What mattered was the point. Not the numbers.

These were not the only elections in which voters were frequently exposed to false facts.

Long before the Brexit vote and long before Trump's ascent to the US Presidency, Vladimir Putin was was putting out blatant falsehoods. He, too, got away with it. Over the course of 2014, Kremlin-affiliated media outlets were waging what NATO's top commander General Philip Breedlove called "the most amazing information warfare blitzkrieg we have ever seen in the history of information warfare" Pomerantsev 2014. In February and March of 2014, an army of 'little green men' in military gear without insignia swarmed over the peninsula of Crimea, seizing its Supreme Court, its Council of Ministers, and its parliament building. Putin referred to them as 'local self-defense forces', or as 'volunteers'. In a televised question and answer session on 16 April, Putin said "I will say this clearly. There are no Russian troops in Ukraine" (Oliphant 2015; Engel 2015). A day later, he said the opposite: "of course our troops stood behind Crimea's self-defence forces." (Anischchuk 2014). In a documentary aired on state television a year after the annexation, titled "The Way Home", Putin dropped all pretenses that the little green men were not Russian military. He explained that you need specialists to protect Russians in Crimea from violence and repression by Ukrainian nationalists. "That's why I gave orders to the Defense Ministry – why hide it? – to

deply special forces of the GRU (military intelligence) as well as marines and commandos there ..." (Schreck 2019). As the Soviet-born British journalist Pomerantsev points out, Putin's propaganda is different from Soviet propaganda: "For the Soviets, the idea of truth was important– even when they were lying. Soviet propaganda went to great lengths to 'prove' that the Kremlin's theories or bits of disinformation were fact." Putin barely tried to cover up his own lies.

In July 2014, Channel One, Russia's most popular channel – a state-controlled news channel – aired a false eyewitness account in which a woman claimed the Ukrainian army had publicly nailed a three-year-old boy to a board in a public quare in Slovyansk, a former rebel stronghold. The child's mother was then tied to a tank and dragged through the streets until she died. The woman said she was risking her own life by reporting the story but wated to protect children against Ukrainian army who behaved like beasts and fascists (Ennis 2015). Putin's deputy minister of communication, Alexei Volin, showed no sign of remorse – but of pride – when he was confronted with the false story. "The public likes how our main TV channels present material," he said. "The share of viewers for news programs on Russian TV has doubled over the last two months." (Pomerantsev 2014). In other words, it did not matter if the story was true or false. What mattered was that it got attention. Putin's popularity peaked during the weeks after the annexation of Crimea (Levada 2019).

Voters in these three elections had two things in common: First, they were inundated with information – some true, some false. Watching state-regulated television did not shield them from fake news. Entire campaigns were spreading false information; presidents lied to the people

who elected them. Second, in all three elections voters had access to accurate news. Throughout these campaigns and presidencies journalists and fact-checking charities went to great lengths to verify information. False facts were debunked on a daily basis. Most voters had access to the internet, and in many cases, the facts were only just a google search away. This reflects the conundrum of post-truth campaigns: Factual inaccuracy does stand in the way of electoral success. And so we blamed the voters: How can anybody in their right mind vote for Donald Trump? Or for Brexit? Had they seen them lie? Hadn't they heard them lie? Didn't they know they were lying? Didn't they care? In his book, 'Post-Truth' Matthew d'Ancona quotes the 1992 article that first spoke of 'post-truth'. In it, the Serbian-American writer Steve Tesich wrote that after Watergate and the Iran-Contra scandal the American people were so traumatised that they had started to discredit the value of truth: "We are rapidly becoming prototypes of a people that totalitarian monsters could only drool about in their dreams. All the dictators up to now have had to work hard at suppressing the truth. We, by our actions, are saying that this is no longer necessary, that we have acquired a spiritual mechanism that can denude truth of any significance. In a very fundamental way we, as a free people, have freely decided that we want to live in some post-truth world."

This dissertation does not assume that voters on either one side of the Brexit or Trump electorate were any more gullible than voters on the other side. Hence, it challenges an implicit assumption of much of the ensuing post-truth talk: that Brexit voters believed in the Brexit campaign's

lies and that Trump voters believed in Trump's lies. Instead, this dissertation seeks to investigate mechanisms that might lead any group of voters to tolerate false facts in political campaigns. The starting point is the observation that the information situation in post-truth campaigns places voters in a difficult position. Modern voters are in the middle of a battle in which facts compete with alternative facts, facts and alternative facts compete with fact checks, and all stories are vying for attention. The truth is out there – but getting to it requires investing some serious time and effort. This thesis explores two sets of research questions. The first set is about reactions to false claims coming from politicians: How do voters react to information that looks like it might not be entirely accurate? Do they believe it? Do they tolerate it? What makes them tolerate it? The second set of research questions is about false beliefs that voters themselves hold: What do voters make of fact-checks? Do they have any interest in learning what is true and what is false? Given the sheer amount of false facts in modern campaigns and the failure of fact-checks in these recent campaigns to change the tide, is fact-checking even worth the effort? And how do post-truth surroundings affect the way we react to fact-checks? In short, if we live in a post-truth era and if voters are bombarded with false information is there any hope for *facts* to reach them?

This dissertation builds on two bodies of literature: research on misperceptions and research on motivated reasoning. Following (Nyhan and Reifler 2010) I understand misperceptions to refer to "beliefs that are unsupported by clear evidence and expert opinion". The belief that the UK sends GBP 350 million to the EU each week (held by 42 per cent of people who had heard of the claim (**Stone2018**)) and the belief that Donald

Trump's inauguration was better attended than Obama's (held by 7 per cent of US citizens) (Edwards-levy 2016) fall under this category. Motivated reasoning theory explains why people hold on to beliefs that are unsupported by clear evidence and expert opinion. As its name suggests, it departs from the rational choice assumption that all reasoning is objective. Instead, it posits that "all reasoning is motivated": "people are more likely to arrive at conclusions that they want to arrive at" Kunda 1987; Kunda 1990. Kahan 2016b defines motivated reasoning as "the tendency of individuals to unconsciously conform their assessment of information to some goal collateral to determining its truth." Kahan notes several goals: to maintain a positive self-conception, to rationalize self-serving behaviour, to avoid anticipated stress or anxiety of unwelcome news, to perceive coherence rather than complexity in pieces of evidence relevant to important decisions and, in particular, to protect one's identity. Kahan 2016b, p.2. Whether I credit information or discredit a new piece of information – such as, for instance, an expert saying the UK does not send 350 / million to the EU each week – does not depend on how true it is. It depends on what crediting or discrediting this piece of information does to my standing in a group of people who share identity-defining political commitments Kahan 2016b, p.4. The classic demonstration of motivated reasoning and one of the first studies in the field is Lord, Ross and Lepper's death penalty study Ross, Mark R. Lepper, and Hubbard 1975. The authors sought out individuals with strong views in favour or against capital punishment and had them read two purported investigations on the effectiveness of death penalty as a deterrent against crime – one study

reported it was effective; the other study reported it was not. Both opponents and proponents of death penalty argued that the study that was consistent with their own opinion was most persuasive, and found fault with the respective other study.

A large body of literature testifies to the resilience of political misperceptions (Barrera et al. 2018; Lewandowsky, Stritzke, et al. 2005; Nyhan 2010; Schaffner and Roche 2017; D. Flynn and Krupnikov 2019; Jerit and Barabas 2012; Eric C. Nisbet, Cooper, and Garrett 2015; Nyhan and Reifler 2010; Nyhan and Reifler 2017; Thorson 2015). This dissertation seeks to contribute to relatively recent question: To what extent do people who claim to believe in false information actually believe in false information? And to what extent do they really only tolerate false information, without actually believing it? Classic motivated reasoning theory assumes that if an individual says that they believe in a false claim then they really do believe it.A few recent studies cast doubt on this assumption. Schaffner and Luks point to a motivation other than belief to agree with facts you know to be false: partisan cheerleading (Schaffner and Luks 2018). The authors constructed a survey experiment around the controversy about the crowd witnessing Donald Trump's v. Barack Obama's inauguration. They presented respondents with two photos, labelled 'Photo A' and 'Photo B' and asked, "Which photo has more people?". Photo A had been taken during Trump's inauguration; photo B had been taken during Obama's inauguration. The correct answer was clear and obvious: Photo B had more people. And yet 15 per cent of Trump voters selected image A as showing more people. Among the highly educated (i.e. those who were most likely to recognize the photos and know about the controversy) 26

per cent of Trump voters chose the wrong photo. The authors took this as evidence of partisan cheerleading: Their respondents *could see* which photo had more people. It was obvious. If they chose the wrong one then that must have been a statement of support, not belief. In a similar vein, Bullock et al. 2015 and Prior, Sood, and Khanna 2015 tried offering financial incentives for correct answers. They worked: In both studies, receiving money for correct answers reduced partisan polarization, and increased the percentage of correct answers. If motivated reasoners can be motivated to give the correct answer then this suggests that among people who hold a certain misperception there will usually be at least a percentage who knows, deep down, that the information they claim is true is fishy.

That is the starting point for this dissertation. Following (Schaffner and Luks 2018; Bullock et al. 2015; Prior, Sood, and Khanna 2015) I do not assume people who claim to believe in false and debunked information to be absolutely convinced that those claims are true. I see belief in false and debunked facts as a continuum ranging from 'definitely false' to 'definitely true'. My question is: If people know (or could know, or should know) that a fact is false – then what makes them tolerate these false facts? In chapters 1 and 2, I test two mechanisms that might make people tolerate false facts: stress and low status.

Chapter 1 examines the effect of stress on belief in false and debunked facts: Does stress make people rely on false information that supports your vote choice and silences voices reminding you of fact-checks? I tested this mechanism in an online survey fielded shortly after the Brexit referendum. The core was a timed quiz designed to subliminally stress

respondents. Immediately after exposure to the quiz respondents were shown a number of false facts that had been spread by the Leave campaign and were asked to rate them on a scale from 'definitely true' to 'definitely false'.Overall, the stress treatment did not increase belief in false facts. However, I found heterogeneous treatments: Respondents who saw themselves on the lower echelons of a subjective social status ladder rated the false facts as truer when stressed than those who saw themselves on the higher echelons of the ladder.

Chapter 2 examines the effect of low status on belief in false facts. This was tested in a laboratory experiment in the summer of 2018: I used a pub quiz-style geography quiz to induce status differences and then tested tolerance of clearly false claims coming from one of two alleged feedback givers: One who sided with the advantaged team and one who sided with the disadvantaged team. Low status affected tolerance of false facts: When the person who sided with the low-status team made the false claims they went un-noticed. However, when the person who sided with the other team made false claims the low-status team was more likely, on average, to notice them.

Chapter 3 tests the effect of post-truth surroundings on reactions to unsolicited, counter-attitudinal expert advice. Here, the assumption is that experts on television or fact checks such as the BBC Reality Check rarely have the last word. They operated in an environment where news are heavily commented. Does exposure to comments telling you not to take the expert too seriously make you more likely to reject unwanted statistics? This project is a co-authored project with Rob Johns and John Bartle. We identified false beliefs, then exposed all participants to a fact-check

and some to a post-truth comment. We found that fact-checks had a positive effect, reducing the extent of belief in false facts. Post-truth comments eliminated some (but not all) of that positive effect – even if the comments were attributed to a person who was as unauthoritative as a blogger and even if their sole point was that there is was much information out there that it is best to trust your instincts.

An electronic version of this dissertation, as well as supporting documents (all questionnaires, all datasets, and all R code) is available at https://github.com/julchenhagenstroem/.

Chapter 2

The Effect of Stress on Belief in False Facts in the Brexit Debate

2.1 Abstract

This study investigated the effect of stress on belief in false facts in the leadup to Britain's EU referendum. An online survey was distributed via twitter to United Kingdom residents in the immediate aftermath of the EU referendum. Participants (n=380) were randomly assigned to a stress treatment (a timed quiz about the EU) or a control condition (the same quiz, not timed). Contrary to expectations this study did not find any evidence that stress increased belief in false facts among the whole sample. The treatment did however increase stress among those low in subjective social status: Among those who saw themselves in the lower echelons of British society (both leave and remain voters), the stressed were significantly more likely to rate factually incorrect campaign statements as "true" than their peers in the control group. Those who saw themselves in the higher echelons of British society were not affected by the stress

treatment. These findings illustrate the impact of personal circumstances and of perceived social status on individuals' ability to assess factual information under stress.

2.2 Introduction

If the United Kingdom's EU referendum was to be described in two words many would say: False facts. Ever since the beginning of the campaigns both camps accused each other of lying. The camp that won was the one that was most accused of making false claims: The Leave campaign. This raises an obvious question: Did voters vote for Leave because of the false claims or in spite of the false claims? And if they believed in those false claims what made them believe in them? This project explores the effect of stress on belief in false claims.

In an ideal world, voters would listen to the two sides in a campaign, think about their arguments and policy propositions, seek more information where they need it, fact-check any claims that sounds odd to them, accept the claims if they are correct and reject and forget the claims that are incorrect. In the real world, citizens rarely have the time, energy, or motivation to think carefully about every snippet of political information they are (consciously or subconsciously) exposed to, let alone to fact-check them. More often than not we are too busy or too stressed to bother. This study examines belief in false facts propagated by the Leave Campaign in the United Kingdom's EU referendum. It seeks to capture the effect of feeling rushed on leave and remain voters' ability to assess the factual accuracy of false campaign statements in the Brexit debate.

Dual process models of persuasion reflect this continuum of thinking carefully about a message v. not thinking carefully about it. They suggest that one way of processing information is to carefully evaluate it – this is the "central route" in Petty and Cacioppo's Elaboration Likelihood Model (ELM, cf. Petty (1994)), Chaiken and Eagly's Heuristic-Semantic Model (HSM) speaks of "systematic processing". Another way is to rely on heuristics, or mental shortcuts – this is the "peripheral route" in the ELM or "heuristic processing" in the ESM. According to the ELM, involvement, motivation and ability predict the likelihood of elaboration. More recently, Lodge and Taber (2013) developed a dual-process model of political information processing that accounts for the fact that cognition is inextricably linked with affect: These authors claim that "all thinking is suffused with feeling, and these feelings arise automatically within a few milliseconds of exposure to a familiar sociopolitical object or event" (Kraft, Lodge, and Taber 2015, p.127). This applies to politics: As Taber (2011, p.8) point out, "Citizens are rarely, if ever, dispassionate when thinking about politics" (Taber 2011, p.8).

Much is known about misperceptions. A number of studies confirm the conventional wisdom that people prefer facts that conform to their political opinions to facts that challenge them Kull, Ramsay, and Lewis (e.g. 2003), Gaines et al. (2007), Fischle (2000), De Vries, Hobolt, and Tilley (2015), Sides and Citrin (2007), Nyhan, Reifler, and Ubel (2013), Jacobson (2010), and Nyhan and Reifler (2015a) . People's bias toward facts that confirm their opinions and their resistance to facts that challenge them is generally traced back to motivated reasoning: "reasoning biased to produce emotionally preferable conclusions" (Westen et al. (2006, p.1947),

see also Festinger (1957), Lord, Ross, and Mark R Lepper (1979), and Kunda (1990)). A large body of literature in the cognitive dissonance tradition has found evidence for motivated reasoning biases in political thinking (e.g. Kull, Ramsay, and Lewis 2003; Jacobson 2010; Lodge and Taber 2013).

Less is known about how to fight political misperceptions. (cf. D. J. Flynn, Reifler, and Nyhan (2016) for an overview). Some findings suggest that the circumstances in which one evaluates information affects reasoning. For instance, preparedness (knowing that information might be incorrect), and incentives (financial incentives or imploring on subjects to be accurate) help (cf. Lewandowsky, Stritzke, et al. 2005; Bullock et al. 2015; Bolsen, Druckman, and Cook 2014; Prior, Sood, and Khanna 2015). This suggests that even if (as Lodge and Taber (2013) claim) affect and cognition go hand in hand there must be circumstances that tilt the balance in one direction or the other (i.e. in favour of affect or of cognition). If incentives to think carefully about an issue alleviate motivated reasoning biases then disincentives to think carefully about an issue should increase motivated reasoning biases.

This hypothesis tested in this study is simply: It suggests that the general cognitive overload of modern life, the time pressure we live in, and the mental fatigue that follows from it nudges heuristic or affective reasoning. Stressed people are too tired to engage in cognitive reasoning. Given the stressful nature of modern life it is surprising that the effect of stress on processing false information has not yet been investigated. This study seeks to contribute to filling this research gap.

Krosnick (1991) describes the effect of cognitive overload on techniques to answer long, and burdensome interviews: as people become tired, impatient, and distracted they start "satisficing". They no longer think about interview questions but interpret them superficially and give what they believe will appear to be a reasonable answer. Research findings on stress reflect the negative effect of stress and of overwhelming cognitive demands on performance (cf. Staal 2004; Starcke and Brand 2012, for a review of research on decision-making under stress). Stressors such as large workloads, time pressure, or the anticipation of a stressful event (e.g. a speech) have been found to impair memory performance (cf. Dougherty and Hunter 2003; Keinan, Friedland, and Ben-Porath 1987). When stressed, people scan fewer alternatives and are more likely to rely on familiar responses regardless of previous success of these responses and to rely on the judgment of others (cf. Keinan, Friedland, and Ben-Porath 1987; Driskell and Salas 1991). [1]

If stress prevents people from carefully evaluating facts then it is very likely to have the same effect on their ability to evaluate false facts. This study hypothesizes that stress impairs people's cognitive ability to accurately assess the validity of false facts that are in line with their political opinions (H1). Assuming a positive relationship between stress levels and the extent to which one relies on affective cues when evaluating facts, this study hypothesises that motivated reasoning biases are especially pronounced among those who feel strongly about an issue (H2) and among those who are more stressed to begin with (H3).

[1] Note that some evidence points to an interaction effect of gender on performance under stress: Preston et al. (2007) and Bos, Harteveld, and Stoop (2009) found that only male participants took disadvantageous decisions under stress.

To examine the effect of stress on belief in false facts this study piggybacked on the United Kingdom's EU referendum. Brexit lent itself to an analysis as a polarizing political issue on which false facts had been widely spread and corrected. The Brexit campaign made it possible to study the effect of stress on belief in false facts without creating or correcting false facts – both false facts and fact checks were circulating; and it was plausible to assume that voters had been exposed to both. This study focused on false facts spread by the leave side. This is not because the Remain campaigned sticked to the truth at all times but because the false or misleading claims in the Remain campaign were mainly exaggerated predictions about the economic impact of Brexit. Obviously, those predictions were not falsifiable at the time.

An online survey experiment was conducted to test to what extent a subliminal stressor made Leave voters believe in the false facts of the Leave campaign. Results showed that among the whole sample the stressor – a short timed quiz about the EU – had no affect at all on belief in false facts. However, it found heterogeneous treatment effects: subjective social status moderated the effect of the stress treatment on belief in false facts. Subjects who saw themselves at the bottom of British society performed worse under stress. This was true for leave voters; and, surprisingly, also for remain voters. Similarly, citizens who expected their personal finances to get worse as a result of Brexit and citizens who were low in self-esteem showed significantly higher levels of belief in false facts than their peers in the control group. Implications of the difficulty of these subgroups (or of the left behind) to evaluate facts under stress are discussed.

2.3 Method

Data were collected through a qualtrics survey fielded to twitter users shortly after the 23 June EU referendum in the UK. Twitter was chosen as a recruitment tool in order to obtain a sample of British residents who held strong opinions about Brexit and who, therefore, had reason to engage in motivated reasoning when confronted with facts about the EU. The survey was embedded in a larger questionnaire about political opinions and tweeting behaviour; it was fielded by Rob Johns, Heinz Brandenburg and Marcel van Egmond.

We created five matching twitter accounts that we used to contact Brexit tweeters. To maximise transparency and trust in the academic nature of the survey we used Twitter names for our accounts that reflected the intent of the survey [2]. In lieu of a bio we wrote a short description, "A research project from the Universities of Essex, Strathclyde & Amsterdam. We don't take sides, but we're keen to hear which side you're on." A link to the Government Department at the University of Essex was provided underneath this description. A picture of an EU and a British flag was used as a profile picture, 10 Downing Street was used as a background picture to capture the attention of subjects interested in politics (see**??**).

Potential respondents were identified based on their tweeting behaviour.

[2] BrexitStudy, BrexitSurvey, eurefstudy, EURefSurv2016, and brexit_survey

We sent out direct messages to 6651 individuals who had used Brexit-related hashtags in the weeks before the referendum. Our messages included a short invitation text [3] and a link to our survey on qualtrics.com. The link was indivualised, or linked to each person's twitter account, and could only be taken once. The survey was fielded on 25 June and remained online until 2 August. Most subjects participated in the first two weeks after the referendum (see figure **??**). Participants were not paid. Data were pre-screened for completeness and residence in the United Kingdom. 278 UK residents completed the survey using links connected to their twitter accounts. We received a few messages from individuals who had problems accessing their individual links or who were interested in re-tweeting the link to the survey. These individuals were sent a direct link to a slightly adapted version of the same survey. [4] 103 participants completed the survey using this open link. Their twitter names were scanned for personal acquaintances. One personal acquaintance and nine subjects not residing in the UK were excluded. The samples accessing the survey via individual links and the sample accessing it via the open link were similar across the main political and demographic measures; they were pooled. [5]

[3] The wording of the invitations sent to our subjects varied slightly; one of five versions read "@TwitterName: Would you participate in a survey on the EU referendum? Please click here: [link]"

[4] As the open link was not associated with any twitter accounts we included a question asking for subjects' twitter names.

[5] 70 per cent of eligible voters accessing it via an individual link and 64 per cent of eligible voters accessing it through the open link had voted to remain in the UK. 90 v. 88 per cent lived in England. The mean age was 39 years in the sample accessing the survey via direct links (sd=13.89), and 42 years in the sample using the open link (sd=14.17). In both datasets, the main source of income (71 and 80 per cent, respectively) was employment, only 1.5 and 1 per cent received unemployment benefits). Gender, however was not balanced: 66 per cent of participants using individual links and 53 per cent in the sample using the open link were male. (Most of the individuals who

The final (pooled) sample consisted of 357 subjects, aged 16 to 75, 63 per cent male. 319 lived in England, 26 in Scotland, nine in Wales, and three in Northern Ireland. The ratio of leave/remain voters did not mirror the results of the referendum (52 to 48 per cent): a large majority of 225 subjects, i.e. two thirds of the sample reported they had voted to remain. 99 subjects had voted to leave, 5 had not voted. (Hence, 42 per cent of the sample were male remain voters.) As in the referendum, the ratio of leave voters to remain voters increased with age. Among participants in their teens, twenties, and thirties, less than a third voted to leave. In contrast, a third of those in their thirties, a forth of those in their forties, and half of those in their sixties and seventies indicated they had voted to leave.

As expected, the sample was highly interested in the referendum and highly opinionated. 93 per cent of both leave and remain voters were "very interested" in the EU referendum. Similarly, 93 per cent of leave voters and 95 per cent of remain voters were sure or absolutely sure of their vote choice. Two thirds of leave voters saw no reasons at all or not very many reasons to remain, only two per cent saw "very many" reasons to remain. Remain voters were even more convinced of their cause: 90 per cent saw no reasons at all or not very many reasons to leave, only one per cent saw "very many" reasons to leave. Leave voters' forecasts for a post-Brexit Britain testified to a high level of belief in the campaign's promises: 59% of leave voters (v. 16% of remain voters) believed immigration would be curbed after Brexit, 56% (v. 1%) thought that the NHS

contacted us to share the link were women who passed it on to more female than male friends).

would be better off; 61% (v. 2%) expected the general economic situation to improve, and 25 % (v. 1%) anticipated that their personal financial situation would improve.

The invitation message we sent out contained a link to the survey on qualtrics. It began with questions on party identification, interest in the EU referendum, vote choice, a measure of of attitude strength: "How sure were you about your referendum vote choice?"), and expectations about what would happen after the referendum Next, we asked about their feelings about the outcome of the referendum, whether they had tweeted about it, sources of information (newspaper / radio / internet), and demographics. We assessed a few personality measures: the Ten-Item Personality Inventory, trust, self-esteem, subjective social status, life satisfaction, need to evaluate, and need for closure.

The MacArthur Scale of Subjective Social Status was used to capture subjects' perception of their own social status across SES indivators. It shows a ten-step "social ladder"; participants place themselves on of the ten steps. The main interest in this measure was to identify individuals of low subjective social status who were assumed to feel relatively deprived. Relative deprivation, "subjective feelings of anger, resentment, and frustration in response to negative social comparisons with relevant others" has been found to predict stress (Smith et al. 2012). [6]

[6]The description of the ladder read: "Think of this ladder as representing where people stand in the United Kingdom. At the top of the ladder are the people who are the best off – those who have the most money, the most education, and the most respected jobs. At the bottom are the people who are the worst off – who have the least money, least education, and the least respected jobs or no job. The higher you are on this ladder, the closer you are to the people at the very top; the lower you are, the closer you are to the people at the very bottom. Where would you place yourself on this ladder? (Please write below a number between 1 and 10.)"

Next, subjects were randomly assigned to a treatment group (n=183) or a control group (n=174). All respondents answered the Eurobarometer's three EU knowledge questions ("The EU currently consists of 28 member states."; "The members of the European parliament are directly elected by the citizens of each member state." and "Switzerland is a member state of the EU."). The control group saw these questions in the same format as all other questions; they were introduced as "True or False?" The treatment group was given a visible timer that started ticking as soon as they reached the page. The questions were introduced as "True or false?We've introduced a time limit to make this a bit more exciting! You have 35 seconds to answer these questions." Because this was a subliminal primer no manipulation check was added. (In addition, adding any questions between the stress treatment and the false facts would have delayed the questions about the false facts questions and therefore weakened the feeling of stress).

Immediately after the stress manipulation participants were presented with a battery of seven statements about the EU. They were asked to rate each statement as "definitely true", "probably true", "probably false", or "definitely false". A "don't know" answer was provided. Four of these statements (the ones of interest to this study) were factually incorrect campaign statements that were widely dissipated by the Leave campaign, and widely criticised and corrected by the remain campaign as well as by British and international media: "Leaving the EU frees up £350m a week for the NHS"; "If Britain had remained in the EU it would have had to accept Turkish membership"; "The EU could have forced British soldiers to join a European army", and "The EU could have made

Britain join the Euro". The remaining statements were distractor items. Two were correct: "The pound plunged to its lowest level since 1985 after Britain voted to leave the EU", and "All council areas in Scotland voted to remain". The final statement had been subject to speculation ever since it was reported in a tabloid: "The Queen backs Brexit". The seven statements appeared in a random order. A hidden timer measured how long subjects spent evaluating them. The survey ended with questions on how much they thought they had in common with poeple who wanted to leave and with people who wanted to remain (apart from what they thought about Europe) and, finally, measures of curiosity, need for cognition, and need for evaluation.

2.4 Results

This study set out to investigate the effect of a mental stressor – a timed quiz on the EU – on the ability to assess factually incorrect statements about the EU. The aim of the stress manipulation was to mimick the situation people are usully in when they are exposed to false facts in politics, i.e. to nudge affective rather than cognitive processing. It was hypothesized that stress increases motivated reasoning biases, i.e. the tendency for leave voters to rely on affective cues and believe in any statements coming from their side, regardless of factual accuracy.

The main independent variable was subjects' assessment of the four false facts about the implications of leaving or remaining in the European Union. A simple additive index was created to measure belief in these false facts. For the claims that disregarded EU legislation (Britain would have had

to accept Turkish membership; the EU could have made Britain join the Euro, or a European army) each "definitely true" was assigned a four, each "probably true" was assigned a three, and each "probably false" was assigned a one. These numbers were doubled for the statement that leaving the EU would have freed up £350 a week for the NHS. Belief in this false fact was deemed to weigh particularly heavily because it was based on a wrong figure and because it was probably the most fiercely corrected claim. Hence, for this statement, "definitely true" was assigned eight points, "probably true" was assigned six points, and "probably false" was assigned two points. Scores were added up so that the resulting index ranged from 0 (all four statements were rated as "definitely false") to 20 (all four statements were rated as "definitely true"). The index was reliable (Cronbach's α = .8 with a 95 per cent confidence interval of .73 to .87, item-rest correlations ranged from .55 (the NHS claim) to .70 (the EU army claim).

As expected, leave voters reported greater levels of belief in false facts about the EU than did remain voters (cf. distribution in figure **??**). Yet contrary to expectations (H1), stressed Leave voters were not any more likely to report belief in false facts about the EU than were unstressed leave voters. Figure **??** shows virtually no difference between stressed and unstressed leave voters. Similarly, it shows that stressed Remain voters were not any less likely to believe them than unstressed remain voters.

Did at least the most convinced leave voters react to the stress treatment? Hypothesis 2 held that the most convinced Leave voters were most likely to rely on affect or heuristics when stressed and believe in any negative

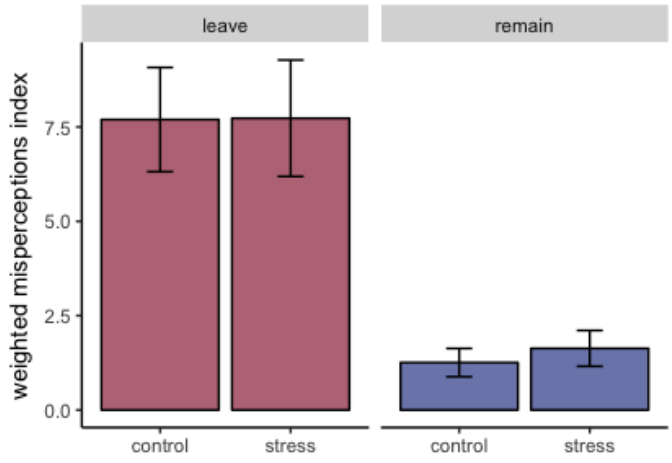


FIGURE 2.1: Belief in false facts about the EU by vote choice

information about the EU. Figure ?? shows the effect of stress by attitude strength (wherein die-hard Leave or Remain voters were leave or remain voters who were absolutely sure of their referendum vote choice, i.e. who ticked the maximum of 6 on the scale). It shows that even the most convinced Leave and Remain voters were unaffected by the stress treatment: the average level of belief in false facts was almost the same in the treatment and remain groups. However, the graph shows slightly different results for those who were not 100 per cent sure of their vote choice. Less convinced Leave voters did slightly *better* in the stress condition, less convinced Remain voters did slightly worse under stress. In neither case did the difference reach statistical significance. Hence, this data showed no evidence for H2, either.

The final hypothesis stated that those who were most stressed to begin with ought to show the greatest motivated reasoning biases when stressed. This is the only hypothesis for which the data shows at least some suggestive evidence: A few indicators of chronic stress moderated

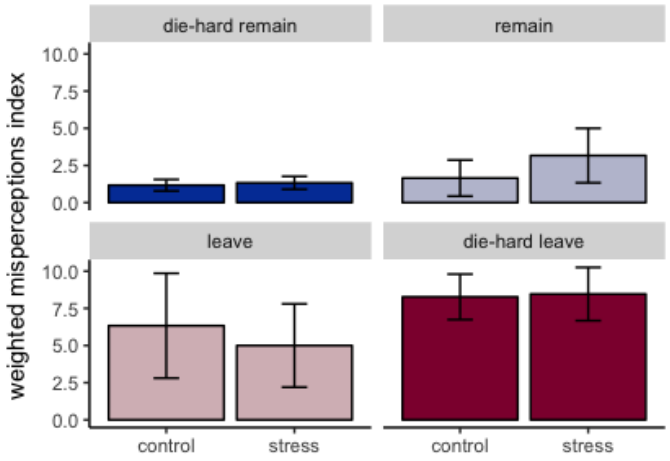


FIGURE 2.2: Belief in false facts about the EU by vote choice and certainty about vote choice

the effect of the stressor on belief in false facts: subjective social status, the anticipation of losing money after Brexit, and life satisfaction.

To capture the effect of different levels of subjective social status a dummy variable and a three-level factor variable was created. For the three-level factor variable those who saw themselves on step 1,2,3, or 4 of the social ladder were defined as low in subjective social status; those who believed to be on steps 5, 6, or 7 were labelled as medium and those who were on steps 8, 9, or 10 were seen as high in subjective social status. The dummy variable distinguished between subjects of low subjec- tive social status (steps 1 to 5) and subjects of high subjective social status (steps 6 to 10). Figure **??** shows two things: First, it shows a downward trend in belief in false facts: On average, Leave voters who placed themselves on the upper steps of the social status showed lower levels of belief in false facts than Leave voters who placed themselves on the lower steps. This most probably reflects the effect of education and knowledge about the EU. Second, it shows no evidence that social status moderates the effect of

stress on Leave voters: those at the bottom, in the middle, and at the top of the ladder seem to respond to the treatment in a similar way. However, the small sample size and the large confidence intervals suggest that for Leave voters the social status variable should be broken up into smaller categories.

The three-category variable shows interesting results for Remain voters. Among this group social status did seem to moderate reactions to stress: Stressed Remain voters who saw themselves at the very bottom or the very top of British society showed higher levels of belief in false facts than their peers in the control group. In both subpopulations (that is, Remain voters high and low in subjective social status) the difference in means between misperceptions among the stressed and unstressed was significant at a 95 per cent confidence interval [7]

The fact that among some groups of Remain voters the treated showed higher levels of belief in the Leave campaign's false facts than the control group is peculiar: if, as hypothesized, stress nudges reliance on affect or heuristics then stressed Remain voters ought to be even *less* inclined to believe in any negative information about the EU than unstressed Remain voters. (All four false facts put EU membership in a negative light.) One possible explanation is acquiescence bias: The time pressure may have increased the likelihood that these respondents agree with *any* statement. Somewhat unfortunately all false facts were coded in the same direction, with "Definitely true" as the first option on the left-hand side and "Definitely false" as the last option on the right-hand side. If the

[7] For subjects low in subjective social status: $M_{stressed}$=2.00), $M_{control}$=0.75 t(25.77) = -2.05, p = 0.05; subjects high in subjective social status: $M_{stressed}$=1.84), $M_{control}$=0.44 t(22.28) = -2.08, p = 0.05.

driving mechanism is acquiescence bias then timed Leave and Remain voters ought to have shown higher levels of belief in each of the facts, including the distractor claims, and in particular the statement to which the answer is unknown: The Queen backs Brexit. This was not the case On average, stressed Leave voters showed slightly higher levels of belief in this claim, however, differences in means were far from significance. Among Remain voters, however, there was no significant difference in how the treatment group and the control group evaluated the three distractor items.

Next, subjective social status was broken up into a dummy variable. Figure **??** shows the effect of stress on on subjects who saw themselves on steps 1 to 5 of the ladder and on subjects on steps 6 to 10 of the social latter. Contrary to the three-level factor his dummy showed some suggestive evidence that stress increased misperceptions among leave voters who believed to be of average or low subjective social status. The difference did not however reach statistical significance ($M_{stressed}$=9.84), $M_{control}$=8.08 t(47.06) = -1.19, p = 0.24). Leave voters who saw themselves in the upper echelons of British society did slightly better when stressed. Here, too, the difference in means was not significant ($M_{stressed}$=5.38), $M_{control}$=7.24 t(42.80) = 1.41, p = 0.16). This may reflect different effects of the stress manipulation: It may have stressed those at the lower levels of the ladder, and woken up those at the higher levels of the ladder.

The questionnaire collected a few other indicators of prior stress: unemployment was used to measure economic hardship. A question asking if respondents anticipated losing money as a result of Brexit was used to measure fear of economic hardship. Due to the small sample size the

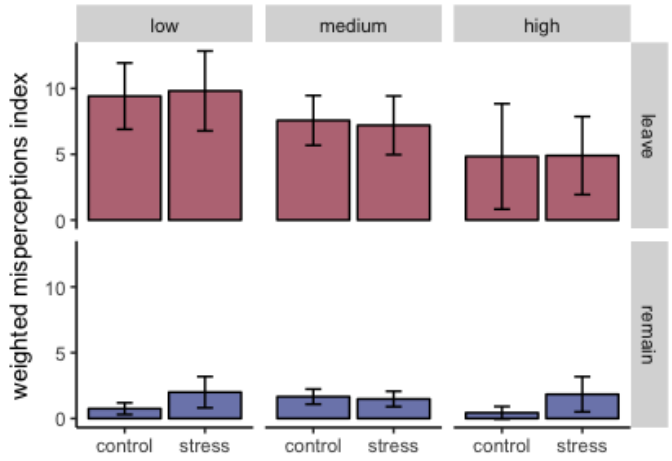


FIGURE 2.3: Belief in false facts about the EU by vote choice and subjective social status

moderating effect of unemployment could not be analysed: in this convenience sample, only 4 Leave voters and 16 Remain voters were unemployed. Fear of economic hardship did not moderate the effect of stress on Leave voters – but it did moderate its effect on Remain voters: Remain voters who anticipated losing money because of Brexit did worse under stress; the difference in means was just short of statistical significance ($M_{stressed}$=1.59), $M_{control}$=0.99 t(149.04) = 1.41, p = 0.06). This, too, defied expectations – remain voters had no reason to believe in false facts that put EU membership in a negative light. If, as hypothesized, stress increased reliance on affect or heuristics then Remain voters should have outrightly rejected any negative information about the EU under stress.

To further investigate the effect of stress among subpopulations of remain voters I looked at a few other control variables. The stress manipulation did not only impair performance among remain voters who were low in social status, and who feared losing money: it also had a negative effect on Remain voters who were not satisfied with their lives (cf.

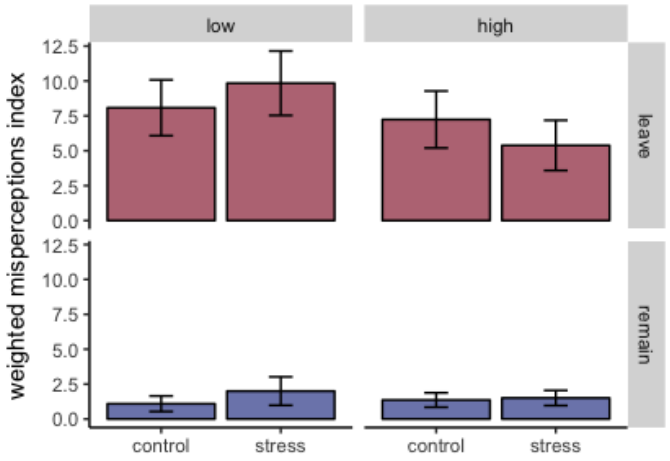


FIGURE 2.4: Belief in false facts about the EU by vote choice and subjective social status (dummy)

??), and who scored low on social trust (cf. figure ??). In both cases the differences in means were statistically significant at a 90 per cent confidence interval. Stressed remain voters low in self-esteem did also worse on average than unstressed low-self-esteem remain voters; yet here the difference was only significant at a 90 per cent confidence interval.

8

Neuroticism, measured as a dummy variable, had no impact on responses to stress (neither for Leave nor for Remain voters), neither did any other Big Five personality factors, need for cognition or need to evaluate.

[8] Life satisfaction was measured as a single item measure ("On the whole, how satisfied are you with the life you lead?"). It moderated the effect of the stressor on remain voter's performance: $M_{stressed}$=1.66), $M_{control}$=0.95, t(104.04) = -1.97, p = 0.05).

Generalised social trust was measured as a dummy variable, either agreeing that "most people can be trusted" or that you "can't be too careful in dealing with people". Among remain voters who ticked the latter box, the stress treatment had a negative: $M_{stressed}$= 1.70, $M_{control}$=095; t(104.04)=-1.97, p = 0.05.

Self-esteem was measured on a seven-point measure of the question "How far would you say that this is true of you: 'I have high self-esteem.'" Among remain voters who scored low or medium, i.e. who tickeed points 1-4 the unstressed outperformed the stressed: $M_{stressed}$=2.51), $M_{control}$=1.43, t(58.93) = -1.69, p = 0.10

FIGURE 2.5: Belief in false facts about the EU by vote choice and life satisfaction (dummy)

2.5 Discussion and Conclusion

Contrary to expectations this study found no evidence that a simple stress manipulation designed to induce mental fatigue increased motivated reasoning biases among leave voters. For two reasons, these null findings ought to be taken with a pinch of salt: First, the sample was small, unrepresentative and unusually opinionated – among leave voters levels of belief in these false facts were so high that at least some of the non-effect may be due to ceiling effects. (Belief was too high for the stressor to have an additional effect). More importantly, it is highly likely that the stress manipulation was too weak. The sample was highly educated and so three questions may not have been very difficult to them. The time limit was most probably too short: 35 seconds may not have been too long to stress respondents. Hence, it is very probable that this treatment did not mimick the rush or the mental fatigue it was supposed to mimick. It may even have made respondents more alert, acting as a wake-up call in a

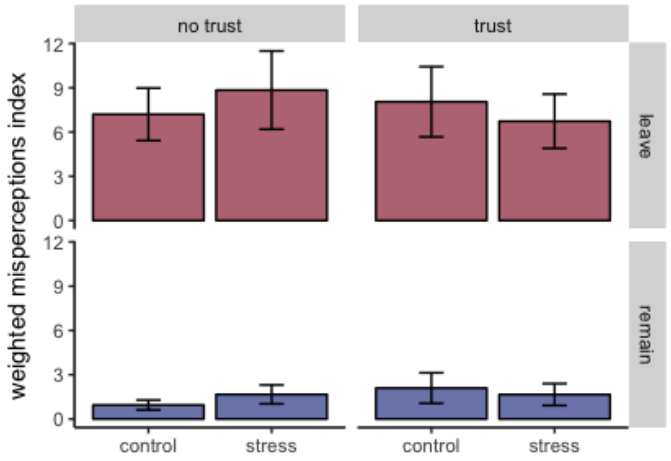


FIGURE 2.6: Belief in false facts about the EU by vote choice and generalised social trust

long study. If so then the effect of this alertness may have offset the effect of stress. Therefore, this study ought not to be seen as a evidence that stress has no effect on belief in debunked facts. Future research with a more stressful stress manipulation is needed.

Nonetheless, even this relatively un-stressful stressor did affect evaluations among a few subpopulations. It never improved judgment. If it had an effect, it impaired judgment. There were three groups of people who showed higher levels of belief in the false facts under stress: individuals who were low in subjective social status, individuals who anticipated their personal finances to get worse after Brexit, and individuals low in self-esteem. Notably, this was true for Leave *and* Remain voters. (Among the subset of Remain voters generalised social trust and life satisfaction also moderated the effect of the stress treatment on belief in false facts: stressed Remain voters who were low in trust or low in life satisfaction did significantly worse on average than unstressed remain voters who were also low in trust or life satisfaction.)

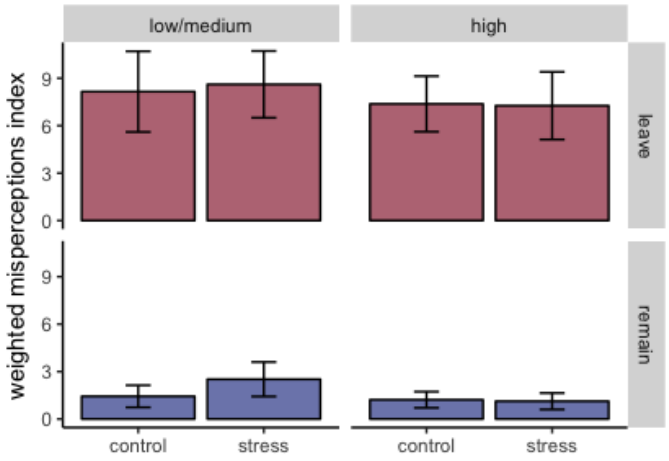


FIGURE 2.7: Belief in false facts about the EU by vote choice and self-esteem

These characteristics that moderated the effect of stress – feeling at the bottom of society, being less educated, having worse jobs and less money, and expecting it all to get worse are signs of feeling left behind. These are the people that the populist campaigns of the past few years – be it Trump in the US, Brexit in the UK, the National Front in France, or the AfD in Germany – have targeted. If those who feel left behind are vulnerable to a stress treatment that does not affect those who don't feel left behind they may be even more vulnerable to real-life stressors when evaluating information. How low social status affects information processing is an important avenue for future research.

Chapter 3

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

3.1 Abstract

Many recent electoral events have been characterised by false claims which, despite abundant fact-checking, were often widely believed. This paper asks why voters do not punish false facts in political campaigns. Building on findings that suggests a link between status threat and Trump/Brexit voting it investigates a simple hypothesis: That low group status or threatened high group status makes voters susceptible to overlooking false claims coming from politicians who boosts their group identities. A laboratory experiment (n=277) was conducted to test this hypothesis. The set-up was a pub-quiz style geography quiz. Unbeknownst to the participants the likelihood of drawing easy or hard questions was varied so that one team would do worse and receive a lower payoff than the other

team. After the quiz, both teams were shown two pieces of feedback: One person called the quiz 'fair play'; the other person called it 'unfair' and demanded a pay rise for the disadvantaged team. One of the two feedback givers were randomly assigned to make false claims. Results showed that the disadvantaged overlooked false claims coming from the person who sided with their team: There was no significant difference in how disadvantaged team members rated the accurate and the inaccurate version of the 'unfair' feedback. If, however, the false facts came from the person who sided with the advantaged team the disadvantaged were more likely to notice the false facts: Those on the disadvantaged team who saw the version of the 'fair play' feedback that included false claims were less agreed (p=0.02), thought it was less accurate (p=0.07), and were slightly (but not significantly) less likely to rate its author as a good team representative (p=0.22) than their peers who saw the factually accurate version. If generalisable to electoral campaigns these results are sobering: If candidates recognize the disadvantage of those who feel disadvantaged then lying will not cost them any votes among that group.

3.2 Introduction

Politics is not known to be the most ethical of professions. For the longest time, a lie (or two) in a political speech did not raise an eyebrow. But in 2016, the amount of lies in political campaigns seemed to have crossed a threshold: The factual inaccuracies in the UK's Brexit campaign and, subsequently, in Donald Trump's election campaign led to a public outcry over politicians spreading false information. What is startling is that

on both sides of the Atlantic the campaigns that were most accused of making false or misleading claims carried away the elections: in June, the Leave campaign won the EU referendum, having toured the United Kingdom in a bus with a false figure of the the weekly cost of EU membership on it. In November, Donald Trump, whose claims about attendance at his inauguration would lead to the coining of the infamous word "alternative facts", won the White House. This project is about the citizen voters confronted with false facts in political debates. It asks: Why would voters vote for a candidate who makes claims that are, simply, false?

This paper challenges the implicit assumption of much of the post-truth talk: that Brexit voters believed in the Brexit campaign's lies and that Trump voters believed in Trump's lies. Instead, it proposes an alternative mechanism to explain how voters might overlook false facts in political campaigns. It suggests a link between low group status or threatened group status and tolerance of false facts in political campaigns that side with one's in-group.

Both the Trump campaign and the Leave campaign argued that 'ordinary citizens', and the British or, respectively, the American people had had the rough end of the stick for too long. Both campaigns appealed to segments of their societies who were concerned about natives losing jobs to immigrants, and promised to improve the status of natives relative to immigrants. Both campaigns also appealed to segments of the British and American society who mourned the loss of Britain's or America's status in the world, and promised to recover a once-held and now lost superior status of their nation. It seems that this campaign strategy paid off: Voters who were concerned about the status of natives or the

loss of America's or Britain's status were more likely to vote for Trump, or, respectively, Brexit. What did these voters make of the false claims in these two campaigns? The motivated reasoning-informed hypothesis advanced in this paper is: nothing. They overlooked them. The hypothesis is simple: The mere feeling that the status of a group individuals identify with is lower than it ought to be or lower than it used to be makes voters susceptible to overlooking false facts coming from politicians who offer positive social identity, for instance by recognizing a group's hardship or promising to improve their status.

A laboratory experiment is conducted to test the causal mechanism on a non-political issue in a controlled laboratory setting at a UK university. The focus is on low status groups: How do groups that are lower in status react to false facts in statements from people who promise to raise their group status?

3.3 Theoretical Background

This study borrows from two theoretical traditions: social identity theory, and motivated reasoning theory. Social identity theory argues that people make comparisons between not only themselves and others, but between their groups and other groups. It is based on a number of group experiments, in which subjects favoured in-groups and discriminated against out-groups even if group assignment was as flimsy and unimportant as preference for a painter (Tajfel 1974; Tajfel 1982; Tajfel and Turner 1986). Tajfel and Turner's social identity theory assumes that part of peoples' self-image is based on group membership (this is their social

identity) and that groups are associated with (socially consensual) positive or negative value connotations. Comparing favourably with relevant out-groups leads to high prestige and positive social identity; comparing unfavourably leads to low prestige and negative social identity (Tajfel and Turner 1986, p. 40). The authors (Tajfel and Turner 1986) theorize that individuals seek to achieve or maintain positive social identity: "The aim of the differentiation is to maintain or achieve superiority over an out-group on some dimensions." (Tajfel and Turner 1986, p. 41).

Unfavourable comparisons (i.e. negative social identity) are said to lead to three types of behaviour: (1) individual mobility (leaving a group to join a more positively distinct group), (2) social creativity (e.g. comparing themselves on a different measure or with a different out-group, or reinterpreting characteristics formerly seen as inferior), or (3) social competition (fighting to improve the group's negative image or position). Much of social identity theory is concerned with the conditions under which negative social identity generates conflict over scarce resources (cf. Tajfel and Turner 1986, p. 44f and Tajfel 1974, p. 76f).

According to Tajfel and Turner's theory, two types of groups are particularly prone to conflict: low-status groups who perceive the status differences as unstable and/or illegitimate and high-status groups whose higher status is threatened. Low status groups are competetive toward the dominant group to the degree that "(a) subjective identification with the subordinate group is maintained; and (b) the dominant group continues or begins to be perceived as a relevant comparison group" (Tajfel and Turner 1986, p. 45). They are most likely to rise up if they are unable to join other groups (i.e. individual mobility is unavailable) and if

they perceive status difference as both unstable and illegitimate. (Tajfel and Turner (1986, p.45) argued that this was probably the set of conditions that underlay the development of ethnocentrism among Black Americans, French Canadians, and New Zealand Maoris.) High-status groups whose high status is under threat are most likely to 'react in an intensely discriminatory fashion' to attempts by the subordinate group to reverse their group status when they perceive their superiority as legitimate (Tajfel and Turner 1986, p.45-46). This paper does not seek to explain conflict. It takes a step back and considers the cognitive consequences of low status and threatened high status in election campaigns.

On this front, Motivated Reasoning Theory offers helpful insights. It argues, as its name suggests, that "all reasoning is motivated": "people are more likely to arrive at conclusions that they want to arrive at" (Kunda 1987; Kunda 1990). Kahan (2016b, p.2) defines motivated reasoning as "the tendency of individuals to unconsciously conform their assessment of information to some goal collateral to determining its truth." Abundant research confirms that individuals see information through an ideological lens. (cf. also Fischle 2000; Bolsen, Druckman, and Cook 2015; Kull, Ramsay, and Lewis 2003; Nyhan, Reifler, and Ubel 2013; Bartels 2002; Kahan 2016b). Lodge and Taber 2013 present a number of experiments that provide consistent evidence that people do not only defend their prior attitudes but that this happens outside of conscious awareness: we know how we *feel* about a piece of information before we know what we *think* of it. They theorize that conscious deliberation is but a way of rationalizing these unconscious processes to find reasons why we feel the way we feel about something (Kraft, Lodge, and Taber 2015, p.129). Westen et al.

(2006) provided the first neuroimaging evidence for motivated reasoning. During the 2004 U.S. presidential elections, they conducted experiments in which they asked Bush and Kerry supporters to consider consonant and dissonant information about both candidates. Magnetic resonance imaging showed that the reasoning areas of partisans' brains shut down when they were considering information that was threatening to their own candidate. In contrast, when they saw consonant information the emotional areas of the brain lit up.

Drawing on social identity theory and motivated reasoning theory this paper assumes that (a) voters identify with groups, (b) voters have a need to see their in-groups compare well with out-groups, and that (c) voters are unable to assess information that threatens their in-group in an even-handed way. This paper explores the cognitive side-effects of coping with negative social identity – the feeling that one's in-group does *not* compare well with relative out-grous – in a political environment in which lying politicians offer positive social identity. It suggests that the low or threatened group status blurs the lens through which one sees politicians who side with one's in-groups.

The Trump campaign and the Brexit campaign seemed to appeal to people who felt economically disadvantaged, or left behind. It also seemed to appeal to people who felt that the status of their people was threatened. This paper focuses on two types of status threat: Global status threat (their country's status in the world was under attack) and racial / ethnic status threat (the status of ordinary (implied, white) citizens was under attack).

3.3.1 Global Status Threat

Both campaigns addressed their electorate in terms of their national (British or American) identity. On its own, this is unremarkable. What makes it remarkable is that both campaigns presented their national identity as being under threat. For instance, Nigel Farage, the key figurehead for the Leave campaign declared that "We are British; we are not going to be bullied by anybody" – implying that Britain had been bullied before (ITV 2016). Similarly, the slogan 'Take *Back* Control' implied that Britain had lost control. The same way, Donald Trump promised to 'Make America Great *Again*', implying that America had lost at least some of its greatness. Trump's campaign launch speech in New York City in June 2015 was a vivid example: 112 words into the speech, he declared "We don't have victories anymore. We used to have victories, but we don't have them" (Phillips 2017). In the same speech, Trump exclaimed that 'Sadly, the American Dream is dead.' – shattering an idea, the American Dream, that is at the very heart of America's national identity.

Both Trump and the Leave campaign offered a way out of the dilemma, promising to free this besieged American or British identity. The Trump campaign conveyed the impression that other nations, including China, had trampled on America's status in the world. The Brexit campaign conveyed the impression that the EU had trampled on Britain's status in the world. Both campaigns promised to un-do this injustice and to return their nations to their rightful position as great nations (implied, better than other nations). Given the history of the American and the British nation it is unsurprising that this rhetoric fell on fertile ground: The Leave

campaign operated in a former (one might argue a fallen) empire, a former colonizer that has ceded at least some of its power to the European Union; Trump spoke to a current superpower, the world's largest economy and the mightiest military – but also a country that has tarnished its image in the wars of the past few decades and is losing ground, politically and economically, to other rising powers – above all, China.

Populist campaigns beyond Britain and the US reveal similar patterns. For instance, when Tayyip Erdoğan and Vladimir Putin address their diasporas in Europe they consistently depict their countries as belittled by the West. The message they have been sending (be it in person or through their various state-controlled media channels) has been clear: their diasporas belonged to a great nation and ought to be proud of it. (And, to express their Turkish or Russian identity, they ought to vote for them.) Both Erdoğan and Putin have been surprisingly successful in gathering support amongs their expat communities across continental Europe, in particular in Germany, where the social and economic status of citizens of Turkish and Russian decent tends to be below average.

3.3.2 Ethnic status threat

Trump and Brexit campaigners also appealed to people in terms of their identity as 'ordinary' citizens – which, at least in the US context, is often understood to mean 'white' (Tesler 2016b). Both campaigns bemoaned, subtly or not so subtly, the declining status of white men.

For instance, the Leave campaign sided with 'working people'. In a 3

June 2016 interview, Michael Gove warned that "you can say, 'their concerns don't matter' [...]. You can dismiss the concerns of working people. [...] You are dismissing the concerns of working people." In good populist fashion he continued, 'You're on the side of the elites, I am on the side of the people" Sky News 2016. Two weeks later (15 June), he pointed directly at European immigrants: "At the moment all our public services – the NHS and education are under strain, as a result of unlimited free movement from the EU" BBC 2016. Meanwhile, Nigel Farage complained that "it is wrong, wrong, wrong, that for average, decent families in this country, their living standards have fallen by 10 per cent over the course of the last few years. And it is about time [...] we started thinking about [...] not just about the rich getting richer, but about ordinary, decent Britons who have got a rotten time. And they really have, too. " ITV 2016. Nigel Farage's 'Breaking Point' campaign poster was less subtle: It showed a photograph of migrants crossing the Croatia-Slovenia border in 2015. The only white person in it was obscured by a box of text, saying 'We must break free of the EU and take back control of our borders' (see Stewart and Mason 2016). Hence, leading Leave campaigners appealed to voters in terms of their identity as 'working people'; they appreciated their hardship, and they vowed to fight to improve their living standards.

Donald Trump's rhetoric was similarly to the point. He promised to bring back 'our' jobs, 'our' manufacturing (namely, Ford), and 'our' military. In his campaign launch speech Trump shouted, "We need money. (...) And we need the right people." In the same speech he described his (mainly White) audience as 'the best and the finest', and contrasted them with 'the people Mexico sends. (...) they're not sending their best. (...) They're

bringing drugs. They're bringing crime. They're rapists' (Phillips 2017).

3.3.3 Status Threat and Vote Choice

Post-election research showed that perceived discrimination against whites was indeed correlated with support for Brexit and Trump. Sides, Tesler, and Vavreck (2018) examined British Election Study data. They found that white voters who thought there was a lot of discrimination *against* white people were over 60 percentage points more likely to support Brexit than white voters who thought there was a lot of discrimination in *favour* of white people in the UK (Sides, Tesler, and Vavreck 2018, p.216-217).

Tesler (2016a) analysed the 2016 American National Election Pilot Study. They found that in the US, whites who thought it was extremely likely that "many whites are unable to find a job because employers are hiring minorities instead" were over 50 points more likely to support Trump than their country men and women who disagreed (Tesler 2016a). Fowler, Medenica, and Cohen (2017) coined the term 'white vulnerability': "the perception that whites, through no fault of their own, are losing ground to other groups". To measure it, the authors constructed a three-item scale, asking whether whites were "economically losing ground through no fault of their own"; whether discrimination against whites was "as big a problem as that against Blacks and other minorities"; and whether minorities overtaking whites as the majority of the U.S. population by 2050 would "strengthen or weaken the country." Controlling for the usual predictors of Trump support (e.g. partisanship, racial resentment, living in the South, gender, and employment status) white vulnerability strongly

and significantly predicted support for Donald Trump. In the US, the perception that whites are being discriminated against is growing; and it is increasingly tied to voting Republican (Sides, Tesler, and Vavreck 2018, p.170f). Americans who voted for Trump saw whites as more discriminated against than Muslims of black Americans (Edwards-levy 2016).

Naturally, none of these correlational studies can establish if white vulnerability, or status threat *caused* citizens to vote for Trump. However, there are a few experimental and panel-data studies – and they do point to a causal connection: First, and before the elections, Major, Blodorn, and Major Blascovich 2016 found experimental evidence that reminding whites that they would soon be outnumbered by non-whites made white Americans high in ethnic identification more supportive of Donald Trump. (It also increased support for anti-immigrant policies and opposition opposition to political correctness among this group.)

Second, and after the elections, Diana Mutz found panel data evidence that rising white and global status threat made Americans shift toward Trump. Mutz argued that white Americans were facing two types of status threat: they were losing in numbers, and they were losing in global status (the US is gradually losing its status as the number one world power). As hypothesized, both types of perceived status threat led people to vote for Trump: People who scored higher on social dominance orientation in 2016 than they did in 2012 (a proxy for racial status threat) were more likely to shift toward Trump. Similarly, people who saw China as more threatening in 2016 than they did in 2012 and who grew less supportive of free trade agreements in 2016 were more likely to shift toward Trump.

3.3.4 Proposed mechanism

This paper suggests a causal connection between the kind of group status threat many UK and US voters felt in 2016 (be it the status of 'ordinary' people, 'white' people, or the status of the 'American' or 'British' nation) and the fact that they overlooked dubious information in campaigns that promised to raise or recover their group status.

False claims coming from politicians who have boosted one's in-group's status poses a dilemma: On the one hand, individuals have a need for factual accuracy (Kunda 1990; Taber and Lodge 2006). On the other hand, according to social identity theory, they have a need for positive group distinctiveness. Acknowledging that a politician is wrong about one thing (e.g. that they don't get their facts right) would mitigate the positive effect of their status-boosting speech on in-group members' social identity. If, as motivated reasoning research has shown, in-group members are willing to sacrifice money to see their in-group fare better than out-groups (Tajfel 1970) (and if, as history has shown, people sacrifice their lives in wars protecting their in-groups) it seems plausible that they will – at least to some extent – sacrifice their need for factual accuracy as well.

Politicians can nudge group categorization by addressing individuals as members of an in-group. They can nudge negative social identity by pointing out to a low-status group ways in which they do *not* compare well with relevant outgroups or by pointing out to high-status groups that its higher status is threatened. (Both happened in 2016.) And then

they can alleviate these feelings of negative social identity. If, for instance, politicians praise an in-group's hard work, recognize their hardship, or promise to improve their socio-economic standing relative to other groups then they effictively provide a straw for in-group members to clutch at. In-group members will *want* them to be right because they have made them feel better about the groups they belong to. But what if these political elites then say or do something that *ought to* set off alarm bells? What if they use a piece of information that isn't entirely true, or if they advocate a political idea that you do not quite subscribe to, or if they make an over-the-top statement about other groups? What if, ultimately, they propagate political extremism or violence? The very beginning of this chain is politicians making benign but false claims. This paper suggests, simply, that in-group members have a motivation to overlook such false claims – not necessarily to believe them, but to overlook them.

This process ought to work at different levels of conscious awareness. At a pre-conscious level individuals will simply not see or feel any signs of fishiness. At a somewhat higher level of conscious awareness citizens may sense something suspicious but may chose not to spend too much time or too many thoughts on it. At an even higher level of awareness people may sense the fishiness, look into the facts, acknowledge them to be false, but decide to vote for the candidate nonetheless. Put differently, they will over-value the positive impact of the candidate's identity-justifying information and under-value the importance of his or her factual accuracy for their vote choice.

This does not negate the well-established research finding that people do have accuracy goals (Kunda 1990; Taber and Lodge 2006). It does,

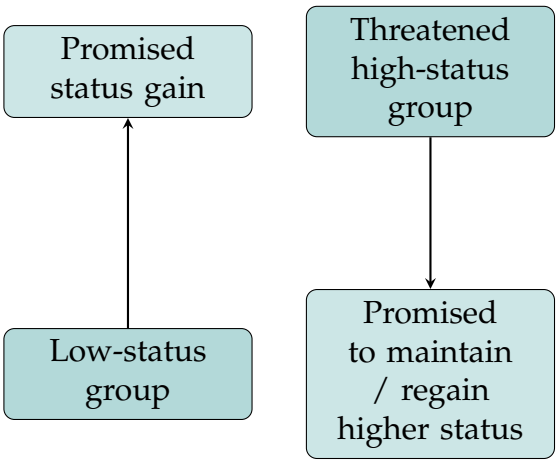


FIGURE 3.1: Groups of interest

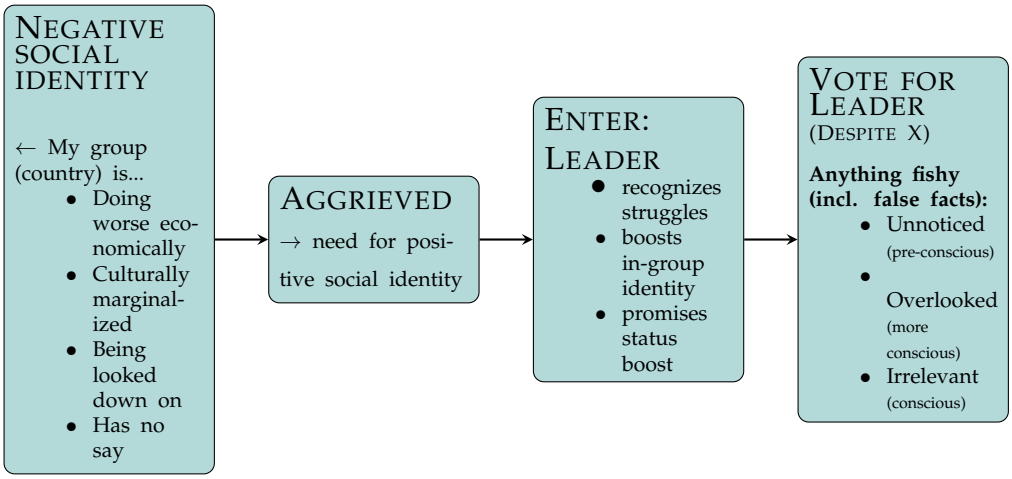


FIGURE 3.2: Proposed causal process – Low-status groups

however, posit that people's need for positive group distinctiveness (or the need to recover lost status) exceeds their need for factual accuracy. Citizens may not believe in everything their leaders say. But as long as these leaders promise to boost their in-group's low or lost status then the nitty-gritty of their talk becomes irrelevant.

A laboratory experiment was designed to test the causal mechanism in a non-political context focussing on economic differences. [1]

[1]The oTree code is available here; the pre-registration is available here.

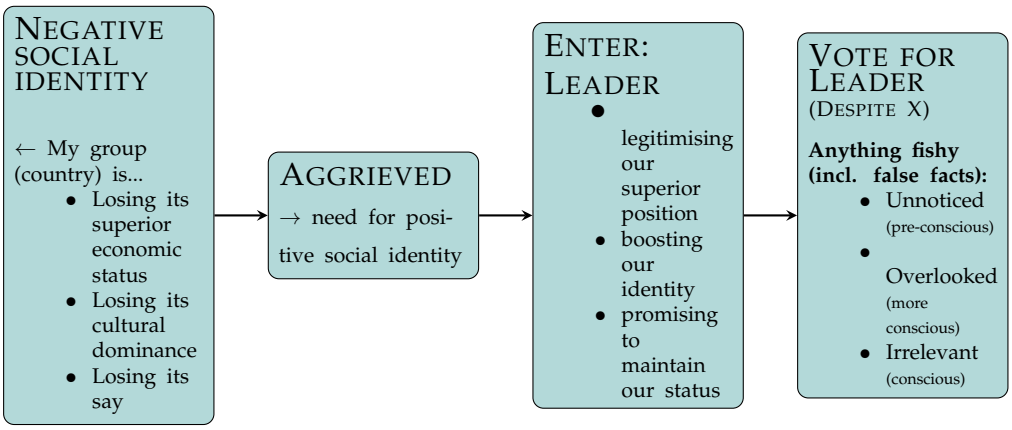


FIGURE 3.3: Proposed causal process – Threatened high-status groups

3.3.5 Method

In a 2*2 design, this study varied a) advantage or disadvantage at a money-paying real effort task (a pub quiz-style geography quiz) and b) exposure to false facts by a person siding with the advantaged or the disadvantaged team. The experiment began with a pub quiz-style geography quiz, desigend to establish status differences. Subjects were randomly allocated to an advantaged team (presented as 'Team A') or a disadvantaged team (presented as 'Team B'). Both teams answered 12 multiple-choice geography questions (3 rounds * 4 questions). The questions were randomly drawn from two sets of thirty questions: an easy set containing questions such as 'What is the capital city of Germany?' and a difficult set containing questions such as 'What is the capital city of Liechtenstein?'. Unbeknownst to the participants, the odds of drawing easy or difficult questions varied by team. For each question, the advantaged team had an eighty per cent chance of drawing an easy question, and a twenty per cent chance of drawing a difficult question. The disadvantaged team had an eighty per cent chance of drawing a difficult question, and a twenty per

cent chance of drawing an easy question. (Therefore, on average, the advantaged team would answer 9 easy questions and 3 difficult questions, whereas the disadvantaged team would answer 9 difficult questions and 3 easy questions.) The quiz was timed: Participants had thirty seconds to tick one of the four answers and hit a 'submit' button. (If they failed to submit their answer within thirty seconds they were automatically forwarded to the next page.) After submitting each question and before moving on to the new question respondents saw the respective question for the other team (see page 125 in the appendix for example questions). The quiz was divided into three rounds with four questions each. After each round players were shown summary statistics for that round: the correct answers to both teams' questions and, for each question, the percentage of respondents who submitted a correct answer. After the last round, i.e. after 12 questions players were shown their own payoffs and the mean payoffs for both teams (see page **??**). Payoffs were designed to depend on participants' own performance as well as their teams' performance: Each correct answer was worth GBP 1.00. Participants kept half of their earnings and contributed the other half to their team's pot, which was evenly divided amongst all group members. To mimic a pub quiz group sizes were set at 4-8 team members.

Following the quiz, respondents were asked for feedback. (They received three questions, asking if they thought any questions were too easy (i), too difficult (ii), and what they thought about the thirty second time limit (iii). Next, they were asked to consider two other people's feedback. One

person's feedback called the game 'fair play', saying Team A (the advantaged team) had had the 'luck of the draw' (see page 126); the other person called the game 'unfair', suggesting that Team B (the disadvantaged team) be paid an additional GBP 5.00 on top of their payments (see page 129). Crucially, one of the two feedback pages were randomly assigned to contain one false claim and one gross exaggeration. The false claim stated that "Some of the places in that quiz don't even exist." The exaggerated claim varied depending on team membership: If assigned to the 'unfair' treatment, the last feedback point said the thirty seconds were "barely enough time to read the questions." If assigned to the 'fair play' treatment, it said they were "more than enough", and that "10 seconds would have been plenty." [2]

This study attempted to capture tolerance of false facts on a spectrum ranging from *overlooking* false claims to *believing* them. First, subjects were asked if they 'generally agreed' with both people's feedback (see page 129). Second, they rated to what extent both authors were 'factually accurate', and 'a good representative of my team' (see page 129). (To disguise the intent of the question participants were also asked to rate how 'educated' the author was.) Third, respondents rated each point, including the false claim, on a four-point scale from 'definitely true' to 'definitely false' (see page 129). The experiment ended with a short questionnaire (stage 4) including demographics and various control questions.

[2]NB: Informal feedback from participants on the advantaged team showed that the exaggerated claim in the 'fair play' feedback was slightly less exaggerated than intended: Some of their questions were so easy that they did not need the entire 30 seconds. (Nonetheless, 10 seconds would not have been enough to answer any of the directions questions or the questions from the difficult set.) Generally, answers to the open-ended feedback question about the time limit revealed that almost all respondents thought 30 seconds was a fair limit.

	False 'fair play'	False 'unfair'
Advantaged team	1	2
(Team A)	n=69	n=64
Disadvantaged team	3	4
(Team B)	n=83	n= 61

TABLE 3.1: Research Design - Study 1

A difficulty that is inherent in studying respondents' susceptibility to overlooking false claims is that it is imperative that they see the false claims (and do not just skim-read the treatment without noticing the false claim). At the same time, one cannot make the false claim stand out too much and one cannot simply ask by way of a manipulation test if they read it. (Because printing a false claim in red letters or even just asking if they read this claim would draw their attention to it. That of course would defeat the purpose of measuring if they noticed it.) To maximise the likelihood that respondents read the feedback it was kept short. The four sentences that included a false claims were layouted as bullet points. Hence, even skim-readers ought to at least have skim-read the false claim. A second difficulty that arises when respondents rate two different people's opinions is that they need to remember which person they are rating. Therefore, when respondents answered questions about either feedback they were able to scroll down to see the respective feedback.

3.3.6 Hypotheses

This experiment sought to create two status groups: an economically disadvantaged group and an economically advantaged group. The general

hypothesis depends on status: For the low-status group (the disadvantaged team), it was hypothesized that exposure to a message recognizing the team's disadvantage and suggesting to raise their payoffs makes team members overlook false claims by the source of the message. For the threatened high-status group (the advantaged team), it was hypothesized that exposure to a message legitimising their higher status makes team members overlook false claims by the source of the message. The following hypotheses will be tested across three dependent variables: general agreement, rating the feedback as 'factually accurate', and rating its author as a 'good team representative'.

Hypothesis 1 considers the hypothetical scenario in which a person who sides with your team either makes false claims or doesn't: It is hypothesized that factual accuracy has no effect on perceived suitability as a team representative:

Hypothesis 1a: Disadvantaged team members who see a factually inaccurate 'unfair' feedback rate the author of the 'unfair' feedback more favourably than their peers who see a factually accurate 'unfair' feedback.

Hypothesis 1b: Advantaged team members who see a factually inaccurate 'fair play' feedback rate the author of the 'fair play' feedback more favourably than their peers who see a factually accurate 'fair play' feedback.

Hypothesis 2 considers the hypothetical scenario in which a person who sides with the *other* team either makes false claims or does not. In this case, it is hypothesized that in-group members will notice (and punish)

the false claims in the other camp:

Hypothesis 2a: Disadvantaged team members who see a factually inaccurate 'fair play' feedback rate the author of the 'fair play' feedback more favourably than their peers who see a factually accurate 'fair play' feedback.

Hypothesis 2b: Advantaged team members who see a factually inaccurate 'unfair' feedback rate the author of the 'unfair' feedback more favourably than their peers who see a factually accurate 'unfair' feedback.

Hypothesis 3 zooms in on the groups that were randomly assigned to see false facts in the feedback that sided with their group (and no false facts in the other person's feedback).

Hypothesis 3a: Those on the disadvantaged team who are exposed to a false 'unfair' feedback and an accurate 'fair play' feedback will rate the author of the (false) 'unfair' feedback more favourably than the author of the (accurate) fair play feedback.

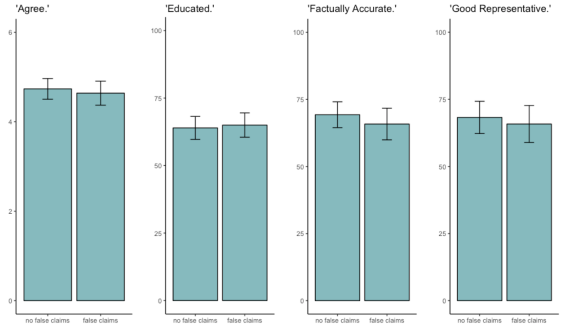
Hypothesis 3b: Those on the advantaged team who are exposed to a false 'fair play' feedback and an accurate 'unfair' feedback will rate the (false) 'unfair' feedback more favourably than the author of the (accurate) fair play feedback.

3.3.7 Results

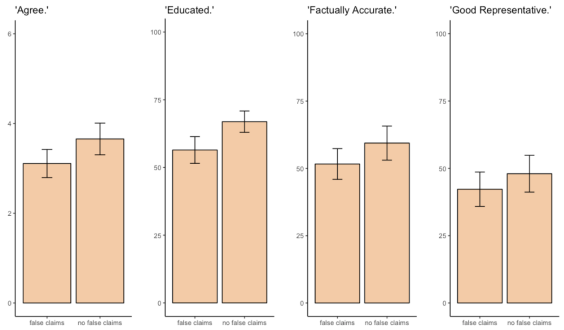
The lab experiment was fielded to a convenience sample sample of 277 subjects at ESSEXLab, Colchester, UK in the summer of 2018 (June-Sep). [3] Subjects were recruited from ESSEXLab's participants' pool. 48 per cent were undergraduate students; 62 per cent female, 54 per cent British. [4]

[3] Results presented here exclude a first round with 264 participants, conducted June 5th-7th, 2018. The first round showed the disadvantaged group was slightly more tolerant of facts by the person who recognized their in-groups' hardship than by the person who called the game 'fair play'. In contrast, the advantaged game was slightly more likely to tolerate false facts in the 'unfair' feedback than in the 'fair play' treatment. However, group differences were far from statistical significance. Post-experimental feedback from individual participants suggested that the null results stemmed from cognitive overload: the treatment, that is, the two feedback pages were fairly long (around 11 lines), and many participants only skim-read them. What is more, there was no 'back' button so that participants who did not read the feedback pages were unable to refer back to them as they were answering questions about the two feedback givers. Two measures had been taken to ensure participants knew which feedback they were rating: First, respondents were shown each feedback before they were asked questions about ("As you will remember, this was feedback 1/2. (...)". Second, each feedback was assigned a colour, which was not associated with any particular political party: The 'fair play' feedback and all questions about it were shown on a teal-coloured background, whereas the 'unfair' feedback and all questions about it were shown on a beige background. Because these measures proved to be insufficient two further measures were taken before the study was re-run: First, to encourage respondents to read the feedback in the first place both feedback pages were shortened; important sentences were printed in bold, and the feedback points that contained false or exaggerated claims were formatted as bullet points. Second, to ensure respondents knew which feedback they were rating and could refer back to it if they had not read it they were shown the feedback on two of the three pages on which they were asked questions about it. Hence, when answering if they generally agreed with a feedback and when assessing how true or false each individual claim was respondents were able to scroll down to view the respective feedback, introduced as: "For your reference, this was feedback 1:". The feedback was not however reprinted on the second page, asking if 'The points this person makes are factually accurate' and if 'This person is a good representative of my team.' These questions were intended to measure to what extent subjects pre-contiously overlooked the false claims. In this case, showing the feedback again would have allowed respondents to check how factually accurate the feedback was and, thereby, would have defeated the purpose of the question.

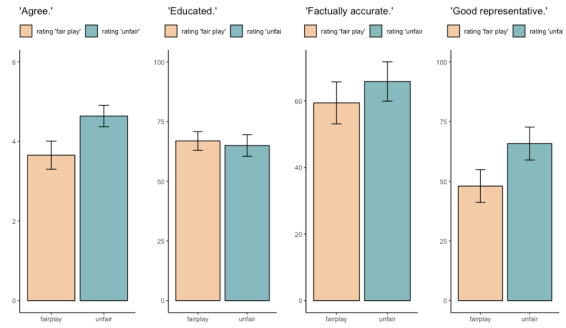
[4] 34% postgraduate students, 10% administrative staff, 3% faculty, 84% affilated with the University of Essex, mean age=31 years (min age=19 years, max age=80 years, two thirds in their twenties), 54% British (15% Eastern European, 6% Southern European, 4% Western European).



(A) Disadvantaged team rating the 'unfair' feedback

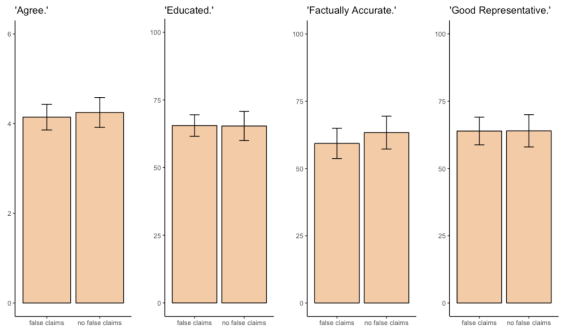


(B) Disadvantaged team rating the 'fair play' feedback

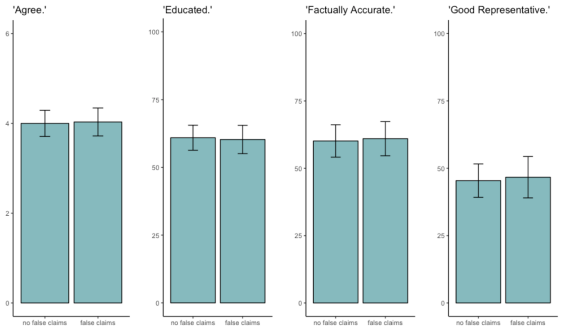


(C) Disadvantaged players who saw false claims in the 'unfair' feedback (group 4) rating the (false) 'unfair' feedback and the (correct) 'fair play' feedback

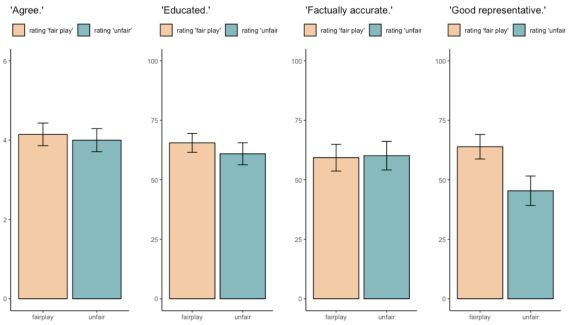
FIGURE 3.4: Disadvantaged team



(A) Advantaged team rating the 'fair play' feedback



(B) Advantaged team rating the 'unfair' feedback



(C) Disadvantaged players who saw false claims in the 'unfair' feedback (group 4) rating the (false) 'unfair' feedback and the (correct) 'fair play' feedback

FIGURE 3.5: Advantaged team

Table 3.1 shows the number of participants in each group: 133 participants were randomly assigned to the advantaged team. 69 of them saw false claims as part of the feedback that called the game 'fair play'; 64 saw false claims as part of the feedback that called it 'unfair'. 144 participants were assigned to the disadvantaged team. 83 of them saw false claims in the 'fair play' feedback; 61 saw false facts in the 'unfair' feedback. [5]

In all sessions, the average payoffs for advantaged players exceeded that of the disadvantaged. Advantaged team players earned GBP 8.90 (min=GBP 5.75, max=GBP 11.80) on average; disadvantaged players earned GBP 4.80 (min=GBP 2.25, max=8.40)). [6] [7] A large majority (260/277) of players correctly remembered that on average, Team A had received a higher payoff than Team B.

Did players notice the false claim? Did they punish it? Three questions were asked to assess to what extent players noticed, or overlooked this false claim: On a first page, respondents were asked if they agreed: 'Generally speaking, do you agree with the author of this feedback?' (1='Strongly Disagree', 2='Disagree', 3='Slightly Disagree', 4='Slightly Agree', 5='Agree', 6='Strongly Agree'). Next, respondents were to rate

[5] (Due to a mistake one session was larger with 9 participants on the advantaged team and 10 on the disadvantaged team.)

[6] Calculated payoffs differed from paid payoffs in two instances: First, because of a minimum payoff policy respondents who earned less than GBP 5.00 (n=78) were paid GBP 5.00. Second, due to maintenance problems two studies started 5 and 10 minutes late, respectively. Subjects in those studies received a GBP 1.00 or GBP 2.00 top-up to compensate for their extra time.

[7] The difference in payoffs varied quite substantially: In two sessions, it was less than 1 GBP (0.21 and 0.80, respectively). This was mainly due to the fact that the game manipulated the *chances* of drawing easier or more difficult questions, not the numer of questions.

the author of each feedback along three dimensions: education (a distractor), factual accuracy ('The points this person makes are factually accurate'), and suitability as a team representative ('This person is a good representative of my team.'). All three questions were measured on a scale from 0 (not at all) to 100 (very). Due to space contraints the results reported here below focus mainly on suitability as a team representative; the full results are shown in the appendix.

Figure 3.4 shows results for the hypotheses for the disadvantaged team (1a, 2a, and 3a). Figure 3.5 show results for the advantaged team (1b, 2b, and 3b). Beige bars represent ratings of the 'unfair' feedback; teal-coloured bars represent ratings of the 'fair play' feedback.

The main question this project sought to answer was: Do players overlook false facts that appear in the feedback that sides with their team? As hypothesized, they do. The first (top) graph in figure 3.4 shows evidence for the disadvantaged team: It compares how the disadvantaged team rated the accurate version of the 'unfair' feedback ("no false facts") and the inaccurate version of it ("false facts"). For each evaluation the graph shows two bars, representing those who saw (and rated) the accurate version ("no false facts") and the inaccurate version of it ("false facts"). As expected, there is absolutely no difference between the two groups. Whether a disadvantaged member saw an accurate version of the 'unfair' feedback or a version with false claims in it had absolutely no effect on how they evaluated it. Those on the disadvantaged team who rated an 'unfair' feedback that contained false claims gave it an average rating of 66/100. Their peers who rated the factually accurate version of the same feedback gave it a mere two points more (68/100). The difference

in means was far from statistical significance (t(131)=-0.53, p = 0.59). (The results for the other two dependent variables are shown in the appendix.)

The same was true for the advantaged team: As shown in figure 3.5 those on the advantaged team who rated a 'fair play' feedback that contained false claims evaluated it just as favourably as their peers who had seen the factually accurate version of it. Those on the advantaged team who rated a 'fair play' feedback that contained false claims gave it an average rating of 45. Their peers who rated the factually accurate version of the same feedback gave it an average representation rating of 47. Again, this two point-effect of factual accuracy was far from significance (t(123)=-0.24, p = 0.80).

Hypotheses 1a and 1b are therefore confirmed. Factual accuracy has no effect on how players rate the feedback that favours their team. This finding implies that the effect of lying – or *not* lying – on individuals' chances as being seen as a good representative are negligeable.

What about false claims in the *other* camp? Do players notice and punish false facts that appear in the feedback that favours the *other* team? The disadvantaged do (at least a little); the advantaged do not. The middle graph in figure 3.4 shows evidence for the disadvantaged team, comparing ratings of the 'no false facts' and the 'false facts' version of the feedback that favoured the other team. It shows that the disadvantaged were generally less agreed with the author of the inaccurate 'fair play' feedback; they rated them as less accurate and an (even) worse representative than the author of the accurate version of it. However, only the difference

in general agreement reached statistical significance at a 95 per cent confidence interval; the difference in accuracy ratings reached significance at a 90 per cent confidence interval (see appendix). The difference in representation ratings (42 v. 48/100, on average) did not reach significance (t(135)=-1.23, p = 0.22). This is suggestive, but not conclusive evidence – Hypotheses 2a cannot be confirmed.

When it comes to the advantaged team the data is clear (middle graph in figure 3.5): Contrary to hypotheses there is no evidence that the advantaged punished the person who favoured the other team for lying. Advantaged team membes gave the 'unfair' feedback accuracy ratings in the lower 60s and representation ratings in the mid-40s – regardless of whether or not that feedback contained false claims . Hence, hypothesis 2b is rejected.

Hypotheses 1 and 2 examined counterfactual situations in which a potential representative either makes false claims or does not make false claims. Hypothesis 3 examines a more common situation: one in which one person lies and another person tells the truth. If the lying person is the one who favours an out-group then in-group members have every incentive to notice the lies. But what if the lying person is the one who favours one's in-group? Half the players in this experiment were nudged to be motivated reasoners: Those who saw false claims as part of the feedback that favoured their team (groups 1 and 4 in table 3.1). If motivated reasoning extends to overlooking false claims then these ought to turn a blind eye to the false claims and rate the feedback that favours their own team more favourably than the feedback on the other side. That is exactly what they did.

As shown in the bottom graph of figure 3.4, disadvantaged players who saw a lying feedback favouring them and an honest feedback favouring the other team showed significantly higher levels of agreement with the former than the latter. They rated it as slightly (but not significantly) more accurate, and deemed its author a much better team representative than the author of the feedback that favoured the opposing team ($M_{\text{correct 'fair play'}}$=48/100, $M_{\text{false 'unfair'}}$=66/100, t(87)=-3.66, p = 0.00.).

The same was true for the advantaged team: As shown in the bottom graph of figure 3.5 those who were nudged to be motivated reasoners, that is, those who were confronted with a 'fair play' feedback that contained false claims and an 'unfair' feedback that did not contain any false claims evaluated the former a better team representative than the latter: $M_{\text{correct 'unfair'}}$=45/100, $M_{\text{false 'fair play'}}$=64/100, t(132)=4.58, p = 0.00. This confirms hypotheses 3a and 3b.

8

3.4 Discussion

This laboratory study was designed to test the effect of boosting in-group status on tolerance of false facts among the in-group. It asked two main research questions: First, do groups that are economically disadvantaged relative to other groups overlook false facts by leaders who recognize their disadvantage and who promise to improve their group's relative status? Second, do advantaged groups that fear losing their advantage

[8] ADD Bonferroni adjustments.

overlook false facts by leaders who justify their group's advantage and promise to maintain it? The core of the experiment was a pub quiz-like geography quiz in which the chances of receiving easy or difficult questions were manipulated: One (advantaged) team had a higher chance of receiving easier questions, and therefore a higher average payoff than the other (disadvantaged) team. After the quiz, respondents were exposed to two people's feedback. One person called the quiz 'unfair'; the other called it 'fair play'. One of the two feedback pages was randomly assigned to include a false claim, complaining that "some of the places don't even exist", and an exaggerated claim, complaining that the thirty second time limit was too short ('barely enough time to read the question') or, if assigned to the 'fair play' feedback, too long ('10 seconds would have been plenty').

Hypothesis 1 examined the (counterfactual) scenario in which a person who favours one's in-group either makes false claims or does not make any false claims. As hypothesized, it didn't affect their standing: Both the disadvantaged and the advanted agreed with the feedback that favoured their team, evaluated their author as a good team representative and stated that 'the points this person makes are factually accurate' – regardless of whether they were accurate or not. This finding confirms recent findings that loyalty trumps honesty Hildreth and Cameron Anderson 2018 Implications for political campaigns are sobering: If grossly exaggerating or making false claims does not affect perceived leadership qualities then politicians have no incentive *not* to lie. As long as they are loyal to their target in-group and as long as their target in-group is large enough to win them the election they will get elected.

Hypothesis 2 examined the (counterfactual) scenario in which a person who favours the *other* team either makes false claims or does not make any false claims. In this case, it was hypothesized that players would notice and punish those false claims. Results are inconclusive. The disadvantaged team showed some signs of sanctioning lies in the opposing team; the advantaged team showed no signs at all. On average, the disadvantaged team 'slightly agreed' with the accurate 'fair play' feedback. They 'slightly disagreed' with the inaccurate 'fair play' feedback. Here, the difference in means was statistically significant (see results in the appendix). When asked to assess how accurate the feedback was they gave the accurate feedback a score of 60, and the inaccurate feedback a score of 50. Here, the difference was significant at a 90 per cent confidence interval (p=0.07, see appendix). Yet when it came to evaluating how apt the person was to act as a team leader respondents seemed to care a little less about how honest they were: The disadvantaged rated the author of the incorrect 'fair play' feedback as slightly, but not significantly less accurate and as a slightly, but not significantly worse team representative than the author of the accurate version of the 'fair play' feedback.

Hence, hypothesis 2 could not be confirmed. For two reasons, these (partially) null findings ought to be taken with a grain of salt: First, there were clear signs that the disadvantaged rated the feedback that favoured the advantaged less favourably when it contained false facts. The fact that group differences did not reach significance across all indicators ought to be seen as a reflection of the small sample size. A larger sample size would most probably have pushed the them toward significance. Second, and more importantly, there was a flaw in the design. Somewhat

unfortunately, the false claims were not 'neutral'. The statement that 'some of these places don't even exist' was clearly false. However, it implied that the game was rigged – and members of the disadvantaged team must have felt cheated. This claim recognized their hardship – even if it appeared as part of the 'fair play' feedback. Future research ought to investigate reactions to more 'neutral' false claims.

Similarly, the fact that the advantaged team let false facts in the 'fair play' feedback slide should not be interpreted as final evidence that status threat does not affect the way individuals evaluate a person who favours the out-group. It is likely to the failure of this study to credibly threaten team A's status.

Hypothesis 3 zoomed in on those who were nudged to be motivated reasoners. The disadvantaged behaved as expected: They rated the lying 'unfair' person as a far better team representative and as slightly more accurate than the honest 'fair play' person. If generaliseable to election campaigns this is consequential. In our experiment voters did not have to choose between two candidates – in real elections they do. This makes it all the more likely that they will overlook false facts. Therefore, these results suggest that if voters who identify with a low-status group have the choice between a candidate who wants to raise their group status but lies and a candidate who wants to maintain the status quo and is honest then they will overlook the false claims and vote for the former.

The high-status group behaved in the same way: Advantaged players who saw a 'fair play' feedback that contained false claims rated the author of that 'fair play' feedback as a better team representative (albeit not

as more accurate) than the author of the (correct) 'unfair' feedback. This, too, is alarming. When voters who identify with a high-status group have the choice between a candidate who promises to maintain their relatively higher status and a candidate who wants to level out status differences they, too, will overlook the false claims and elect the candidate on their side.

Two limitations ought to be noted. The main caveat concerns the advantaged group. This study was designed to create a low status group and a threatened high-status group. It did create status differences. The low status group's low status was built up during a twelve-question quiz that culminated in a very low payoff – on average, four pounds lower than the other team's payoff and, in many cases lower than the minimum payoff. (Which made them eligible for the lab equivalent of welfare benefits; i.e. the un-earned minimum payoff of GBP 5.00). The advantaged group's higher status was similarly consolidated: Their success was consolidated: They beat the other team across most of the twelve questions, and made more than they may have expected to make in a 30-minute study. However, the attempt to threaten the high status group's relatively higher status seems to have been less successful. The status threat depended entirely on a few sentences in the 'unfair' feedback: The demand to top up Team B's payoffs was assumed to act as a status threat to Team A. This is unlikely to have worked. The advantaged group had nothing to lose: their own payoffs were safe and even in the unlikely event that the experimenter had topped up team B's payments they were still part of the winning team. It is very likely that the threat to their higher status did not felt very real. Future studies ought to create a more realistic sense of

status threat, ideally coming from above, not from a fellow participant.

A second caveat is the small sample size. Larger studies are needed to detect small effects and heterogenous treatment effects. In particular, the sample size was too small to examine the effect of status-boosting speech among players who according to social identity theory were most likely to engage in violence: those who strongly identified with their team, who thought status differences were illegitimate and who believed it was possible that the status differences that resulted from the quiz might be overturned. (I would expect these players to be particularly immune to false claims in the feedback on their side.)

3.5 Conclusion

This study contributes to research in an important way: It showed that low status makes individuals vulnerable to overlooking factual inaccuracies in statements from people who sympathize with their group. In light of the rising levels of inequality in Western Democracies and the rise of populist leaders appealing to voters who believe to be socially or economically disadvantaged this is important to know. If feeling disadvantaged makes voters vulnerable to overlooking false facts in campaigns that promise to raise their status then it is unlikely that the present tide of populist politicians with agendas that are based on facts dubious quality will turn.

This laboratory experiment was designed to test a causal mechanism. As with any laboratory study the flipside of isolating the mechanism in a

controlled lab environment is that the setting deviates from a real-world setting. For external validity, the next step in this research avenue is to test this mechanism in a setting that is closer to the everyday life of disadvantaged voters. For instance, online survey experiments could manipulate low group status and then expose respondents to a hypothetical election in which one campaign recognizes their hardship and seeks to level out inequality. It could vary the factual accuracy of their campaign promises.

In addition, more research is needed on the effect of status threat. Both laboratory and survey experiments could be used to test if status threat makes voters vulnerable to overlooking false facts: Do voters who believe to have lost in status or who fear losing status overlook false facts from politicians who promise to move their group up again? Here, too survey experiments around hypothetical elections are a promising route for future research. Future studies could try to nudge feelings of ethnic or global status loss and investigate the effect of that feeling on tolerance of false facts in political campaigns.

Future studies should also tease out the effect of status threat and status boost: Is threatening status sufficient to make voters overlook false claims? Is boosting status sufficient? How do the two interact? What are the minimal conditions under which voters overlook false facts? Could in-group membership or siding with the right team be a sufficient condition? In this case, If none of the teams had been disadvantaged and if both teams had had the same payoff would team members still have overlooked false claims from a person who sided with their team? If so, how far does this willingness stretch: Would minimal (e.g. Klee/Kandinsky)

groups do the trick, too?

In addition, future studies should investigate the effect of *perceived* status differences: Do subject who *believe* to be disadvantaged but are not actually disadvantaged also overlook false claims by leaders who promise to raise their group status or fight off threats to it? Exploratory analyses from this study suggest that the driving factor is not actual disadvantage but perceived disadvantage: Because this study varied the likelihood of receiving difficult or easy questions (but not the number of difficult or easy questions) the level of advantage and disadvantage differed across the games. Both the *somewhat* disadvantaged and the *very* disadvantaged overlooked false claims in the feedback that favoured their team. (This question could easily be answered in a variant of these experiments in which both teams receive the same number of easy and difficult questions; and a flat payoff.)

Finally, scholars ought to investigate the extent to which voters tolerate false claims coming from a person who recognizes their struggles: If they are willing to support someone who makes false but benign claims are they also willing to support them if the claims become less benign? For instance, will they continue to support their candidate if the candidate makes derogatory comments about out-group members, if they condone violence against out-group members, or, ultimately, if they incite violence?

3.6 Acknowledgements

This research was supported by an ESSEXLab Seedcorn grant as well as research funding by the Eastern Academic Research Council. I gratefully acknowledge this support.

Chapter 4

A Post-Truth Public? Investigating the Mechanisms of Resistance to Factual Correction

4.1 Abstract

Many recent electoral events have been characterised by false claims which, despite abundant fact-checking, were often widely believed. This led to much talk about 'post-truth' politics . Meanwhile, an extensive literature confirms that political misperceptions are often highly resistant to correction. But how far does that tendency stretch? Is there any evidence of a 'post-truth' mindset? And how do post-truth surroundings affect the way people respond to expert information – not any expert information but information that challenges political beliefs? We conducted a representative survey experiment in Britain (N=2900) concerning common misperceptions – both liberal and conservative – about immigration. We follow the classic setup of misperception-correction studies but add a

twist: First, we identify false beliefs about immigration and provide expert information countering one of those false beliefs. Second, we seek to approximate 'real world' conditions, where expert information is not the final word: We show respondents a comment from a blogger or a professor giving one of three reasons to 'take these statistics with a big pinch of salt'. Finally, we ask respondents to re-assess the false claims and to answer questions explicitly testing for a post-truth mindset. Results show that fact-checks worked: The expert statement significantly reduced belief in the false fact. However, the post-truth comment worked, too: If the fact-checker did not have the final word – if respondents read a post-truth comment before they re-evaluated the facts – then they kept the false fact on the 'true' side of our scale. The post-truth comment did not however cancel out the *entire* effect of the correction: Even if it stopped our respondents from reaching 'don't know' they still moved closer to the midpoint of our scale. Finally, we find concerningly high levels of agreement with post-truth statements. Well over half of our sample – even those who skipped the post-truth comment – agreed that 'it is OK to disagree with the facts, if that's what you believe.'

4.2 Introduction

Many recent electoral events have been characterised by false claims which, despite abundant fact-checking, were often widely believed. This led to much talk about 'post-truth' politics . Meanwhile, an extensive literature confirms that political misperceptions are often highly resistant to correction. But how far does that tendency stretch? Is there any evidence

of a 'post-truth' mindset? And how do post-truth surroundings affect the way people respond to expert information – not any expert information but information that challenges political beliefs? We conducted a representative survey experiment in Britain (N=2900) concerning common misperceptions – both liberal and conservative – about immigration. We follow the classic setup of misperception-correction studies but add a twist: First, we identify false beliefs about immigration and provide expert information countering one of those false beliefs. Second, we seek to approximate 'real world' conditions, where expert information is not the final word: We show respondents a comment from a blogger or a professor giving one of three reasons to 'take these statistics with a big pinch of salt'. Finally, we ask respondents to re-assess the false claims and to answer questions explicitly testing for a post-truth mindset. Results show that fact-checks worked: The expert statement significantly reduced belief in the false fact. However, the post-truth comment worked, too: If the fact-checker did not have the final word – if respondents read a post-truth comment before they re-evaluated the facts – then they kept the false fact on the 'true' side of our scale. The post-truth comment did not however cancel out the *entire* effect of the correction: Even if it stopped our respondents from reaching 'don't know' they still moved closer to the midpoint of our scale. Finally, we find concerningly high levels of agreement with post-truth statements. Well over half of our sample – even those who skipped the post-truth comment – agreed that 'it is OK to disagree with the facts, if that's what you believe.'

4.3 Introduction

In an interview with the guardian, Arron Banks, co-founder of the Leave.EU campaign explained the 'American-style media approach' his campaign had taken. Quoting the political strategists Goddard Gunster he had hired to help run the campaign, Banks commented, "What they said early on was 'facts don't work and that's it.". "The Remain campaign featured fact, fact, fact, fact.", he said. "It just doesn't work. You've got to connect with people emotionally. It's the Trump success."

Dominic Cummings, campaign director of Vote Leave, told a similar story. In a January 2017 blogpost, he described the foundations of 'our story'. Among them was the infamous £350 million claim: 'The official bill of EU membership is GBP 350 million per week – let's spend our money on our priorities like the NHS instead'.

What is unusual about this campaign is not that some of the claims were factually inaccurate. What is unusual though is the public reaction: The fact that independent fact-checking charities, public broadcasters, and even politicians repeatedly denounced the £350 million as false did not seem to have the slightest effect on public opinion polls. It is exactly this public apathy, this indifference to factual accuracy that is at the heart of a phenomenon commonly referred to as 'post-truth'. Oxford Dictionary defines 'post-truth' as shorthand for "circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief.'" The British journalist Matthew d'Ancona described it as a "crash in the value of truth, comparable to the collapse of a currency or a stock". (D'Ancona 2017, p.8) "Honesty and accuracy

are no longer assigned the highest priority in political exchange". Stanford's Francis Fukuyama describes the 'post-fact world' as a world, in which "virtually all authoritative information sources were called into question and challenged by contrary facts of dubious quality and provenance" (Fukuyama 2017).

This paper contributes to a nascent body of literature investigating the effect of post-truth surroundings on citizen voters in Western democracies. Our goal is two-fold: First, we try to capture the current state of post-truth thinking in Britain: Is there any evidence of a 'post-truth mindset'? Second, we seek to understand the effect of post-truth surroundings on reactions to what ought to be authoritative information: How do people react to expert statements correcting false false claims when they are surrounded by others who use contrary facts of dubious quality and provenance to call their expertise into question? Do fact-checks have any effect at all if they are competing with post-truth comments?

We explored these questions using a representative-sample survey experiment in Britain (N=2,936) concerning common misperceptions – both liberal and conservative – about immigration. We followed the classic setup of misperception-correction studies but added a twist: First, we identified false beliefs about immigration and provided expert information countering one of those false beliefs. Second, we sought to approximate 'real world' conditions, where expert information is not the final word: We showed respondents a comment from a blogger or a professor giving one of three reasons to 'take these statistics with a big pinch of salt'. (Respondents in our control group skipped this step.) Finally, we asked respondents to re-assess the false claims and to answer a few

questions explicitly testing for a post-truth mindset.

Our results showed that expert corrections successfully reduce misperceptions. Post-truth comments cancel out a small but significant part of that effect: Those who saw the expert only shifted from evaluating a false fact as probably true to placing it right between 'true' and 'false'. Those who received the additional post-truth comment placed it as well within the 'true' side of the scale, the second time around. (Albeit closer to 'don't know' than the at the beginning of the survey.) This was true for respondents who were generally sceptical of immigration and for respondents who were generally supportive of it, and it was true for each of the three post-truth comments we tested. Contrary to our expectations, the source of the comment did not moderate its effectiveness: People responded to a post-truth blogger in about the same way that they responded to a post-truth professor. Finally, we found high levels of agreement with two indicators of a post-truth mindset: on a scale from '0 - Purely a matter of fact' to '6 - Purely a matter of opinion' about a third of our sample rated the false fact they had read about as a 'matter of opinion'. 46 per cent agreed that 'It's OK to disagree with the facts, if that's what you believe'. 18 per cent strongly agreed.

4.4 Reactions to uncongenial expert information

The post-fact-check failure of many leave campaigners to drop or adapt the 350 million claim points to a very human trait: we don't like to be proven wrong. An extensive literature confirms that people continue

to rely on outdated information after it is retracted (Ross, Mark R. Lepper, and Hubbard 1975; Carretta and Moreland 1983; Johnson and Seifert 1994; Johnson and Seifert 1998; Wyer and Budesheim 1987; Wilkes and Leatherbarrow 1988; Wilkes and Reynolds 1999; C.A. Anderson 2007; Ullrich K H Ecker, Lewandowsky, Fenton, et al. 2014). False beliefs about political issues are particularly resilient – especially if they are tied to people's opinions (Kull, Ramsay, and Lewis 2003; Thorson 2016; Jacobson 2010; Hart and Erik C. Nisbet 2012; Nyhan, Reifler, and Ubel 2013; Nyhan, Reifler, Richey, et al. 2014; Thorson 2015; Rampell 2016).

The drivers of this resistence to counter-attitudinal facts are well-known: Human beings are motivated reasoners. We engage with new information with a motivation to reach a particular conclusion Kunda 1990. Kahan 2016b defines *politically motivated reasoning* (PMR) as "the tendency of individuals to fit their assessments of evidence to beliefs that cohere with their political identities." In Bayesian terms, the authors argue, it involves deriving the weight one assigns to new information to how congenial it is with one's political beliefs. Westen et al. 2006 provided the first neuroimaging evidence for motivated reasoning. During the 2004 U.S. presidential elections, they conducted experiments in which they asked supporters of Bush and Kerry to consider consonant and dissonant information about both candidates. Magnetic resonance imaging showed that the reasoning areas of partisans' brains shut down when they were considering information that was threatening to their own candidate. When they saw consonant information, the emotional areas of the brain lit up. Hence, a certain proclivity to post-truth reasoning is in our DNA.

Much of the misperceptions research is concerned with finding ways to

make fact-checks more credible. A number of approaches have been proven effective. For instance, wording matters: affirming a correct claim rather than negative a false claim helps reduce misperceptions. Graphs help get across numbers. Sources make a difference: If a correction comes from a person who shares your values or identifies with the same party, then you are more likely to believe their corrective information Nyhan and Reifler 2012. Unlikely sources have also been shown to increase people's willingness to reject misinformation (Berinsky 2015). Earlier studies have detected a need for fluency: False information is often unnoticed if it is part of a fluent story. The much-replicated Moses illusion shows that when asked, "How many animals of each kind did Moses take on the Ark" most people answer "two" – even if they know that the biblical actor was Noah Erickson and Mattson 1981. As a remedy to prompt more careful processing scholars have suggested breaking the fluency of the story: a font that is difficult to read, a less familiar question wording, even a fishy smell helps detect inaccuracies Song and Schwarz 2008; Lee, Kim, and Schwarz 2015. Similarly, if a factual correction leaves a causal gap in the story then this causal gap must be filled – otherwise, people will continue to refer to the outdated information to explain the outcome. For instance, in a variant of the classic warehouse fire story Johnson and Seifert 1994 had their participants read a story about a warehouse fire and retracted a crucial detail: that volatile materials had been stored in a closet. They found that only respondents who were subsequently told that arson materials had been found elsewhere dropped the volatile materials in the closet.

In many ways, post-truth comments do the opposite of what the past few

decades of misperception scholarship has recommended: They smooth out an incoherent story; they add fluency to a false narrative; they provide alternative facts to fill causal gaps. Therefore, we expect them to have the opposite effect of a well-designed fact-check, increasing motivated reasoning, and, thereby, decreasing citizens' willingness to submit to expert information. Given the cognitive and emotional effort it takes to engage with counter-attitudinal information we speculate that hearing even just one person say that that they don't trust the expert will tilt the internal battle between accuracy goals and directional goals in favour of the latter (see Kunda 1990, p. 482 on accuracy goals and Kahan 2016b; Kahan 2016a; Lodge and Taber 2013 on partisan goals). The idea is simple: if I don't like the information to begin with I will take any excuse not to trust it. If this is true then even a seemingly senseless comment such as 'it is best to trust your instincts' will suffice as an excuse not to let the facts get in the way of my opinons.

Our paper is among the first studies that systematically assess the effect of post-truth surroundings on reactions to uncongenial information. We add to a few early studies about belief in human-caused climate change among U.S. Americans. These studies found that exposure to messages that question the authority of experts reduced or even eroded the effect of factual information: McCright et al. 2016 investigated belief in anthropogenic climate change (ACC). They found that a denial counter-frame ("some scientists ... are quick to point out that the Earth hasn't actually warmed in the last decade") significantly reduced belief in the reality of ACC. Linden et al. 2017 looked at the effect of misinformation about scientific consensus about climate change on the persuasive effect of a first,

correct piece of information: They exposed one group of respondents to a persuasive piece of information (a pie chart stating that "97% of climate scientists have concluded that human-cause climate change is happening) and, subsequently, to a piece of misinformation (a petition urging the US government to reject the Kyoto protocol because "there is no convincing scientific evidence that human release of carbon dioxide ... will ... cause catastrophic heating of the Earth's atmosphere ..."). Among this group, seeing the petition completely cancelled out the (positive) effect of the first, correct piece of information. We are not aware of any studies assessing the effect of second opinions or post-truth comments on reactions to expert opinions in the immigration debate or in the UK.

We ask two main research questions.

First, we investigate the effect of post-truth surroundings on reactions to expert information. We ask: Does exposure to post-truth comments affect, or even erode any positive effect of expert information? If so, does the effect of post-truth comments depend on how authoritative the source is, or what they say? Are certain groups of people more vulnerable to post-truth comments than others? Does exposure to post-truth comments create something of a post-truth citizen? Does it spill over and increase realiance on other false facts as well?

Second, we attempt to capture the extent of post-truth thinking in the UK. When facts challenge your opinions how do you deal with them? Do you denigrate them as matters of opinions? Do you doubt the expertise of the expert? Do you agree with statements such as 'it is okay to disagree with the facts, if that's what you believe?

	Authoritative (Professor)	**Not authoritative** (Blogger)
Biased source		
Personal experience		
Alternative facts		

TABLE 4.1: Experimental Design. (Treatment groups only. Control group receives no second opinion.)

We investigated these questions in a population-based survey experiment about an issue on which both liberals and conservatives tend to hold false beliefs: immigration.

4.5 Method

In stage 1, we asked a number of background questions. This block included questions about immigration: Should Britain allow more immigrants, fewer immigrants, or should the number of immigrants stay the same? And how important is this issue to you? In a separate question, we asked how important this issue was to the country: Respondents were shown a list of seven items, including immigration They were asked to pick and rank the three most important issues facing Britain today.

In stage 2, we assessed belief in false facts about immigration. Respondent were shown a list of four commonly held misperceptions as well as four distractor statements. Respondents rated each claim on a scale from 'definitely false' (0) to 'definitely true' (6). Two statements buttressed sceptical views about immigration: First, "There has been a sharp rise in the number of people applying for asylum in the UK in the past ten

years." Second, "European immigrants receive more in benefits and services than they pay in taxes." Another two statements buttressed immigration-friendly attitudes: Third, "The majority of crimes in London are committed by white people, not ethnic minorities." Fourth, "Immigration to the UK does not affect the wages of the low-paid.". The distractor items were unrelated to immigration and included items such as "Fracking causes earthquakes" or "England's plastic bag usage dropped 85% since the 5p charge was introduced."

In stage 3, respondents were shown "some detailed information about one of those statements". They saw a short statement ascribed to an Oxford professor that challenged whichever false fact they thought was true. (Answers from stage 2 were used to direct respondents to an article correcting the false fact they gave the highest veracity score. If they gave two or more statements the same veracity score they were randomly assigned to see more information on one of them.) The statements reported official statistics that directly corrected each misperception and added a little more commentary in order to approximate the way that expert evidence is presented by public service broadcasters and fact-checking organisations. They were designed to be as effective as possible, following recommendations from research about correcting mispercpetions (Nyhan and Reifler 2013). All four factual corrections were attributed to the same authoritative source: A professor of Economics at the University of Oxford, and a former Consultant on Population and Demography at the Office for National Statistics (ONS). All articles included graphical information. The false claims were not repeated. Instead, the correct claims were affirmed. Wherever possible, an alternative causal explanation was

provided to displace false causal explanations. (For instance, to counter the claim that the UK had experience a high influx of asylum seekers the article conceded that it was true that there had been a sharp rise of asylum seekers crossing into Europe. Most of them did not, however, reach the UK – and this was simply because they stayed in Germany, Sweden, or Hungary. [1] This section was followed by a battery of demographic questions, designed to distract respondents.

In stage 4, we tested how resilient factual corrections were against fact-free dispute. At this point, the sample was divided: The control group (25% of the sample) moved directly to Stage 5 and received no more opinions about the statistics they had just seen. The treatment groups (75% of the sample) was exposed to a second opinion, saying that 'I would take these statistics with a big pinch of salt'. This comment was ascribed to either a Professor at the London School of Economics or a Blogger. They did not directly contradict the expert's facts. Instead, they used post-truth arguments to encourage respondents to retain their original beliefs. The second source gave one of three reasons to doubt the first source: In the first variant, they suggested the expert was biased. In the second variant, they suggested that the expert's facts were at odds with their own personal experience. In the third variant, they argued that there were so many contradictory facts out there that it was best to trust on one's own feelings.

In Stage 5, we asked our respondents to re-evaluate the validity of the corrected claim. To measure if our factual corrections affected un-corrected

[1] Extensive pre-testing was carried out to identify commonly held misperceptions on issues which respondents deemed important and to identify credible sources of corrective information.

false claims we had respondents re-assess the second false claim that buttressed their immigration opinions, too.

In Stage 6, we assessed signs of post-truth reasoning. We asked three questions: First, was the (false) statement they saw a matter of fact or a matter of opinion? Second, we asked if the statistics about that statement were consistent with what respondents believed. (If not, we followed up and asked them to assess the veracity of the statistics.) The third question was the most upfront one: Is it 'OK to disagree with the facts if that's what you believe'?

Finally, in Stage 7, we ask two questions about the two sources, the Oxford professor giving factual information that disconfirmed prior beliefs and (for the treatment groups) the LSE professor or the blogger giving fact-free comments confirming prior beliefs: (1) How accurate would you say was the information they used and (2) how much would you generally trust what they say on the issue of immigration?

4.6 Hypotheses

An assumption more than a hypothesis is that we assume to see 'fact polarization' (Kahan 2016b): a partisan divide over questions of fact. We expect people to polarise along the lines of immigration attitudes: We expect those who are, generally, sympathetic to immigration to be more likely to believe in one of the two false facts that bolster pro-immigration attitudes. Similarly, we expect those who are, generally, sceptical of immigration to be more likely to believe in one of the two false facts that

bolster anti-immigration attitudes.

Effect of the fact-check on belief in false facts

Following many misperception-correction studies (e.g. Nyhan and Reifler 2015a; Nyhan and Reifler 2015b; Wood and Porter 2016; Barrera et al. 2018) we expect our factual corrections to reduce, but not to eliminate respondents' misperceptions.

H1. On average, respondents move closer to the 'false' side of the 'true' to 'false' scale after they are exposed to the factual correction.

Effect of post-truth comment on reactions to the fact-check

We expect any negative comment about an unwanted piece of information to increase motivated reasoning biases and, thereby, to decrease respondents' willingness to adapt their factual beliefs to the expert information:

H2. Exposure to post-truth comments moderates the effect of the fact-check. We test the impact of three types of post-truth arguments (a) the source of the information is biased, b) the information does not match the commenter's personal experience and c) 'it is best to trust one's instincts even if it looks like the facts are different'. We expect each of these to shrink the gap between pre- and post-correction veracity scores: Among the treatment groups, the difference between pre-and post-correction ratings [diff] is smaller **(a)**, trust in the source of the authoritative correction

trustAuthority] **(b)**, and perceived accuracy of their information [accurateAuthority] **(c)** is lower, on average, than among the control group (see list of **??** in the appendix).

We know that credentials matter: Reputable sources are more effective at correcting false perceptions that not so reputable sources (Berinsky 2015; Nyhan and Reifler 2015a; Schaffner and Luks 2018). We expect the same to hold for post-truth comments.

H2d. Authoritativeness moderates the effect of the factual correction. [diff], [accurate], and [trust] are smaller among respondents who were exposed to a post-truth comment attributed to the LSE professor than among respondents who were exposed to a post-truth comment attributed to a blogger.

Spillover effects

A question that has not received any attention in misperception-correction studies is that of potential spillover effects. We test if a) the factual corrections and b) the post-truth comment have any effect on belief in other, uncorrected false claims.

Intuitively, one might not expect expert advice about one item to affect beliefs about any other, unrelated items. However, given research findings about beliefs being linked (Lewandowsky, Oberauer, and Gignac 2013; Friedkin et al. 2016) it is possible that expert advice on one issue might indeed stretch over and affect other misperceptions about the same

issue. Two mechanisms are plausible: On the one hand, corrective information could make people more alert and, thereby, make them reconsider other beliefs (a). On the other hand, being proven wrong on one item could increase reliance on other information, including wrong information, to justify one's opinions (b). We test both mechanisms:

H3a. Factual corrections *decrease* belief in related false claims. Respondents who rate two false claims as true (4, 5, or 6 on a scale from 0 (false) – 6 (true)) report *lower* levels of belief in the uncorrected false claim is after they read the correction than it is before they read it.

H3b. Factual corrections *increase* belief in related false claims. Respondents who rate two false claims as true report *higher* levels of belief in the uncorrected false claim is after they read the correction than it is before they read it.

The second spillover effect we investigate is that of the post-truth comment. If post-truth comments encourage citizens to disregard expert information about one issue then the effect of these comments may stretch beyond beliefs about that one issue. We suspect that post-truth surroundings make people more suspicious of expert information and increase motivated reasoning biases. There are two ways in which this would become apparent: First, exposure to post-truth comments would decrease peoples' willingness to accept expert advice on other issues. (Our design does not allow to test if this true.) Second, exposure to post-truth comments would encourage people to hold on to other false beliefs they hold. This is the hypothesis we test. We assume that fact-checks that challenge one's beliefs elicit a battle between accuracy goals and confirmation

goals: On the one hand, people have a desire to arrive to correct conclusions (Kunda 1990). On the other hand, people engage in motivated reasoning to justify political attitudes or to show loyalty to groups they belong to. Post-truth comments do not suggest that experts are wrong in as much as they question the existence of a single correct answer or downplay its relevance. We expect them to shift the battle in favour of the confirmation bias, silencing the voices telling you that a false fact is fishy and amplifying the ones telling you that it's okay to ignore the expert.

H3b. Fact-free comments spill over. Respondents who rate two false claims as true and who are exposed to a post-truth comment report higher levels of belief in the *uncorrected* false claim than their peers who are not exposed to a post-truth comment: [diff_i2] is smaller for the treatment groups than it is for the control groups.

Evidence of post-truth reasoning

This study attempted to get a glimpse of the state of post-truth reasoning in the UK. To that end, we asked two questions: First, is there any evidence of a post-truth mindset among British voters? Second, what do post-truth surroundings iincrease post-truth reasoning?

The first question is exploratory. We rely on our control group – that is, individuals who were not exposed to a post-truth comment – to see to what extent UK citizens agree with post-truth arguments. We ask: Do respondents designate the statements as matters of opinion rather than fact? Do they notice if statistics are not in line with their prior beliefs?

If so, do they believe the statistics are correct? Do they believe they are false? Or do they believe the statistics are correct but, nonetheless, believe something different? What proportion of UK citizens are willing to say that it is 'okay to disagree with the facts?'

4.7 Results

Deltapoll fielded the survey to a representative sample of 2,936 British citizens across England, Scotland, and Wales in June 2019. 54% of the sample was female, 51% was university-educated, aged 18 to 99 (mean age: 44 years). At the last general election in the UK in 2017, 26% of the sample had voted Conservative, 30% Labour, 5% Liberal Democratic, 6% UKIP, and 3% had voted for the Green Party.

As typical for a UK sample most respondents were sceptical of immigration: On a scale from -3 to +3, where -3 meant 'many fewer', 0 meant 'no change', and +3 meant 'many more', 47 per cent said Britain should take in fewer immigrants (that is, values between -3 and -1), 32 per cent said the number of immigrants should stay the same, and 21 per cent said Britain should take in more immigrants (values between +1 to +3). About a third of the sample (30 per cent) ranked immigration as one of the top three issues facing Britain today; 9 per cent, overwhelmingly immigration-sceptics, ranked it as *the* most important issue facing Britain today. (Thereby, immigration ranked after the EU/Brexit (31 %) the NHS/Health (28%), crime (10%), and the economy (9%), but before housing (7%) and schools/education (6%) as the most important issue facing Britain today.)

Generally, respondents showed higher levels of belief in the anti-immigration claims than in the pro-immigration claims. Consequently, the number of respondents who read each of the four expert statements differed substantially. Most respondents saw the correction to one of the two anti-immigration claims: 40 percent read the article about the false claim that the number of asylum seekers had risen sharply in the last 10 years; 24 per cent read the article about the false claim that immigrants receive more in benefits than they pay in taxes. Far fewer saw the correction to one of the two pro-immigration claims: 20 per cent read the article about the false claim that immigration to the UK does not affect the wages of the low paid, and only 16 per cent read the article about the false claim that the majority of crimes in London were committed by white people (see figure **??**).

Pre-correction belief in the false claims was fairly high. 40 per cent of respondents were shown statistics about a statement they had rated as '6 - definitely true'. A further 23 and 24 per cent saw statistics about a statement they had rated as '5' or '4', that is, closer to 'true' than to 'false', and 12 per cent were given information about a statement they were not sure about ('3' on the scale from 0 to 6). Less than one per cent received statistics about a statement they had rated as 2, i.e. closer to 'false' than to true. (No one saw statistics about a fact they had rated as 1 or as '0 - definitely true'.) Hence, the overwhelming majority of respondents received statistics that disconfirmed their factual beliefs.

In most cases, the statistics also challenged respondents' political beliefs. Most of the immigration-sceptics (83 per cent) gave the highest veracity score to one of the two false facts that were catered to immigration

sceptics and, therefore, read an expert statement refuting a false anti-immigration claim (i.e. the claim that the number of asylum seekers had risen sharply in the last ten years, or the claim that immigrants receive more in benefits than they pay in taxes). Most of the immigration-friendly respondents (58 per cent) gave the highest veracity score to one of the two false claims that were catered to immigration-friendly citizens and read an expert statement correcting a false proimmigration claim (i.e. the claim that the majority of crimes in London were committed by white people, or the claim that immigration did not affect the wages of the low paid). These individuals had a motivation to reject the expert's statistics.

Notably, this was not the case for all respondents. Those who gave the highest veracity score to one of the false facts catered to the 'other side' also received a correction catered to the 'other side': Among those who thought Britain should take in fewer immigrants, 17 per cent received an article correcting one of the *pro-immigration* claims (7%: white crime; 9% effect of immigration on the low paid). Among those who thought Britain should take in more immigrants 42 per cent received an article correcting one of the a*anti-immigration* claims (26%: number of asylum seekers; 16%: cost of immigration). This is consequential: These respondents received an article that corrected their factual beliefs but that did not challenge their opinions about immigration. They had every motivation to believe an expert telling them they were wrong. Therefore, out main hypotheses (concerning the effect of the post-truth comment, H2a-d, H3b, H4a-d) do not apply to these individuals. They were excluded from the respective analyses.

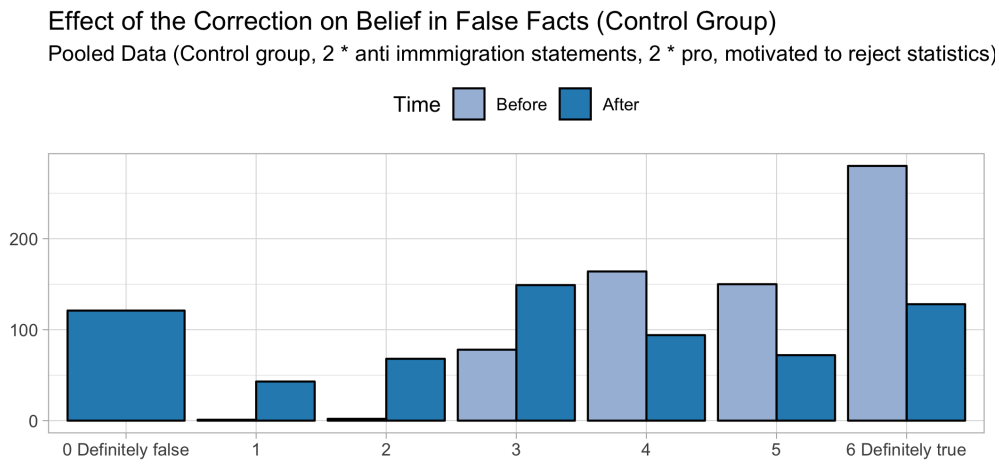


FIGURE 4.1: Effect of the correction on veracity scores (pooled data)

Effect of the correction on belief in false claims

As expected, the expert statement reduced belief in false facts, but did not eliminate it. The histogram in figure 4.1 shows how true or false respondents rated the statements before and after they saw the correction. The figure refers to respondents in the control group (that is, respondents who saw the correction and no further comments) and who were motivated to reject the statistics because they challenged their immigration opinions. On average, ratings shifted from a mean of around 5, that is, just about a point under 'defnitely true' (m=4.91, sd=1.09) to a mean of around 3, that is, the mid-point of the scale (m=3.21, sd=2.05).

Pre-correction belief in the two anti-immigration claims was slightly higher than pre-correction belief in the two pro-immigration claims. Control-group respondents who were sceptical of immigration and who saw statistics refuting a false claim that bolstered sceptical opinions evaluated the false claims as 5.05 the first time around – just a point below the extreme

end of the scale (6 - Definitely True). The second time around, they averaged at 3.09, just about the midpoint of the scale. The mean paired difference was 1.95 points (95% confidence interval: 1.75-2.15, t(393)=18.87).

Control-group respondents who were favourable of immigration and saw statistics challenging a false claim that bolstered favourable opinions started out at 4.83 points, on average, and shifted to 3.33. Here, the mean paired difference was 1.50 points (95% confidence interval: 1.22-1.78, t(210)=10.668). In both cases the factual correction made respondents move from being relatively certain that a false claim was true, to not being certain at all, hovering right between 'true' and 'false'. Given that this was a one-shot exposure to a fabricated expert statement we take this to be a successful correction.

Figure 4.2 shows the veracity scores for each individual statement. It shows slight variations in how effective the four expert statements were at reducing misperceptions. The most convinving statement was the one about the number of asylum seekers in the UK having remained constant over the past ten years (m=2.05, sd=2.12); the least convincing was the statement about the effect of immigration on wage levels in the low-income sector (m=1.23, sd=1.75).

One variable predicted the effect of the correction to almost every false fact: All other things equal, being female increased the difference between pre and post-correction veracity scores by about half a point for the two pro-immigration claims and by over a point for the asylum seekers claim (it did not have a significant effect on the cost/benefit claim, see OLS regression models in table **??**).

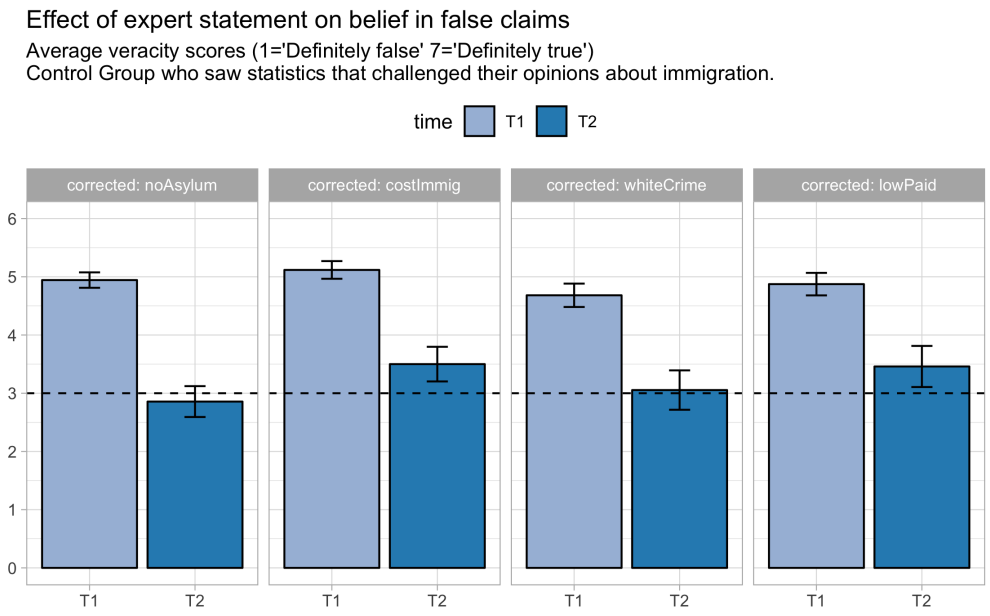


FIGURE 4.2: Effect of the correction on mean veracity scores

Effect of post-truth comments

Eighty pecent of our sample saw not only an expert statement correcting a false fact but also a post-truth comment casting doubt on the expert. To mirror the different kinds of post-truth comments 21st century citizens might be exposed to we varied the reason why the commenter said they would 'take these statistics with a grain of salt': They argued that a) the expert was biased, b) the expert's statistics did not match their personal experience, or c) that when it doubt it was best to trust one's instincts (see questionnaire in online appendix). Did these comments affect respondents' willingness to accept the expert advice they had heard earlier on? They did.

Among immigration-sceptics who saw an expert statement challenging an anti-immigration claim the post-truth comment reduced the difference

between pre- and post-correction veracity scores by about a two thirds of a point (0.73): As mentioned above, those who saw the expert information and no further comment shifted from 5.05 to 3.09 – a 1.95 point difference. Their peers who read the expert information followed by one of the post-truth comment shifted from being fairly certain the false claim was 'true' (5.04) to to being somewhat less certain, but still rating it as 'true (3.82) – a 1.22 point difference. The mean difference in the effect of the expert information on those who did not see the post-truth comment and those who did see it was statistically significant ($M_{\text{diff in control group}}$=1.95, $M_{\text{diff in treatment groups}}$=1.22, t(566)=6.4269, p = 0.00).

Among immigration-friendly respondents who saw an expert statement challenging a pro-immigration claim the post-truth comment reduced the effect of the correction by about one third of a point (0.33): As mentioned above, those who were only exposed to the expert statement shifted from 4.83 to 3.33 – a 1.50 point difference. Their peers who read one of the post-truth comments in addition to the expert statement moved from 4.71, on average, to 3.55 – a 1.16-point difference. Again, the difference in means was statistically significant ($M_{\text{diff no comment}}$=1.50, $M_{\text{diff comment}}$=0.15, t(308)=22037, p = 0.03). Therefore, we accept hypothesis 2a: Exposure to a post-truth comment reduced the effect of the correction on belief in false facts. [2] The content of the message did not affect its effectiveness. All three variants of the post-truth comment reduced the positive effect of the correction significantly and by about the same amount. (Hence, for further analyses

[2]NB: The post-truth comment also reduced trust in the expert (H2b) and in their information (H2c) – results for these dependent variables are shown in the appendix.

the three types of statements were pooled.) [3]

In contrast to earlier studies (Linden et al. 2017), the post-truth comment did not cancel out the *entire* effect of the correction: Respondents who read the comment did still adapt their factual beliefs, moving closer to 'don't know'. However, they fell short of reaching 'don't know'. Hence, the comment constrained the positive effect of the expert statement. It caused respondents to place a corrected false fact well within the 'true' side of the scale when the expert statement on its own would have led them to position the false fact between 'true' and 'false'.

Contrary to our expectations, the source of the post-truth comment did not seem to matter: An authoritative source did not do any more harm than a non-authoritative source. What is more, the small differences that we did observe went in different directions: Those who were sceptical of immigration seemed to be slightly more responsive to non-authoritative sources while those were are generally supportive of immigration seemed to be slightly more responsive to authoritative sources. On average, respondents who thought Britain should let in fewer immigrants were 0.16 points more responsive to a professor as compared to a blogger; respondents who were content with current levels of immigration or who thought Britain should take in more immigrants were 0.2 points more responsive to a blogger as opposed to a professor. Nonetheless, neither of those

[3] $M_{\text{Biased statistics comment}}$=1.14, $M_{\text{No comment}}$=1.95, t(-6.21)=-6.21, p = 0.00; $M_{\text{Personal experience comment}}$=1.24, $M_{\text{No comment}}$=1.95, t(765)=-5.31, p = 0.00; $M_{\text{OK to disagree comment}}$=1.27, $M_{\text{No comment}}$=1.95, t(750)=-5.11, p = 0.00.

two group differences reached statistical significance. [4] Hence, we reject hypothesis 2b: authoritativeness did not moderate the effect of the post-truth comment.

Spillover effects

How far does the effect of a fact-check stretch? If people hold multiple false beliefs that corroborate their immigration opinions does being proven wrong about one of them make them question others, too (H3a)? Or does it make them hold on to other, false facts (H3b)? We find evidence for the former hypothesis. Among respondents who rated the two false facts on 'their' side of the immigration debate as 'true' seeing the expert correction on the one false fact decreased belief in the other false fact, as well. The effect was slightly smaller among those with sceptical views of immigration than among those with favourable views: Immigration-sceptics moved 0.28 points closer to 'false' (paired t-test, 95 per cent confidence interval: 0.13-0.43, t(206)=3.75, p=0.00). Immigration-friendly respondents moved 0.43 points closer to 'false' (95 per cent confidence interval: 0.19-0.68, t(95)=3.53, p=0.00). This finding is encouraging: Fact-checks are not only an effective measure of countering one false fact – they even diminish belief in related, but un-corrected false facts.

[4] Respondents who thought Britain should take in *fewer* immigrants, who believed in one of the false facts that confirmed this opinion, and who received statistics refuting that claim lowered their true/false score by 1.14 points after seeing the statistics and a comment from a blogger saying they should take those with a grain of salt. Their peers who saw the comment from a professor lowered their scores by 1.30 points, (t(1224)=-1.77, p = 0.08. Respondents who thought Britain should take in *more* or the same number of immigrants, who believed in one of the false facts that confirmed this opinion, and who received statistics refuting that claim lowered their true/false score by 1.26 points after seeing the statistics and a comment from a blogger saying they should take those with a grain of salt. Their peers who saw the comment from a professor lowered their scores by 1.06 points, t(611)=1.51, p=0.13.

A second type of spillover effects we investigated were those of the post-truth comments. Do post-truth surroundings increase belief in un-corrected and un-commented false facts? At the beginning of the survey our respondents evaluated four false facts – two that corroborate sceptical views about immigration and two that corroborate positive views. At the end of the survey – after having seen the fact check on one of those items (and, if applicable, a post-truth follow-up comment) respondents re-evaluated the two false facts that corroborated their own opinions: Immigration-sceptics reviewed the claim that the number of asylum seekers in the UK had risen sharply over the past ten years and the claim that the cost of immigration is higher than its benefits. Immigration-friendly respondents reviewed the claim that most crime in London is committed by white people, not ethnic minorities, and the claim that immigration has no effect on salaries at the lower end of the wage distribution. To see if post-truth comments increase belief in other false facts about immigration we looked at a subset of respondents who rated both false facts that confirmed their opinions as 'true' and who saw a post-truth comment (n=630 sceptical; n=256 favourable). We compared their first evaluations to their second evaluations of the un-corrected claim – expecting that the post-post-truth comment ratings of false fact II would be higher, on average, than the pre-post-truth comment ratings of false fact II. We found evidence for the opposite. T Our respondents who subscribed to multiple false facts Our respondents did not become any more confident that the second false fact was true: they became *less* confident that it was true. Respondents who saw one of the false claims that put immigration in a

bad light rated the second false claim as 0.32 points closer to 'false' after they saw the fact-check and the post-truth comment than before (95% CI: 1.17-0.48, t(192)=4.15, p=0.00). Respondents who saw one of the false claims that put immigration in a good light rated the second false claim as 0.24 points closer to false after they the second time around (95% CI: 0.02-0.47, t(81)=2.13, p=0.04). This is surprising: It suggests that the effect of fact-checks is stronger than the effect of post-truth comments. Hence, we reject hypothesis 3b.

Next, we compared the effect of the expert statement on belief in un-fact-checked claims among those who saw the post-truth comment and those who didn't. We did observe slight differences: On average, the difference in pre- and post-correction evaluations was about 0.1 points larger among the control group (who didn't see the post-truth comment) than among the treatment group. However, these group differences did not reach statistical significance. (Those who saw the expert statement only adapted their factual beliefs about the second false fact by 0.33 points; those who saw the post-truth comment adapted their factual beliefs about the second false fact by 0.22 points, t(13656)=1.36, p=0.17.

We conclude that the effect of fact-checks go beyond the fact-checked item: They have a small but significant effect on belief in related false facts – even if competing with post-truth surroundings.

Evidence of post-truth reasoning

Our final quest was to capture the extent of post-truth thinking in the UK We rely on our control group (respondents who saw the expert's statement but no second opinion on it) to explore to what extent British citizens agree with classic post-truth arguments.

First, we asked respondents to rate the statement they had just read an article about on a scale from '0 - purely a matter of fact' to '7 - purely a matter of opinion'. Less than a fifth (17 per cent) rated it as '0 – purely a matter of fact'. 37 per cent thought it was closer to a 'fact' than an 'opinion' (rating it as 0, 1 or 2) while 27 per cent were undecided (3). 35 per cent deemed it closer to a 'matter of opinion' (4, 5, or 6) with 13 per cent saying it was 'purely a matter of opinion' (6) (See the graph on p. B.2a of the appendix). Which of the four statements respondents saw did not seem to affect their opinions about where the statement fell on a

The distribution was very similar across respondents evaluating the four statements that challenged false claims on both sides of the spectrum of immigration attitudes: The percentage of respondents who thought the statement was a 'matter of opinion' only varied between 32 (whiteCrime) and 35 per cent (costImmig). It varied slightly more depending on opinions about immigration: Among those who wanted to raise immigration, 35 per cent said the statement they saw was a matter of opinion; among those who wanted to curb immigration, 40 per cent said the statement they saw was a matter of opinion.

Second, we asked if 'the statistics here were consistent with what you believed.' Among those who had rated the false claim they were corrected

on as 'true' 46 per cent said 'Yes'. Next, we asked those who had said 'Yes': 'Which of these best describes where you stand?'. We gave them three options: 'The statistics are probably right, but I believe something different'; 'I think that the statistics are wrong', and, finally, 'The statistics made me change my mind'. 28 per cent agreed with the first, post-truth argument; 49 per cent thought the statistics were wrong, 24 per cent said they changed their mind. Answers to this question ought to be interpreted with caution: We gave our respondents two admittedly extreme choices: changing their mind or doubting the facts, with a post-truth in-between option. Our goal in forcing them to decide between these three options was to distinguish among those who knew that the statistics were correct: We wanted to see how many of them would rather say that they disagree with the facts than admit that they were proven wrong. We are aware that providing more alternatives (such as, for instance, an option saying they needed more information, or a don't know option) would have decreased the percentage agreeing with the post-truth option.

Finally, we asked if respondents agreed that "It's OK to disagree with the facts if that's what you believe". 18 per cent strongly agreed, 46 per cent agreed, 28 per cent disagreed, and 7 per cent strongly disagreed.

Notably, exposure to post-truth comments did not make people any more likely to rate the corrected false claim as as a matter of opinion rather than a matter of fact or to agree that it is 'okay to disagree with the facts if that's what you believe: There was no significant difference in fact/opinion ratings and in agreement that it is okay to disagree with facts. We suspect that this is due to ceiling effects: Given the extraordinarily high levels of agreement with these post-truth statements in the control group there

simply wasn't any scope for the post-truth comment to increase agreement any further.

Limitations

A few limitations ought to be noted.

One concerns our sample: We had to exclude a significant part of our sample (n=489/2936, i.e. 16 per cent) from the analyses: respondents who believed in false facts on the 'wrong' side of the immigration debate and, therefore, had no motivation to disbelieve the expert that proved them wrong. The main culprit was our asylum claim: Respondents on both sides of the spectrum evaluated the claim that the number of asylum seekers had risen sharply in the past ten years as true. Obviously, this affected the representative nature of our sample. (Given the large size of our sample this is not dramatic: Our purpose was not primarily to estimate the precise extent of fact-avoidance in the electorate, but to explore how individuals react to facts when they have the motivation to believe the post-truth comment rather than the fact.) Nonetheless, future researchers ought to be careful to use false facts that discriminate between opposing camps in a political debate.

Furthermore, the post-truth mindset we encountered in this experiment ought to be interpreted with caution. Our aim was to explore if exposure to post-truth surroundings created something of a post-truth citizen. The people who answered our post-truth questions had just come across information that disconfirmed their factual beliefs and their opinions. Due

to financial contraints we were unable to include a control group that would only have answered the post-truth questions.

Another limitation lies in the difficulty of studying the effect of post-truth surroundings in a 2019 environment where all respondents will have been exposed to some sort of a post-truth surroundings. We are well aware that it is impossible to isolate the effect of post-truth surroundings. Our control group is not a pure control group; and our post-truth comments did not create post-truth surroundings. They made them more salient. Our study provides a first snapshot of post-truth thinking among motivated reasoners in Brexit-era post-truth Britain. To fully understand the extent of post-truth thinking and reactions to post-truth comments we need panel data covering times of high and low post-truth popularity.

What is more, our study was confined to the effect of post-truth comments on belief in debunked false facts. We did not measure how post-truth surroundings affect political beliefs that are based on false facts. It is well known that accepting factual corrections does not usually make people change their political beliefs Gaines et al. 2007. Future studies ought to examine the effect of post-truth surroundings on self-exposure to expert information, on the way people process expert information, on perceptions about beliefs on the other side of the political divide, on willingness to engage with people on the other side of the debate and, ultimately, on political opinions. Creative experimental designs experiments are needed to understand how exposure exposure to post-truth surroundings affects political behaviour in as close to real-world settings as possible. Finally, future research ought to investigate ways in which the effect of post-truth surroundings on reactions to expert information

can be restrained.

4.8 Discussion and Conclusion

This project set out to explore the effect of post-truth surroundings on reactions to unsolicited expert information that contradicts prior beliefs. We chose immigration as a salient, and important issue on which voters across party lines hold misperceptions. In a nationally representative survey we tested belief in false facts and exposed respondents to an authoritative expert statement correcting the one false fact people rated as closest to 'definitely true'. To mimick real-world settings in which news rarely go un-commented and, in particular, to mimick the post-truth era we then followed up with a post-truth comment, which we attributed to either a highly authoritative source (a Professor of Economics at the LSE) or a not authoritative source (a Blogger). We varied the reason the gave to 'take these statistics with a pinch of salt': The blogger/professor either suggested that 'just because someone is a professor doesn't mean they don't have an agenda', or that 'that doesn't sound like the world I live in' or, simply that 'there is so much information out there that it can be hard to know what to believe. In that case, I think it's best to trust your instincts even if it looks as if the facts are different.'

Our findings have three important implications: First, not all hope is lost: contrary to many correction-misperception studies, we find that citizens *do* listen to expert information. They may not shift from rating a false fact as 'Definitely true' to rating it as 'Definitely false'. However, our nuanced scale shows that fact-checks do move people away from the 'Definitely

true' end of the scale and closer to 'don't know': On average, respondents shifted from 5 – just a point below 'Definitely true' – to 3, the midpoint of the scale. We are not at all alarmed that they did not shift any further toward the 'Definitely false' side of the scale. Given the relevance of priors in information processing we are perfectly content to see citizens move from thinking they know a false fact is true to knowing they do not know if it is true or false. (We would of course hope to see citizens reach out for more information and, eventually, after having seen more reliable information move toward rating a false facts as 'false'.

Second, people are not immune to post-truth surroundings. Our respondents did not *only* listen to experts providing expert information; they also listened to non-experts providing non-expert information. Wherever the expert information was followed up by a post-truth comment respondents were less willing to adapt their factual beliefs. On average, the post-truth comment cancelled out about half a point of the effect of the expert statement (.33 points among respondents on the favourable side of the immigration debate and about .73 points among respondents on the sceptical side of the debate). This effect is not dramatic: Contrary to prior research (Linden et al. 2017) our second opinion did not cancel out the *entire* effect of the correction. It was, nonetheless, consequential: The post-truth comment caused respondents to remain on the 'true' side of the scale where the expert information on its own would have led them to shift to evaluating a false claim as right between 'true' and 'false'. Notably, people seem to respond to *any* comment telling them to disregard the expert – regardless of how authoritative the source is or what they say. When we attributed the comment that "I would take these statistics

with a pinch of salt" to a professor it did just as much damage as when we attributed it to a blogger. And when they continued saying that the expert was biased they had the same effect as when they continued saying that the statistics did not match their personal experience or that 'I think it's best to trust your instincts even if it looks as if the facts are different'.

Third, we find a level of agreement with post-truth arguments that supports the notion a post-truth public. We asked our respondents to rate the false fact they had just read about on a scale from '0 - purely a matter of fact' to '6 - purely a matter of opinion'. 35 per cent deemed it closer to a matter of opinion than a matter of fact (i.e. rating it as 4, 5, or 6), with 13 per cent evaluating it as '6 - purely a matter of opinion'. Next, we asked if the statistics were in line with what they believed. If they were not we followed up, asking respondents to choose one of three statements to describe where they stood. 32%) said "The statistics are probably right but I believe something different." Finally, we presented respondents with the statement that "It's OK to disagree with the facts if that's what you believe." 46 per cent agreed. 18 per cent strongly agreed.

In his blog, Dominic Cummings wrote: "Would we have won without immigration? No. Would we have won without £350m/NHS? All our research and the close result strongly suggests No. Would we have won by spending our time talking about trade and the Single Market? No way." (Cummings 2017). Democracies depend on voters' ability to aquire accurate information. If politicians use false facts and continue to use them after they have been debunked then this information environment places a great deal of weight on voter's shoulders. If online newspapers, blogs, and social media are slowly replacing state-regulated media and

if, as we found, virtually any post-truth comment erodes trust in experts, then it is not surprising that it has become increasingly difficult for expert information to reach the masses. And if, as we found, the line between fact and opinon has become blurry, if voters believe it is okay to disagree with facts, if that's what you believe, then this does not bode well for the future of democratic decision making.

Our findings allow for one conclusion: Fact-checking is as important as ever. We find that even under the most post-truth circumstances, experts are still heard. We have exposed citizens to an expert statement that disconfirmed their factual beliefs, challenged their political beliefs, and followed up with a post-truth comment saying it was okay to ignore the expert. And yet, the expert statement made them shift slightly closer to 'don't know'. We conclude that the battle is not lost – but it requires some serious fighting. Fact-checking charities and public broadcasters such as the BBC Reality Check are fighting Goliath – they deserve our gratitude and they deserve our public funding.

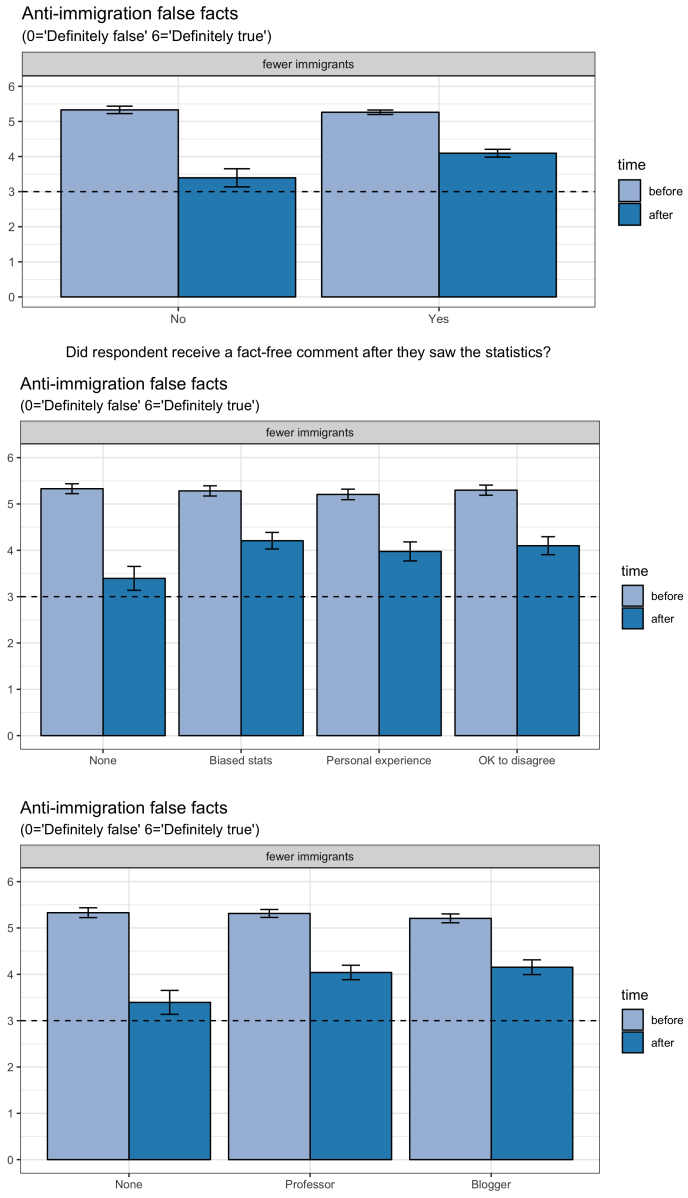


FIGURE 4.3: Effect of the fact-free comment on belief in anti-immigration false facts

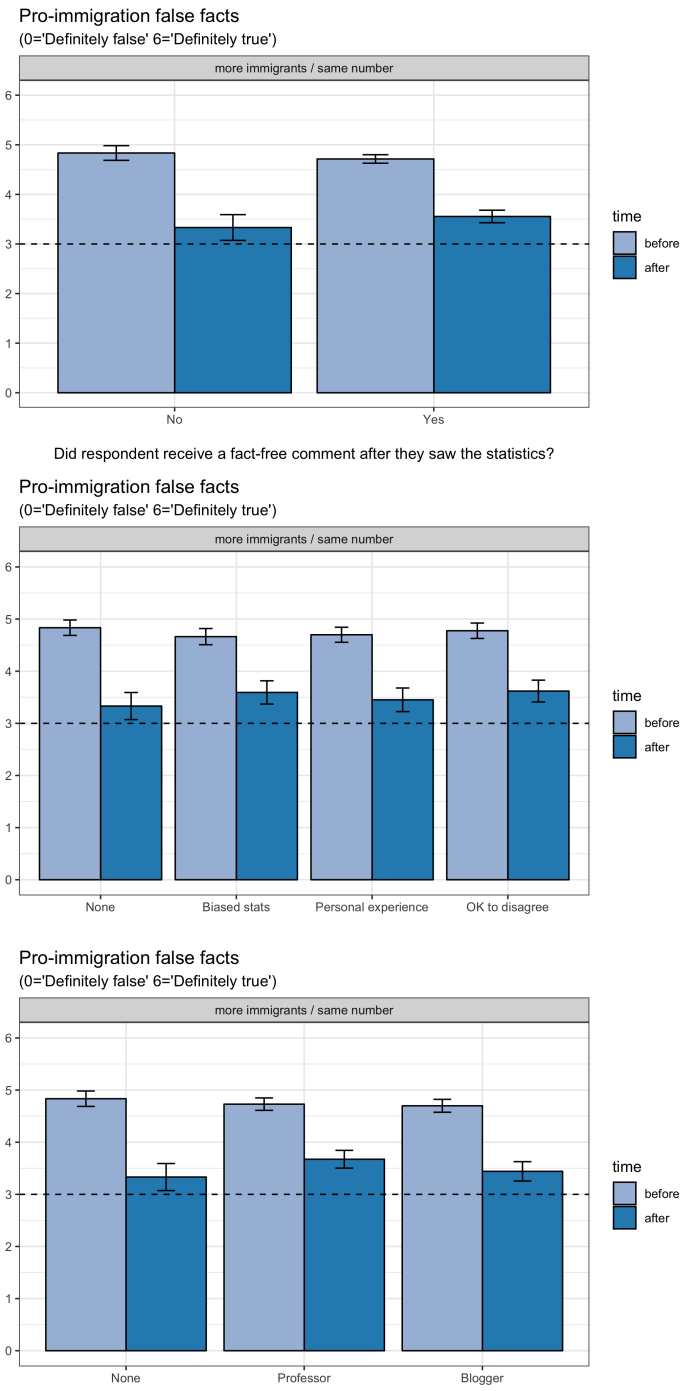


FIGURE 4.4: Effect of the fact-free comment on belief in pro-immigration false facts

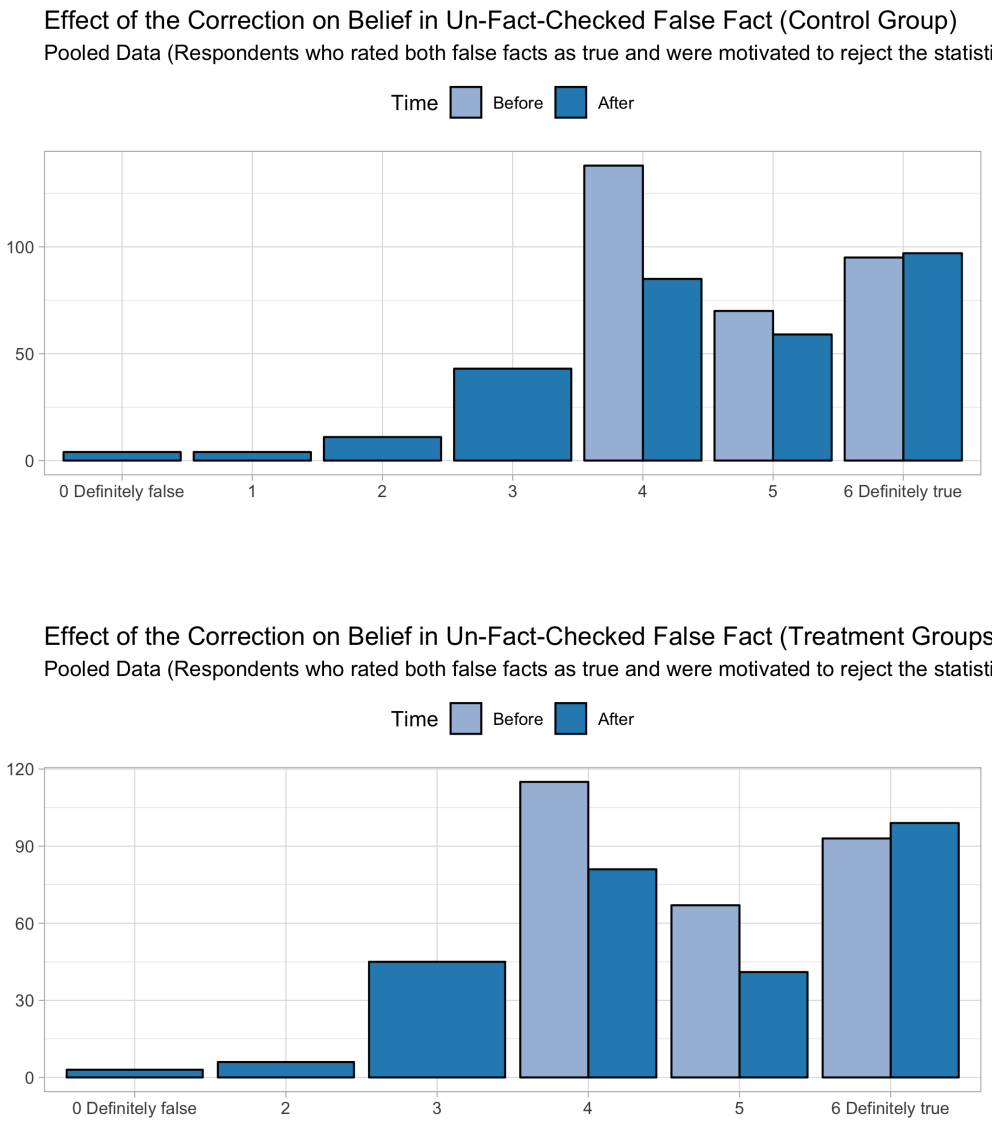


FIGURE 4.5: Spillover effects

Chapter 5

Conclusion

If the recent outcy over post-truth politics has shown one thing, it is that voters – not all, but many – notice deviations from the norm of truthfulness in political speech. Enough of the public has taken offence for this phenomenon to be named: *post-truth*. And if people takes offense at political campaigns printing false numbers on their buses, at state-regulated media channels spreading fake stories, at politicians spreading lies over twitter to create a debate over post-truth politics that has reached its fourth year then this shows that our era is not *entirely* post-truth: Even if the value of truth is under attack we do, still, defend it. Representative democracies depend on information –*accurate* information. A voter who does not have access to accurate information cannot elect the person who best represents their interests.

This dissertation looked at how individuals process false or challenging information. It is situated in a larger body of research on bounded rationality – the idea that decision-making is not perfectly rational, but bounded by the information individuals have, their cognitive ability, and the time they have to make a decision (Simons 1982). Much research

has shown that being uninformed does not necessarily turn people into bad desion-makers. Heuristics can often make up for people's lack of knowledge (Lewandowsky, Ullrich K. H. Ecker, et al. 2012, p.107). It contributes to the strand of research investigating how being *mis*informed affects people's judgment. Volumes have been published about information updating – to what extent do people update their factual beliefs once they have access to new information? Research has shown that outdated information continues to cloud people's reasoning (Lewandowsky, Ullrich K. H. Ecker, et al. 2012; Ullrich K H Ecker, Lewandowsky, and Apai 2011; Seifert 2002) Within this body of research my project contributes to studies of how people process not *any* new information but new information that challenges their factual beliefs and opinions, or that threatens their sense of identity, such as information that disconfirms beliefs commonly held in the groups people belong to. How do people react to information that is, truly, challenging? Unsurprisingly, they don't like it. A large body of research confirms that misperceptions are difficult to correct – the closer they are to people's world views the more difficult they are to accept.

The driver of this bias is well known: an established literature on motivated reasoning theory finds that people tend to interpret information in a manner that is consistent with their prior beliefs (e.g., Kunda 1990; Lodge and Taber 2013). In the context of political misperceptions, Kahan Kahan 2017 developed the paradigm of politically motivated reasoning, arguing that the criterion voters use to evaluate new information is not truthfulness but compatibility with political predispositions. They

explain "identity protective cognition" as "the tendency of culturally diverse individuals to selectively credit and dismiss evidence in patterns that reflect the beliefs that predominate in their group" (Kahan 2016b).

The most comprehensive theoretical framework to describe how individuals react to identity-challenging information was devised by Milton Lodge and Charles Taber (Kraft, Lodge, and Taber 2015; Lodge and Taber 2000; Lodge and Taber 2007; Lodge and Taber 2013; Kahan 2016b). Deploring that "most of the work in the cognitive dissonance tradition did not consider the strength of prior affect to be critical" Lodge and Tabor established an affect-driven model of political reasoning and attitude formation. Their model, named "John Q. Public (JQP)" is based on neurophysiological evidence that we know how we feel about a poltical issue or political candidate before we think about any facts (Lodge and Taber 2007, p.16). Political beliefs, attitudes and predilections in long-term memory are seen as a network of linked associations between objects, such as persons, events, images, or ideas. (For instance, they write, a hypothical weak Republican would most strongly associate Barack Obama with Democrats and African Americans, appraise him negatively for the oil spill recovery, be angry about the Wall Street bailout, and feel ambivalent about the Iraq war.)

At the centre of Taber and Lodge's theory stands the "hot cognition" hypothesis, coined by Abelson 1963: affect and cognition are inexorably linked; there is an affective tag (positive or negative) on every political object that has been evaluated in the past. These feelings spring to mind unintententionally and uncontrollably upon exposure to the name or image of the object. The second you see a president on television you know

how you feel about them. Feelings are followed by the strongest cognitive associations (Lodge and Taber 2013, p. 60). The "affect transfer" hypothesis holds that spontanoes feelings can be transferred into snap judgments which anchor the evaluation of an object before individuals have engaged in any conscious reasoning. With suffcient time and motivation, the retrieved considerations can trigger the construction of conscious reasons for evaluating an object in one way or another (Kraft, Lodge, and Taber 2015, p.128). The authors maintain that it is possible but rare that citizens "construct evaluations", i.e. consciously consider, deliber, and evaluate. They are much more likely to rationalize – generally, deliberation serves to rationalize rather than cause our thinking. .

Building on Lodge and Taber's model this dissertation seeks to find ways to nudge this rare event in which voters consciously consider, deliber, and evaluate. To get there, I seek to understand in which circumstances people are most likely to rely on affective reasoning or heuristic processing. The idea is simple: If we understand what makes motivated reasoning biases worse we can use that knowledge to find ways to nudge deliberative reasoning. Those findings could then be used to inform journalists, public broadcasters, and educators, or any person who wishes to communicate information that is sensitive to certain groups of people. A large body of research in social psychology affirms the effect of affect on information processing (Baumeister et al. 2007, e.g.) (For instance, research has shown unnoticed negative primes (such as sad music or a foul smell) will promote negative and inhibit positive thoughts while unnoticed positive primes (such as happy music or an attractive spokesman)

will promote positive and inhibit negative thoughts) It is therefore plausible that circumstances that trigger certain emotions (such as, for example feeling disadvantaged, or aggrieved) will affect the way voters react to challenging information. (Other approaches include manipulating specific moods or emotions to study their effect on individuals' ability to detect false facts that confirm their opinion or drop false facts that support them.) In chapters 1 and 2 I tested mechanisms that might affect reactions to false facts that affirm prior attitudes (i.e. in terms of Lodge and Taber's model, positively charged (but incorrect) information): a stressful environment (chapter 1) and a low group standing (chapter 2). Chapter 3 explores a mechanism that might affect reactions to fact-checks that disconfirm prior opinions (i.e. negatively charged (but correct) information): post-truth surroundings.

Chapter 1 investigated the effect of stress on belief in four of the false claims spread by the Leave campaign in the United Kingdom's June 2016 EU referendum. It tested if feeling rushed made Leave voters any more likely to rely on motivated reasoning biases, accepting false facts that confirm their own vote choice. The hypotheses was tested among a small convenience sample of n=99 leave voters (as well as 225 remain voters and 5 non-voters in the UK) recruited via twitter. Contrary to hypotheses I found no evidence that the stress treatment increased belief in false facts among the general sample of Leave voters. I do not however interpret these null findings as evidence that stress does not affect the way voters evaluate false facts that confirm their opinions.

Instead, I interpret these null findings as a result of an ill-devised stress

treatment. The treatment consisted of a timer on the Eurobarometer's 3-item EU knowledge quiz. All respondents answered the three questions. Those in the treatment group were timed; those in the control group were not timed. The timer was a visible timer that started at 35 seconds and started ticking the second respondents hit the page. The EU knowledge questions were chosen as challenging questions. According to a 2015 Eurobarometer, only 27 per cent of British respondents answered all three questions correctly. 84 per cent answered at least one correctly (**Hix2015**). However, my twitter-recruited convenience sample was much more educated than the general public. Unsurprisingly, these respondents (who had been sent an invite to the survey because they had used a Brexit-related hashtag and who had agreed to participate in a survey about the EU referendum for free) showed much higher levels of knowledge of the EU than the general public. Putting a timer on these questions most probably didn't turn them into a stressful quiz. To make matters worse, the quiz was introduced as the fun part of the survey: Those in the treatment group saw the introduction that, "We've introduced a time limit to make this a bit more exciting!". If the timer itself made them feel stressed then this introduction probably overrode any feelings of stress with feelings of excitement. I conclude that for most respondents this treatment probably failed to manipulate stress. Therefore, there remains a need for research on the effect of stress on tolerance of false claims that affirm one's political views.

Nonetheless, this data showed heterogeneous treatment effects. The treatment increased belief in false facts among respondents who placed themselves in the middle or on the lower end of the subjective status ladder,

respondents who thought that their personal finances would get worse after Brexit, and respondents who were low in self-esteem. It is possible that these groups of people *did* feel stressed when subjected to a timed EU knowledge quiz (but, because this study lacks a treatment check, it is impossible to tell). However, this finding pointed to an avenue for future research, examining the effect of low status or feeling left behind on information processing. That is what I did next.

Chapter 2 explored the effect of status differences on information processing. The rationale is simple: if you feel that your group is disadvantaged relative to relevant other groups and if a politician recognizes this disadvantage and promises to raise your group's status (ideally above and beyond the status of other groups) then you will want to believe them. If that leader then says or does something that ought to set off alarm bells – for instance, if they use facts that are, simply, false – then you will put on blinders and ignore it. The same should hold for individuals who are (or believe to be) part of a high-status groups whose higher status is threatened: If you feel that your group is about to lose its superior (e.g. cultural or socio-economical) standing or if feel that your group has already lost their legitimately higher status and if you are then exposed to a politician who promises to prevent or undo this status loss then you will want to believe them. If these leaders say anything that could compromise their credibility then you have every reason to ignore those cues.

To test this hypothesis, I designed a laboratory experiment that a) manipulated status differences and b) tested reactions to false facts coming from people who favoured one's own team or the opposing team. The design proved (very) successful at creating low status: The disadvantaged team

performed much worse than the advantaged team and received a much lower payoff, on average. It was also successful a creating high status: The advantaged team performed much better and made much more. It did not however succeed at threatening the advantaged group's higher status. (Hence, the effect of status threat on information processing remains to be studied.)

The low-status team behaved as expected: They overlooked false facts in the feedback that favoured their own team but noticing (and punished) the same false facts when they appeared in the feedback that favoured the other team. These results do not bode well for the future of factual accuracy in political debates in an increasingly populist world.

Chapter 3 investigated the effect of post-truth surroundings on motivated reasoning biases. Here, the focus was on reactions to information that, according to Taber and Lodge's model will have a negative tag on it. Do post-truth surroundings increase reliance on affective cues? In a large-scale survey experiment respondents were presented with expert opinions refuting a false belief they held. Next, some (but not all) respondents were exposed to a post-truth comment nudging them to disregard the expert advice and return to their original beliefs. As expected, the post-truth cancelled out some of the effect of the expert statement.

This dissertation adds to the growing body of research testifying to the difficulty of accepting identity-threatening information and of dropping

identity-affirming (but false) information. It shows that situational factors, in particular low status, affect the way people respond to challenging information. The main finding of chapter 3 points to the most important avenue for future research: If people listen to comments telling them not to take experts too seriously then they might also listen to comments telling them to do the opposite. Paradoxiacally, the most important factor in persuasion research seems to have received the least attention in the misperceptions literature: People. Future research ought to investigate how people – through mututal respect, contact, and conversation – can convey information that is threatening to other people's identity.

Appendix A

Chapter 2

A.1 Screenshots of the laboratory experiment

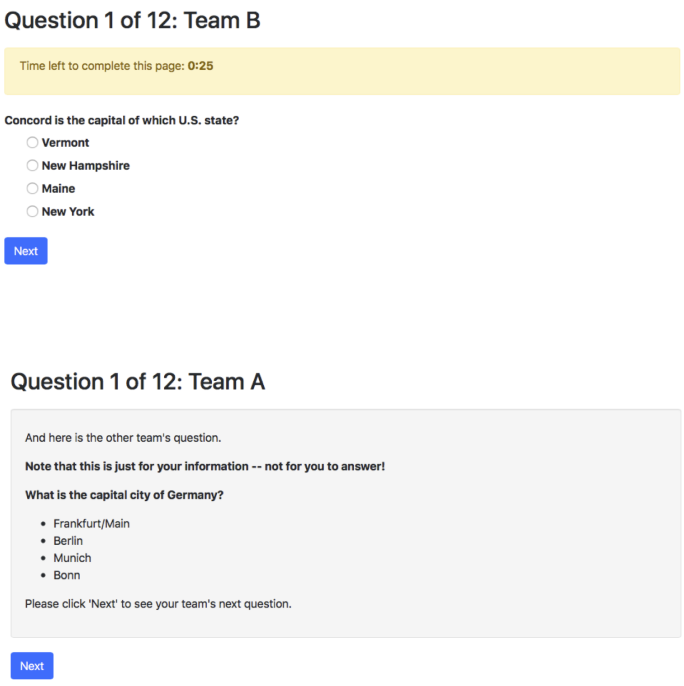
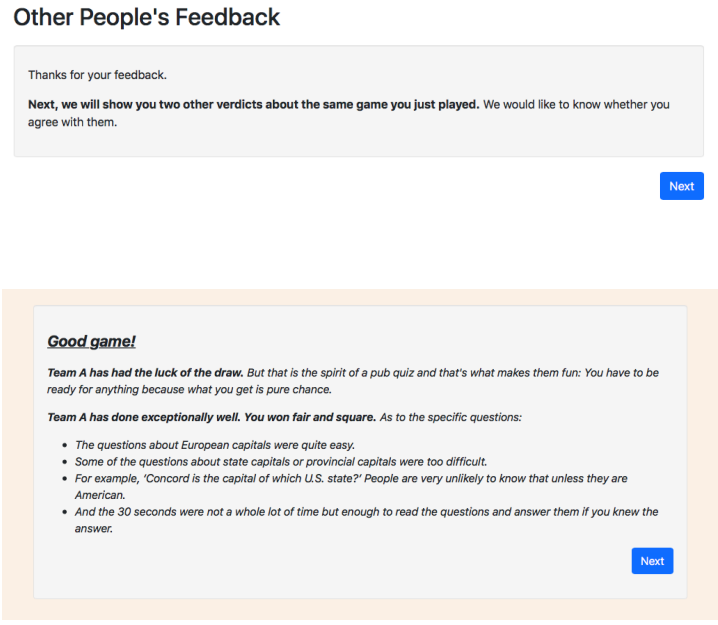
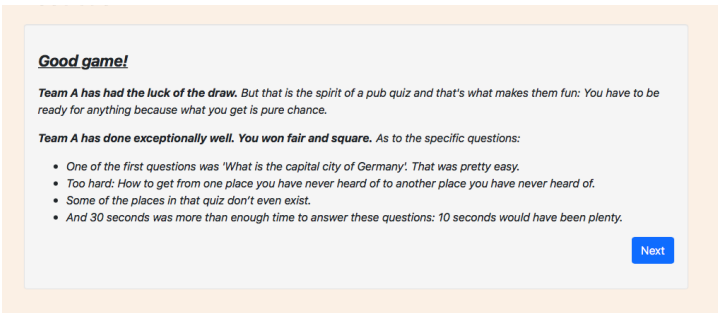


FIGURE A.1: Example quiz questions (as shown to a Team B player)

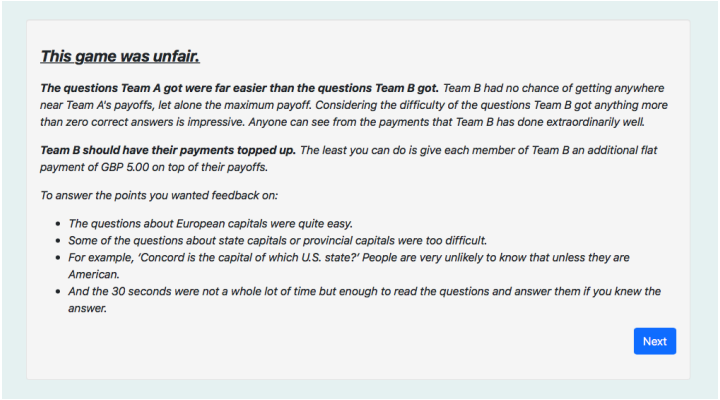


(A) Not including any false claims

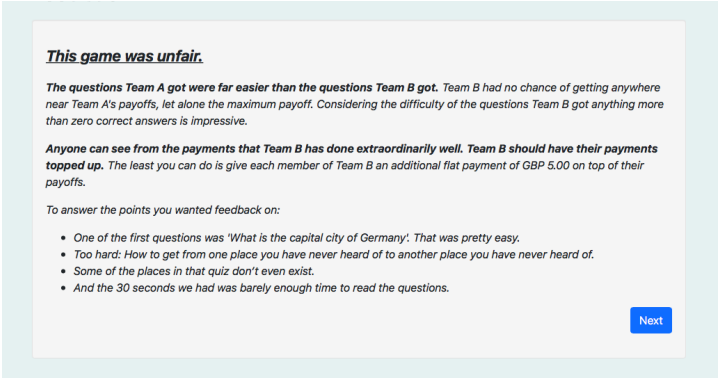


(B) Including false claims

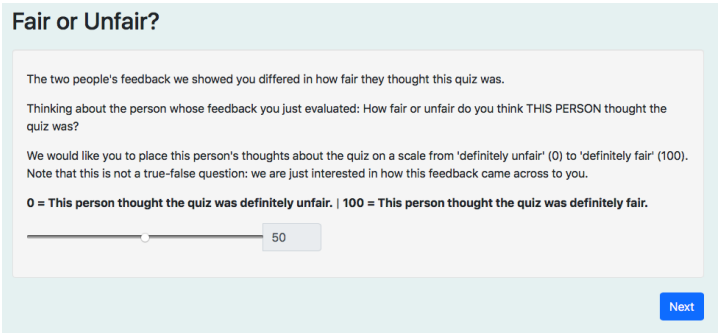
FIGURE A.2: 'Fair Play' Feedback



(A) Not including any false claims



(B) Including false claims



(C) Attention Check

FIGURE A.3: 'Unfair' Feedback

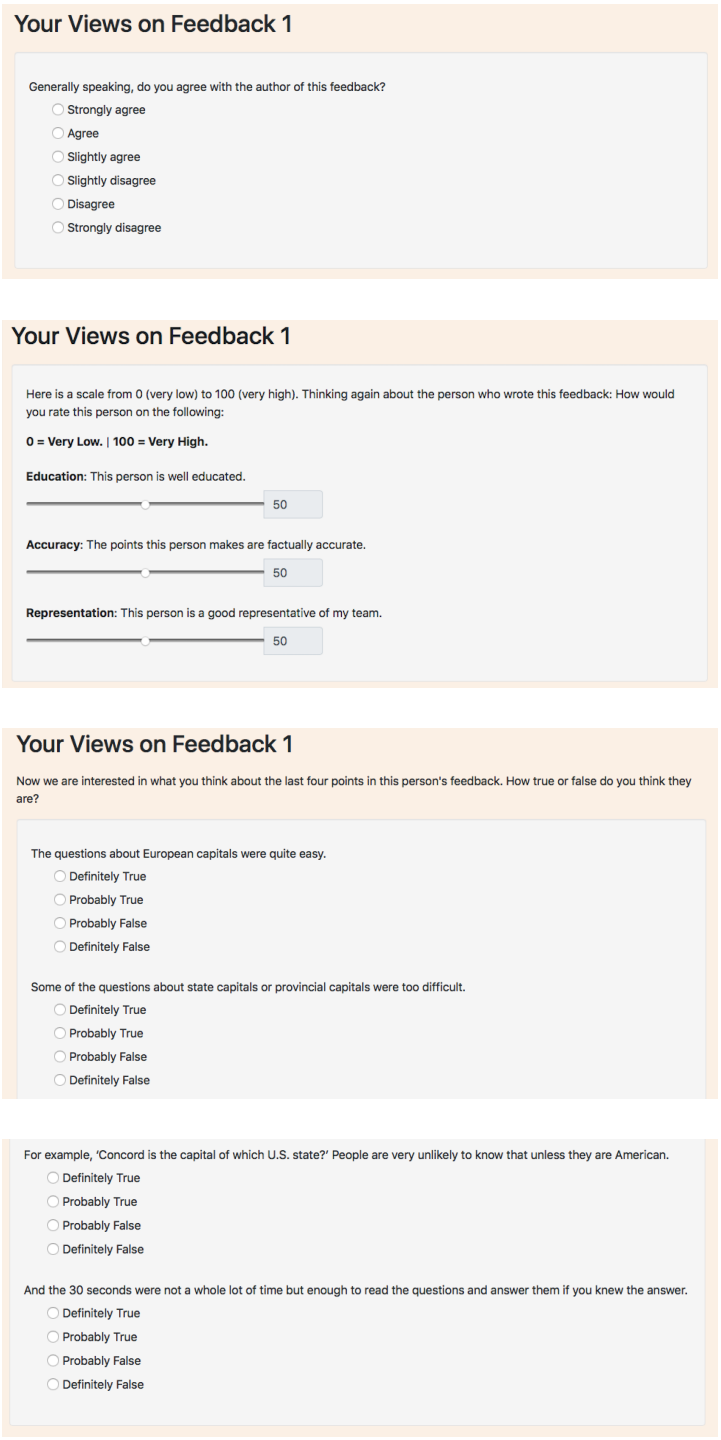


FIGURE A.4: DVs – 'Fair Play' Feedback

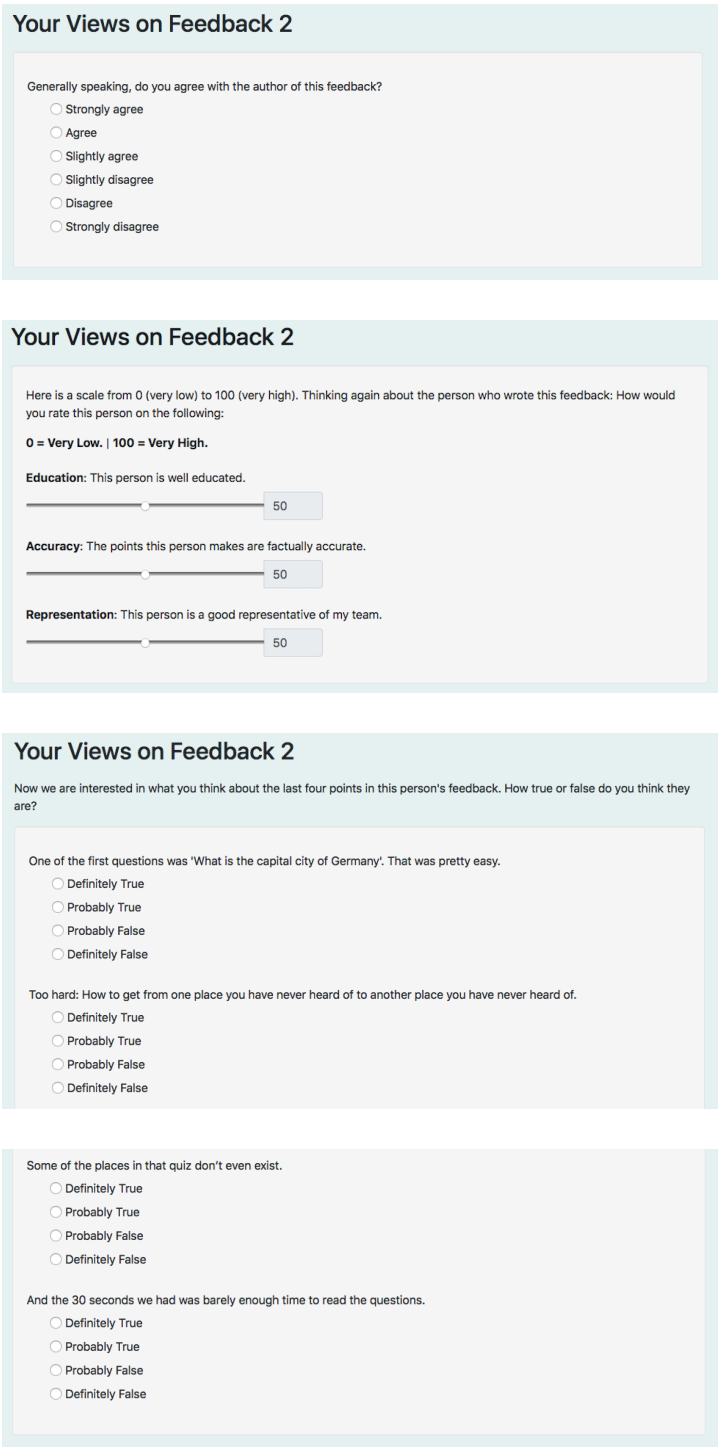


FIGURE A.5: DVs – 'Unfair' Feedback

Appendix B

Chapter 3

B.1 Effect of Post-Truth Comment on Belief in False Claims

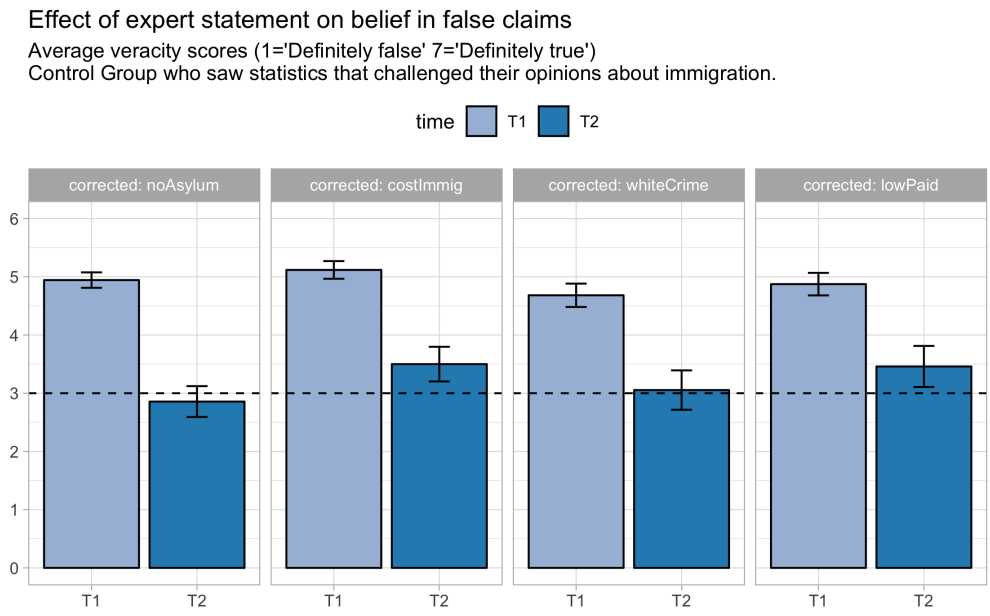
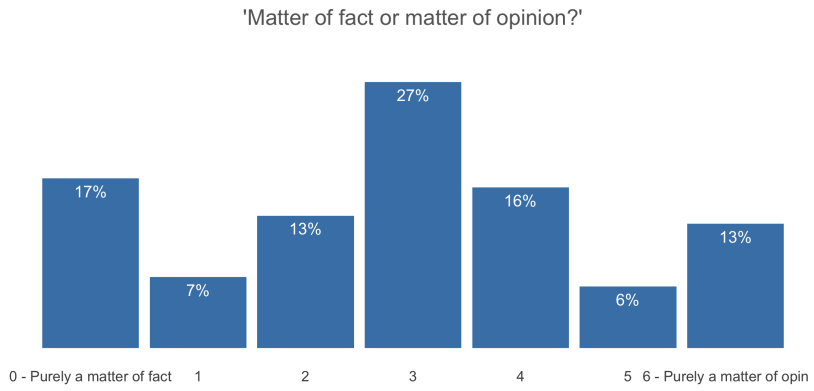
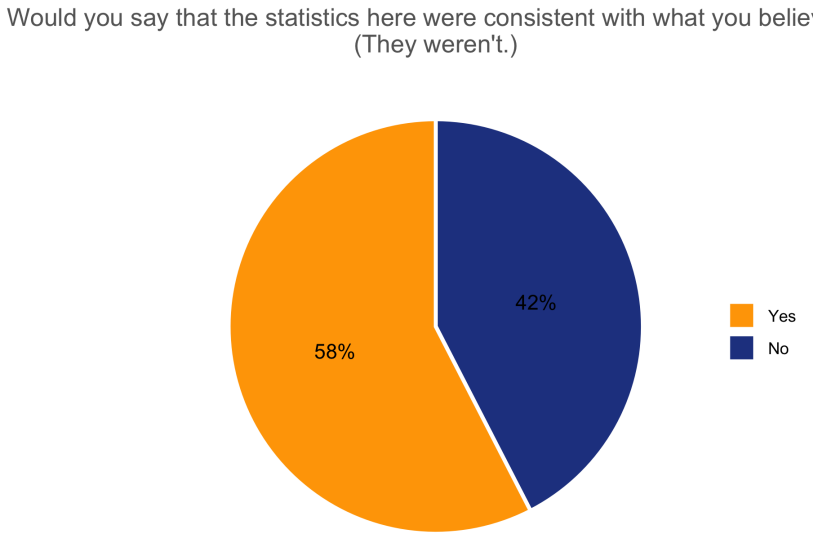


FIGURE B.1: Effect of the correction on belief in false claims

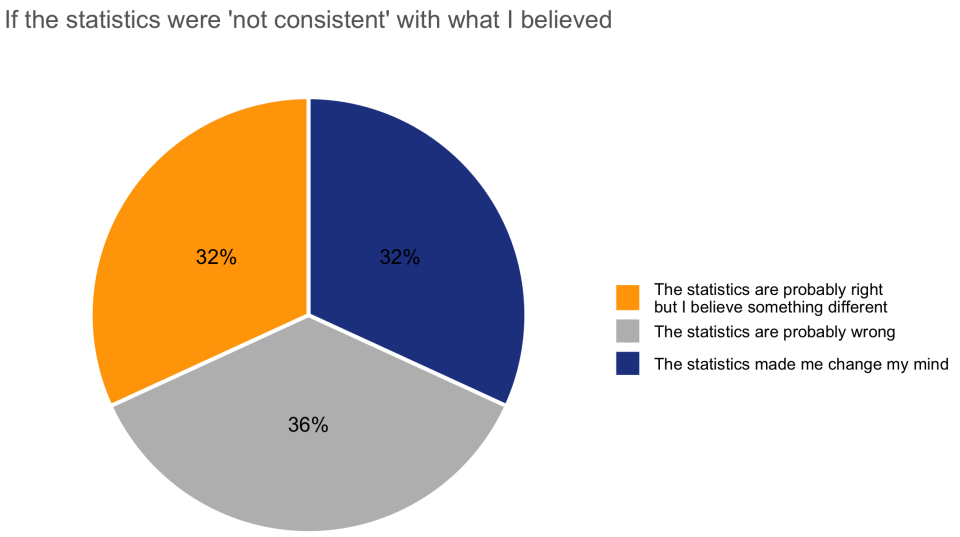
B.2 Evidence of Post-Truth Reasoning in the Control Group



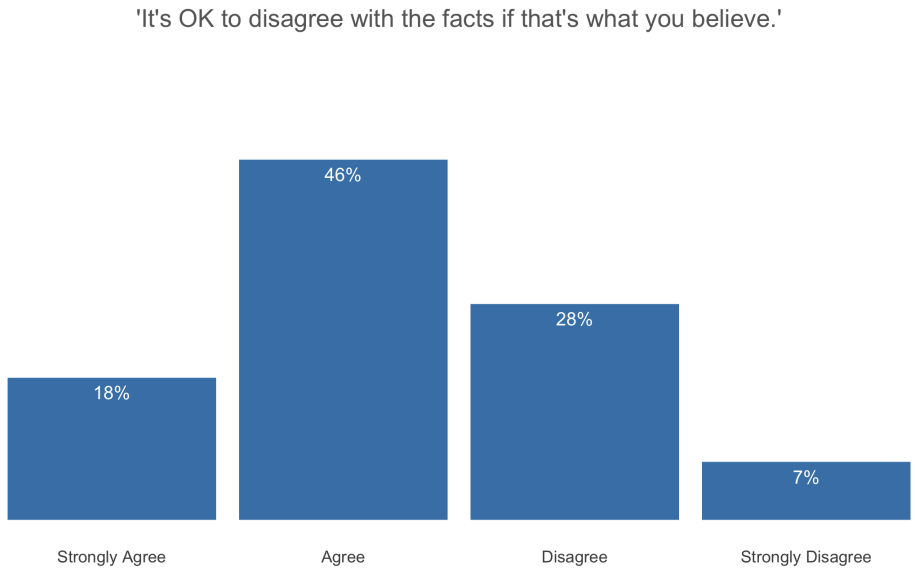
(A) Control group



(B) Control group

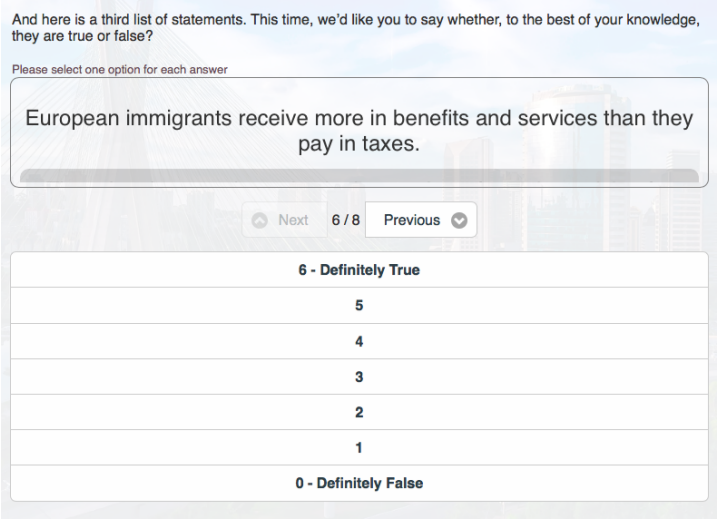


(A) Control group

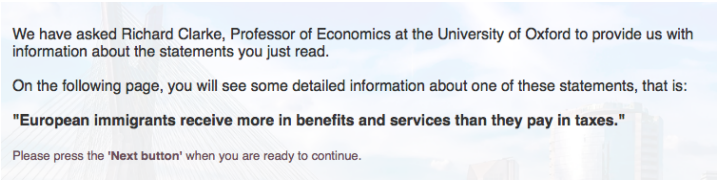


(B) Control group

B.3 Screenshots Survey Experiment

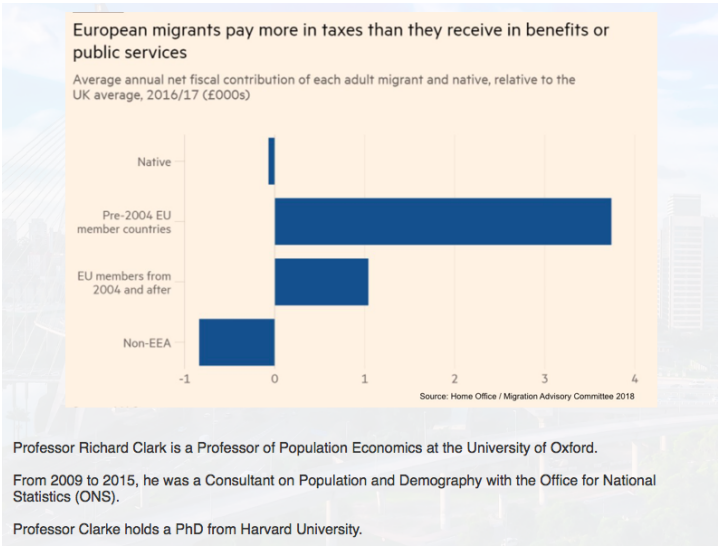


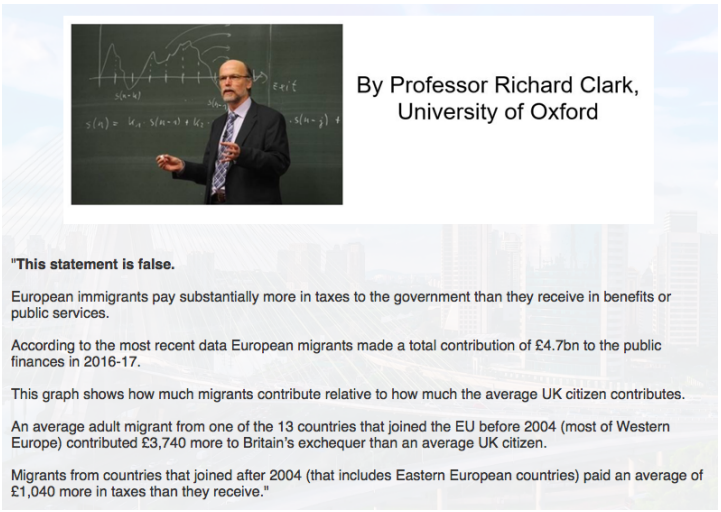
(A) Assessing belief in false facts



(B) Introduction to the expert statement

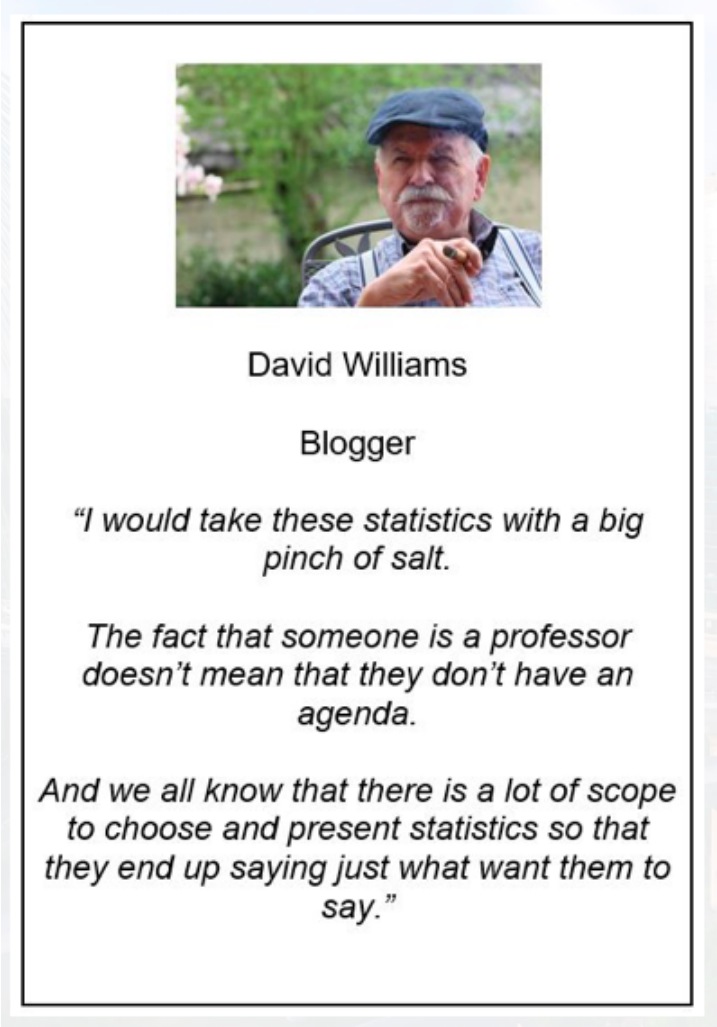
FIGURE B.4





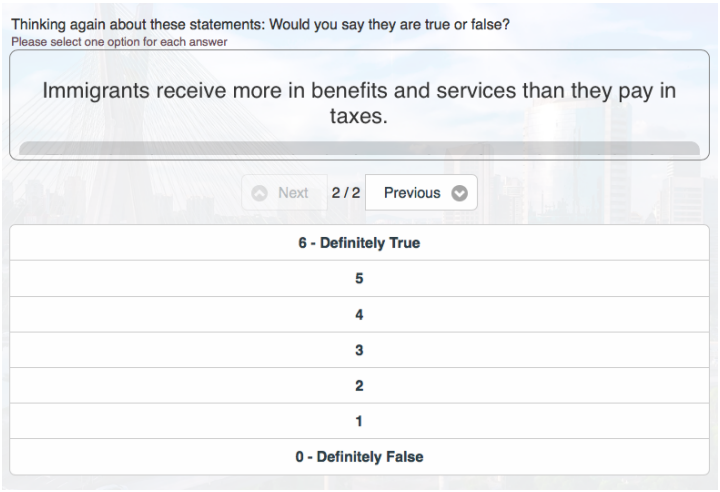
(A)

FIGURE B.5: Expert statement



(A) Post-Truth comment (here: Blogger)

FIGURE B.6: Post-truth follow-up



(A) Post-correction belief in false facts

FIGURE B.7: Post-correction misperceptions