Dispelling Misconceived Beliefs: Insights from Experiments

Jordi Brandts
Isabel Busom
Cristina Lopez-Mayan
Judith Panadés

This version: July 2020
(May 2019)

Barcelona GSE Working Paper Series

Working Paper nº 1096

DISPELLING MISCONCEIVED BELIEFS: INSIGHTS FROM EXPERIMENTS

Jordi Brandts[a], Isabel Busom[b], Cristina Lopez-Mayan[c], Judith Panadés[d]

[a]Instituto de Análisis Económico (CSIC) and Barcelona Graduate School of Economics; [b]Universitat Autonoma de Barcelona; [c]Euncet Business School; [d]Universitat Autonoma de Barcelona and Barcelona Graduate School of Economics

July 2020

Abstract

Some popular views about the workings of the economy are completely at odds with current empirical evidence and congruent theoretical explanations and therefore can be qualified as misconceptions. One consequence is that such beliefs lead to support for harmful policies. Dual process thinking and cognitive biases may contribute to explaining why misconceptions persist even when scientific information is provided to people. We conduct experiments to investigate, for the first time, whether presenting information in a refutational way affects people's beliefs about an important socio-economic issue on which expert consensus is strong, the harmful effects of rent controls. Our refutation text induces a substantial belief change in the direction of expert knowledge, in the laboratory and in the field, although in the former the effect is estimated imprecisely. Measured against a common benchmark, the non-refutational text, the effects of the refutation text are of a similar magnitude. In addition, the persuasiveness of the refutation message varies with individual cognitive traits, and with whether team discussion among participants is allowed for.

JEL: A12, A2, D9, I2.
Keywords: misconceptions; biases; rent control; economic communication; persuasion

Contact: Jordi Brandts; jordi.brandts@iae.csic.es; Instituto de Analisis Economico (CSIC) and Barcelona GSE, Campus UAB, 08193 Bellaterra, Barcelona, Spain. The authors acknowledge financial support from the Spanish Ministry of Economy and Competitiveness ECO2017-88130 (Brandts); Ministry of Science and Innovation, RTI2018-095799-B-I00 (Busom), ECO2017-82882-R (Lopez-Mayan) and ECO2015-67602 (Panadés). Support from Generalitat de Catalunya is also gratefully acknowledged 2017SGR 1136 (Brandts), 2017SGR 207 (Busom), 2017SGR 1765 (Lopez-Mayan and Panadés). Severo Ochoa Program for Centers of Excellence in R&D SEV2015-0563 (Brandts). This work complies with the ethics requirements of Universitat Autonoma de Barcelona and Universitat de Valencia.

1. Introduction

Misconceptions about natural, health, economic and social issues are pervasive in society. Denying climate change, attributing autism to vaccines, and the idea that humans only use a small part of their brain are some illustrative examples of beliefs that are contradicted by scientific evidence. Bensley and Lilienfeld (2017) define misconceptions as "claims about behavior and mental processes that are unsupported or contradicted by high-quality psychological research, that is, they are assertions inconsistent with well-established scientific research. (p. 378)". This definition can be generalized to any field of science. Misconceived beliefs may lead to behavior or choices that have a negative impact both at the individual and social level (Golman, Hagmann and Loewenstein, 2017). It is thus pertinent to ask how misconceptions can be dispelled, whether providing information from experts is sufficient, and whether the way information is provided matters.

Our research contributes, for the first time in economics, new empirical evidence on the effectiveness of a particular communication tool, the refutation text, to help people revise their intuitive beliefs about a prominent economic topic through *slow,* analytical processing of information. This approach is similar to that used to dispel misconceptions in other scientific fields regarding other socially relevant issues. Some examples are found in public health, where international and governmental organizations use misconception-debunking strategies to reduce the serious problem of vaccine hesitancy (World Health Organization, 2017); in natural sciences, where such strategies are used especially regarding climate change denial (Druckman, 2015; Jamieson, Kahan and Scheufele., 2017; Nussbaum, Cordoba and Rehmat, 2017; Yale Program on Climate

Change Communication); and in psychology and STEM education (Kowalski and Taylor, 2009; Masson et al. 2014; Lucariello, Tine and Ganley, 2014).

Studies in cognitive psychology suggest that people often ignore scientific evidence when it contradicts a particular pre-existing belief (Kahneman, 2011; Lewandowsky et al., 2012). Cognitive biases predispose people to accept and hold on to misinformation that spreads through a variety of channels such as the media and social networks, the entertainment industry, the internet and peer groups. Cognitive biases –such as confirmation bias, blind spot bias, self-serving bias and causal illusions–may be at the root of certain beliefs and their persistence. These biases often blind subjects to the data, resulting in misconceptions that are very entrenched and hard to eradicate (Kahneman, 2011). One consequence is that some popular views about the workings of the economy are completely at odds with current empirical evidence and congruent theoretical explanations, and therefore can be qualified as misconceptions.

Work by Caplan (2002), Jacob, Christandl, and Fetechenhauer (2011) and Sapienza and Zingales (2013), among others, provides many examples of the divergence between economic experts' consensus and lay people beliefs. These differences are of concern especially because citizens vote on economic and social policies. Although possibly well-intended, popular misconceptions about economic policies may have a negative impact on welfare, as, for example, when people underappreciate policies' equilibrium effects (Dal Bo, Dal Bo and Eyster, 2017). Several lines of economic research explore the implications of entrenched beliefs and cognitive biases for economic decisions. Some areas of special concern are financial decisions (Lusardi and Mitchell, 2014), education decisions (Lavecchia, Liu and Oreopoulos, 2016; Levitt et al., 2016), and labor market decisions (Cardoso, Loviglio and Piemontese, 2016). All these lines of research owe a lot

to the seminal work of Kahneman and Tversky (2000), Kahneman (2011) and Thaler (2015). However, to the best of our knowledge, strategies for communication with the public about economic policies have only very recently become the focus of research, for example regarding monetary policy (Haldane and McMahon, 2018; Coibion, Gorodnichenko and Weber, 2019).

In this paper we design and test a particular type of intervention –a refutation text– aimed at increasing people's ability to question their own initial, intuitive beliefs. To this end, we use findings from research in cognitive, political and educational psychology, where the problem of misconceptions has been studied for some time. Our motivation is to add to the understanding of how to communicate social science results to broad audiences. We focus here on one specific economic policy: rent controls. There are two main reasons for choosing this topic. One is that evidence and expert consensus about the negative effects of rent controls is very strong, while the belief that they increase the availability of affordable housing is highly popular.[1] And second, it is very hard to eliminate this belief, even after individuals take a formal course in economics, as shown in Busom, Lopez-Mayan and Panadés (2017).

The key feature of a refutation text is how arguments contradicting the misconception are presented (Tippett, 2010). The text must first explicitly state the belief and assert it is a misconception. It then should emphasize the negative consequences of the belief and refute it explaining the arguments and evidence obtained through scientific research. In this way the text distinctly connects this new information to the incorrect information pre-

[1] See for instance the IGM Economic Experts Panel or Mankiw (2015), where about 93% of experts agree that rent controls reduce the amount and quality of housing. In contrast, a poll conducted in December 2018 in London by YouGov/Mayor of London Survey, finds that 68% of the sample suports rent controls while only 16% oppose them.

existing in a person's memory. In addition, the text acknowledges the motivation for the misconceived belief. We discuss this in detail in Section 2.

We study the effects of our refutation text with two types of experiments: the field environment of a college-level economics course, and a laboratory environment that allows us to access a broader young population that is not receiving economic instruction. In the first case, the assignment of individuals to the control or to the treatment group is determined by the year of enrollment. Enrollment year is not a variable of choice by students, as the vast majority enroll after finishing high school or equivalent studies. In our empirical strategy we control for differences across cohorts. Randomizing within the same cohort could create other problems, as we discuss in Section 4.1, and would conflict with university rules. In the case of the laboratory experiment, participants of the same cohort are randomly assigned to control and treatment groups, since the intervention does not interfere with the normal development of their academic activities.

In the field experiment we compare beliefs about rent controls at the beginning and at the end of the semester of two cohorts of first-year college students enrolled in a principles of economics course at the Universitat Autonoma de Barcelona (Spain). Both cohorts are exposed to a standard lecture on price controls and a practice session. The lecture is the same for both cohorts whereas the practice session is different (see Section 4.1). In the control cohort, the practice session consists of solving standard problems on demand, supply and price controls. In the treatment cohort, participants work with the refutation text as described in detail below. The effect of the refutation text is identified by the change in beliefs over the semester across cohorts.

In the laboratory experiment we study the change in beliefs of non-economics students, using the same refutation text. We conduct this experiment at the LINEEX of

the Universitat de Valencia.[2] In this case, we look at changes in beliefs at three points in time, and we study their relationship with measures of cognitive reflection and of confirmation bias. The effect of the refutation text is identified by the change in beliefs over time across conditions. When evaluating the effectiveness of the refutation text the question that arises is what control benchmark to measure it against. In the field environment there is a rather natural benchmark: a lecture based on a standard textbook and problem solving. In the laboratory experiment, the nature of a proper benchmark is not determined by the environment. One possibility is to have participants answer the questionnaire twice, providing only very succinct information about the effects of rent controls. We decided, instead, to write a control non-refutational text for the laboratory, in which we attempted to convey the same information as students in the control group of the field experiment are exposed to.

We find that the refutation text reduces the strength of the rent-control misconception, independently of whether subjects are in the specific environment of an economics class, or whether they are a broader audience in a more controlled environment. The impact of the refutation text in the field is larger than in the laboratory, where the effect is not significant.

After seeing this result we conduct a follow-up field treatment to better understand the roots of the difference in the results between the field and the laboratory. In the new experiment students are given the same control text as in the laboratory to compare its effect to that of the refutation text. The results show that the non-refutation text induces a change away from the misconception, like in the laboratory, compared to standard instruction. However the refutational content induces an additional change away. The

[2] LINEEX stands for Laboratory for Research in Behavioural Experimental Economics. See https://lineex.es/en/home/.

magnitude of the additional effect of the refutation text in the field is consistent with the magnitude in the laboratory, where we test the refutation text against the non-refutation text.

Note that our work is not strictly about field-laboratory generalizability in the sense of Camerer (2015). It is about the impact of different communication formats with various audiences. Our findings show that the refutation texts causes a larger reduction in the misconception than other formats, but the effect is only statistically significant in the field. Being aware of our audience's misconceptions and designing communication tools accordingly has the potential of significantly reducing a misconception on a relevant economic policy issue, and of increasing the demand for evidence-based policies, but using the refutation format may not be the only way to go.

In other domains of science the concern about effectively communicating scientific evidence and, in particular, dispelling misconceptions, has become an active area of research. In economics, awareness of the importance of cognitive biases when communicating economic knowledge is still embryonic but does exist. For example, in his recent book, Tirole (2017) devotes the first chapter to describing how cognitive biases affect our intuitions about economic issues. He highlights how the need to have a positive self-image, to avoid facing unpleasant facts and to follow our first impressions influence beliefs. However, academics often implicitly expect students as well as the public at large to rationally accept the logic of economic models and statistical facts, and disregard the importance of pre-existing views and of cognitive biases. Unless economic communication takes into account these factors, the impact of economic knowledge on society may be rather limited. Our work contributes to this still undeveloped research area in economics.

The rest of the paper is organized as follows. Section 2 discusses how to deal with misconceptions. Section 3 contains our hypotheses and research questions and section 4 describes our experimental design. In section 5 we present the results. Section 6 concludes.

2. Dealing with misconceptions

Reducing the distance between experts' consensus on economic policy issues and popular misconceptions requires investigating whether differences arise because people are not well-informed about economic experts' views –an information gap problem– or whether differences arise because of resistance to changing prior beliefs. Johnston and Ballard (2016) conduct an experiment to shed light on this question. They design different conditions according to whether participants are informed about economic experts' views on particular items. They find a significant opinion change when the item is of technical nature; however, when symbolic and politically salient issues are involved, opinion change is more unlikely. In addition, respondents that initially disagree with experts' consensus strengthen their prior opinion through the belief that economists are unreliable. By contrast, respondents that initially agree with experts' views reaffirm their prior opinion judging economic experts to be reliable. This behavior is consistent with confirmation bias, whereby information that contradicts a subject's prior beliefs is discarded, while information that is consistent with prior beliefs is retained. Other cognitive factors, such as the propensity to engage in analytical thinking, may also contribute to explaining the prevalence and persistence of misinformation (Pennycook and Rand, 2019).

In Johnston and Ballard (2016) participants in the treatment conditions are given information on experts' views about a particular issue immediately before asking for the participant's opinion. A plausible hypothesis is that exposure to economic instruction might be potentially more effective at dispelling misconceptions. We would in particular expect that taking a college level course in economics would achieve this goal. At this point, however, there is scant evidence that speaks to this issue. Most studies in economics evaluate the impact that a variety of instructional methods (classroom experiments, case studies, flipped classroom) have on students' academic performance as measured by grades (List, 2014; Allgood, Walstad and Siegfried, 2015), but do not inquire about beliefs, with the exception of Busom, Lopez-Mayan and Panadés (2017) that shows that course exposure hardly affects prior beliefs.

In contrast, academics in natural sciences and in psychology have long been concerned about misconceptions (Nakhleh, 1992; CUSE, 1997; Lilienfeld et al., 2009; Kowalski and Taylor, 2009; Lilienfeld, 2010; Lucariello, Tine and Ganley, 2014; Masson et al., 2014; Nussbaum, Cordova and Nehmat, 2017). Studies find that several cognitive biases are at the root of the failure of purely expository communication in reducing misconceptions (Bensley and Lilienfeld, 2017). Standard communication methods do not explicitly challenge misconceptions and, in terms of the dual system theory of information processing by the human mind, do not engage System 2, or analytical thinking (Stanovich and West, 2000; Kahneman, 2011).[3] Research on how the brain learns suggests that when people holding views about how the world works are presented

[3]This model postulates that the mind uses two mechanisms to process information and generate judgments: one is automatic, involuntary and fast (System 1), while the second is slow and requires allocation of attention and effort (System 2). The first generates impressions, intuitions and feelings; it can make mistakes because it is prone to exhibit biases that result from mental shortcuts and heuristic processes. The second adopts these intuitions, which become beliefs unless a surprising event occurs and makes it allocate effort to information processing.

with contradicting information, the brain classifies this information as errors, allowing them to stick to their original views (Dunbar, Fugelsang and Stein, 2007; Masson et al., 2014; Byrnes and Dunbar, 2014). Therefore, theoretical and empirical arguments by themselves are not sufficient to dispel misconceptions (Kahneman, 2011). It is necessary to design an appropriate communication strategy.

Some researchers propose using communication methods that first explicitly call up the misconception and then disprove it. A prominent one consists in designing a *refutation text*, defined as a text that "engages, challenges, and remediates common misconceptions" (Kowalski and Taylor, 2009; Tippett, 2010; Braasch, Goldman and Wiley, 2013; Kendeou, Braasch and Braten, 2016; Lassonde, Kendeou and O'Brien, 2016). Although other communication formats are possible, we decide to use the refutation text based on previous success in debiasing misconceptions in other fields, such as climate change (Nussbaum, Cordoba and Rehmat, 2017), psychology (Lilienfeld et al., 2009; Kowalski and Taylor, 2009; Lilienfeld, 2010) and mathematics (Lucariello, Tine and Ganley, 2014).

In our case, as discussed in the introduction, we single out the popular belief that rent controls increase the number of families that find affordable housing. This belief can be qualified as a misconception because extensive empirical evidence contradicts it across time and countries: see Glaeser and Luttmer (2003); Gyourko, Saiz and Summers (2008); Mora-Sanguinetti (2011); Andrews, Caldera Sanchez and Johansson (2011); Andersson and Söderberg (2012); Hilber and Vermeulen (2016); Gyourko and Molloy (2015) and Molloy (2020), and Diamond and McQuade (2019).[4] Although the public may hold many other economic misconceptions, we focus here on the one about rent controls because it

[4] See also the website of the Stockholm Housing Agency, https://bostad.stockholm.se/english/.

exhibits two interesting features. First, it is an issue where experts' and popular opinion exhibit a remarkable disparity and is of social importance. According to the IGM Economic Experts Panel, 95% of panel members disagree with the idea that rent controls lead to an increase in the quantity of affordable housing.[5] In contrast, according to a poll conducted by the Institute of Governmental Studies (IGS) of UC Berkeley, 60% of the state's registered voters favor rent control, while 26% oppose them. In the UK, a survey published by Survation reports that 59% of about 1000 polled people in December 2014 backed rent controls, and only 7% opposed them.[6] Although the questions asked are not identical, these surveys illustrate the significant gap between experts and the public in the support for rent controls.

Second, Busom, Lopez-Mayan and Panadés (2017) document that this misconception is hard to dispel. Notably, they find that this false belief remains even if standard instruction based on theory and solving problems received for a semester in a principles of economics course contradicts it. Indeed, while at the beginning of the semester 67% out of 554 students share this belief, at the end of the semester 74% students hold it. Thus, not only do students who initially have this misconception stick to it, but some of those who do not hold it at the beginning change their view in the wrong direction. Interestingly,

[5] The statement reads: "*Local ordinances that limit rent increases for some rental housing units, such as in New York and San Francisco, have had a positive impact over the past three decades on the amount and quality of broadly affordable rental housing in cities that have used them*". See http://www.igmchicago.org/surveys/rent-control.

[6] The question in the IGS poll was: "*Some people believe rent control laws that give local governments the ability to set limits on how much rents can be increased are a way to help middle and lower income people remain in their communities. Others say rent control leads to fewer rental units being built and this makes the problem worse in the long run. What is your opinion? Do you favor or oppose rent control laws in your area?*" The question in the Survation survey was: "*Would you support or oppose proposals for the government in introduce a "rent control" system in the UK*?". See http://survation.com/public-back-introduction-of-rent-control-survation-for-generation-rent/.

many students who stick to the misconception do well in mid-term or final tests that include a question on this topic.

The refutation tool is designed to dispel a specific misconception and so it requires knowing the existence of that misconception. In our case, our knowledge comes from previous work (Busom, Lopez-Mayan and Panades, 2017) and from surveys such as those conducted by Survation and IGS cited above. Other economic misconceptions might be identified through appropriately designed surveys.

We write the refutation text following guidelines from the psychology literature (Tippett, 2010). According to these, the text should state, or activate, the misconception, affirm explicitly that it is incorrect, provide arguments and evidence as to why this is the case, stressing the negative consequences of the belief and explaining the scientific conception as simply and clearly as possible. Druckman (2015), drawing on his and others' work, lists several factors as essential for offsetting motivated reasoning through communication: capturing the attention of the audience, making clear that the issue is relevant to the individual and her values, and facilitating interactions by asking individuals to explain their opinions to others. Discovering that their belief about a particular matter is a misconception may come as a surprise to people, who then may slow down their thinking process, and become more open to the scientific arguments. At the same time, this should be done in such a way as not to repeat the misconception too often as this could produce a backfire effect.

Our text states the misconception and explains the negative, unintended consequences of a rent control policy as predicted by economic theory. It then provides real data on current rental prices in the areas participants live in (Barcelona in the field and Valencia in the laboratory) with the intention of providing examples close to students'

environment. We write the text placing the rent control discussion in the context of searching for policies that are effective in improving social welfare and fairness. It is reasonable to assume that fairness concerns motivate the belief that rent control is a good and easy solution to the housing problem. We explain the negative effects that a rent control policy –queuing, black market– has had in Stockholm (Andersson and Söderberg, 2012). We choose Stockholm to illustrate that the negative consequences of rent controls arise even in a country especially known for its welfare-oriented policies. The text then explains alternative policies that would increase affordable housing without the negative effects of rent controls. In this way students learn that scientific knowledge enables them to find effective policies that are aligned with their values. In our case we not only refute the misconception but also offer credible policy alternatives that respond to people's legitimate fairness concerns. In the text we use simple, non-technical language, and provide some cues to induce critically thinking about own beliefs. For instance, we include sentences like: "*how things are often more complex than they seem"; "thinking slowly [...]", "ask ourselves the following questions"*. The refutation text is about 1200 words long. The complete text can be found in Appendix B.1.[7]

We wish to highlight here that previous studies on the effects of refutation texts find that they are effective in many but not all cases (Tipett, 2010; Nyhan and Reifler, 2010). It is hence a promising but not guaranteed route to dispelling misconceptions. In particular, it is, a priori, not clear whether a refutation text can be effective in a social science environment where misconceptions may be arguably more entrenched due to the direct connections between certain misconceptions and political views. Our paper is a step towards finding out how to communicate economic evidence to the society.

[7] Nussbaum, Cordoba and Rehmat (2017) use a 1009 word long refutation text on the greenhouse effect.

3. Hypotheses and research questions

We now introduce our formal hypotheses. On the basis of previous evidence about the successful use of refutation texts in other areas, we formulate the following main hypothesis:

Hypothesis 1: The refutation text will have a positive impact on the misconception relative to standard teaching in the field and to a non-refutational text in the laboratory.

We want to find out whether a refutation text in the spirit of those used in psychology and natural sciences can be convincing both for college students taking an economics course and for a broader public. The details of the respective experiment designs are explained in section 4.

Hypothesis 1 is about the direction of the change in beliefs both in the field and in the laboratory, not about the magnitude. Camerer (2015) reports on good comparability between the laboratory and comparable field settings with respect to the direction of the effects, but magnitudes need not be similar.[8] In our case, specific features of the respective environments may influence the magnitude of the impact, but we do not have priors about how these factors affect it. First, in the field experiment participants are students taking a principles of economics course, and are therefore exposed to economic reasoning for a whole semester. In the laboratory experiment our participants are college students who have not taken any economics courses. We intentionally exclude students with exposure to economics from participating to make our participant pool more similar to the general population in terms of educational and professional interests. Second, in

[8] In some cases they might be. Herbst and Mas (2015) compare estimates of peer effects on worker output obtained in a number of laboratory experiments and in field studies. They find that the average magnitudes in both environments are similar in terms of the mean and variance of productivity spillovers.

the field the refutation text is presented to students by a professor, an authority figure, and students might be inclined to respond as they think the instructor expects them to, even if they know that their responses do not affect grades. In contrast, the laboratory experiment is not administered by professors but by the professional laboratory staff members who interact with the participants only for a short time.

The laboratory experiment allows us to test several additional explicit hypotheses related to the conditions under which the refutation text is more effective. First, we conduct separate laboratory sessions with and without group discussion and this allows us to investigate whether group discussion makes a difference relative to individual reading of the text. In the field experiment it was not possible for us to have both groups with and without group discussion, due to university regulations that require all students in a cohort to be taught in the same way. There is some evidence that group discussion may reinforce the effect of the refutation text. For instance, Cooper and Kagel (2005), Charness and Sutter (2012), and Druckman (2015) find that in many cases groups make better decisions. However, depending on the group composition, team discussion may strengthen the misconception if the majority of the group has it and persuades the minority to hold it (Levendusky, Druckman and McLain, 2016). These two forces may offset each other and we thus posit the following hypothesis:

Hypothesis 2: There will be no difference in the impact of the refutation text with a group discussion and without it.

An important issue is the duration of the potential effect of the refutation text. To investigate this in the laboratory we ask for participants' beliefs both right after the refutation text is administered (immediate effect) and some weeks later (delayed effect). In the field we are only able to measure the delayed effect because, unlike in the

laboratory, in the field we are subject to a time constraint of a 90-minute class session. Within this time span we have to on the spot randomly allocate students to teams and have them work with the refutation text. There is then no time left for filling out a survey in the same session.

Previous work has studied the effects of refutation texts at different points in time (Broughton, Sinatra and Reynolds, 2010; Kowalski and Taylor, 2017; Lassonde, Kendeou and O'Brien, 2016; Nussbaum, Cordova and Rehmat, 2017). This literature suggests that the effect decays over time. In our laboratory study we are able to analyze this issue. Our hypothesis in this case is the following:

Hypothesis 3: The delayed impact will be smaller than the immediate impact.

In addition to the explicit hypotheses 1-3 we dig deeper into the possible effect of the refutation text by exploring the heterogeneity of treatment effects, both in the field and in the laboratory. We ask whether the strength of the change of beliefs induced by the refutation text varies according to the initial agreement with the misconception.

To better understand the mechanisms driving the change in beliefs we investigate whether some cognitive factors may be correlated with these changes. To this end, we add two building blocks in the laboratory. First, we include the Wason selection task (WT), which is a measure of confirmation bias (Wason, 1960 and 1977). Confirmation bias is the tendency to search for, interpret, favor, and recall information that confirms or supports one's prior beliefs or values. The WT has been used for instance in experiments about how people process or seek information individually (Charness, Oprea and Yuksel, 2018) or make decisions in teams (Charness and Sutter, 2012). Second, we include a nine-item Cognitive Reflection Test (CRT), intended to measure the dominance of reflective versus intuitive thinking processes in individuals. The CRT has been used in economics

to study decision-making, strategic behavior and social preferences (Brañas-Garza, Kujal and Lenkei, 2015). The initial test proposed by Frederick (2005) contained three items, but it has been expanded after some revisions (Toplak, West and Stanovich, 2014; Thomson and Oppenheimer, 2016). Based on cited work as well as on Wood, Galloway and Hardy (2016), we posit the following hypotheses:

Hypothesis 4: Participants who score higher on the WT will change their beliefs more strongly in the right direction.

Hypothesis 5: Participants who score higher on the CRT will change their beliefs more strongly in the right direction.

For ease of exposition, results regarding hypotheses 1, 2 and 3 are presented in section 5.2. Results about the heterogeneity of the treatment effect are shown in section 5.3 and results regarding hypotheses 4 and 5 are reported in section 5.4.

4. Experimental design

We design a survey to know participants' beliefs and attitudes about several economic issues. The key statement that refers to the misconception reads as follows: *"Establishing rent controls, such that rents do not exceed a certain amount of money, would increase the number of people who have access to housing facilities."* Participants are asked to indicate their agreement with this and remaining statements on a five-point scale. Items included in the surveys are detailed in Appendix A.

Both in the field and in the laboratory participants fill out the surveys. We also collect socio-demographic information, and in the case of the field experiment, we use

administrative records. A detailed separate description of the design and procedures of each experiment follows in the next section.[9]

4.1 The field experiment

Participants are first-year college students enrolled in a principles of economics course at the Universitat Autonoma de Barcelona (UAB), Spain. This is a compulsory course for first year students in five majors offered by the School of Economics and Business and by the Law School: Business, Business and Law, Economics, and Law and Labor Relations, respectively.[10] At the beginning of the semester students fill out a survey, which includes the statement about rent control. Towards the end of the semester, we have again students respond to a survey that includes the same statement on rent controls. Students are informed that the surveys will not be used for grading, so that they do not feel pressured to respond in a way that does not reflect their true own opinion.

The control group consists of students who took principles of economics in 2015 in the majors indicated above. The treatment group consists of students enrolled in the same course and majors in 2017. Assignment to treatment and control groups is not decided by students; it depends on the year of enrollment, 2015 or 2017.[11] We acknowledge that the time gap of two years between cohorts may be a concern since it does not allow holding everything constant between 2015 and 2017. However within cohort randomization in the field is problematic as well. We would be unable to control for potential contagion–

[9] A common concern in experimental work is the presence of an experimentar demand effect. That is, participants may try to guess the purpose of the experiment and change their behavior accordingly, posing a threat to external validity (Zizzo, 2010). De Quidt et al. (2018) find that the magnitude of these effects is modest.

[10] In Spain, first-year college students enroll directly in specific majors; they can switch fields subsequently. In Spain, legal studies can be a student's major at the undergraduate level.

[11] Potential concern about students' deciding to delay college entry after completing other studies, for instance because of taking a gap year, can be dismissed as in Spain this is a rather rare choice.

spillover–effects among treated and control students. It would not be possible to randomize within major since some majors have only one group. In addition randomizing students within the cohort would violate university rules, which require that the same teaching content and practices be used in all groups. The two-year gap instead of one-year gap avoids the presence of control students in the treatment group because of failing to pass the course in 2015.

Course content and organization is the same for both the control and treated groups. Instructors use the Krugman, Wells and Graddy (KWG) textbook (2015 Spanish edition); class sessions include lectures and practice sessions. Special care was taken to teach course content in the same way to both cohorts, except for the use of the refutation text in the treatment group. Course evaluation is based on in-class graded assignments, a mid-term and a final exam.

About the fourth week in the semester the topic of price controls, where the misconception we focus on fits in, is introduced within the context of supply and demand analysis. The lecture on price controls in all cohorts is standard and based on the corresponding chapter in KWG. It explains with graphical analysis types of price controls (floors and ceilings), effects on the market equilibrium and deadweight loss; examples of price controls are provided.

Regarding the practice session about price controls, the control cohort participates in a standard session, where students are given linear demand and supply functions, have to solve for the equilibrium both numerically and graphically, provide the new market solution when a binding price ceiling is imposed and calculate deadweight loss.

The practice session for the 2017 cohort consists of reading and discussing the refutation text in class (see Appendices B.1 and B.4). Students are randomly allocated to

a team of three or four at the beginning of the session. Each student is given a copy of the text and they are asked to read the text individually and answer, on paper, some questions using the template provided. Each team has to hand just one response template. This session is graded with a weight of around 10% of total grade. The whole exercise can be completed during a 90-minute class[12].

The initial sample sizes are 399 participants in the treatment group, and 508 in the control group. Since our outcome of interest is the change in beliefs at the end of the semester, we exclude from each initial sample participants who did not answer both surveys. Failure to respond to the end of semester survey could be attributed mostly to non-attendance, as in-class presentations or essay deadlines for the different courses tend to be concentrated towards the last two weeks and some students may be under strong pressure. A small number of participants respond the end of semester survey but not the initial one, likely because of late enrollment –students have a period of about four weeks for group switching-out– or non-attendance for personal reasons. As shown in Appendix C, Table C.1, this leaves us with a sample size of 272 (68% of initial sample) in the treatment group, and of 340 (67% of initial sample) in the control group. Attrition rates are practically identical, but the composition of the two samples may still differ. Column (5) in Table C.2 shows that the assigned-to-treatment group and the control group are very similar in terms of gender, share of non-Spanish students and retakers, and, importantly, in the College Admission Test (CAT) grade, which is a proxy for general academic ability at college entry. We observe significant differences in the share of younger students and scholarship recipients. Composition by major also varies, as the

[12] Prior to the refutation text session, students were suggested to search on their own for information about the number and prices of rental housing available in their town of residence in the Barcelona area, on a widely used website in Spain. Doing this was not a requirement to participate in the refutation text session and it was not graded.

number of groups in the Business major assigned to one of the instructors participating in the experiment was smaller in 2017 than it was in 2015. We control for these differences in our empirical analysis.

Of the 272 participants from the treatment cohort (intended-to-treat) who answered both surveys, 243 received treatment, i.e., attended the refutation text session. The compliance rate thus was 89%. Non-compliant participants are not expected to be a random sample. Table C.2 shows substantial differences between compliers and non-compliers mostly in terms of gender, retaker and major. Compliers are positively selected as proxied by the lower percentage of retakers, although the average CAT grade is not significantly different (see column (6)). We take into account this potential source of selection as we explain next.

We estimate the impact of the treatment, the refutation text, on the change in beliefs about rent controls using the following baseline empirical model:

$$y_i = \alpha + \beta D_i + \varepsilon_i \qquad (1)$$

where (y_i) is the change in beliefs, computed as the difference between a student's response to the statement at the end of the semester and her response at the beginning. The original responses in the five-point scale are transformed into numerical values according to the degree of agreement with the statement: 5 (fully disagree), 4 (disagree), 3 (do not know), 2 (agree), and 1 (fully agree). Hence y_i takes values between −4 (a change from fully disagree at the beginning to fully agree at the end) and 4 (a change from fully agree at the beginning to fully disagree at the end). That is, a positive value obtains when the response varies from agreement towards disagreement. If the student provides the same response in both periods, the change is zero. D_i is a dummy variable equal to one if a student is in the treatment cohort and receives treatment, and zero otherwise.

Differencing beliefs within cohort takes out fixed unobserved heterogeneity over the semester that may affect the results. In order to minimize confounding factors that may differ across cohorts, equation (1) includes the following set of variables that capture differences in cohort composition: gender, age 18 (dummy variable taking value one if the first-year student is 18, zero if she is older), retaker, CAT grade, receiving a scholarship, non-Spanish student, high school track, majoring in Law or Labor relations, and in Economics.

When estimating the β parameter in equation (1) we take into account that the group of participants intended to treat and actually treated is not the same. Twenty-nine (11% of participants in the assigned-to-treatment cohort) do not comply with the treatment. As discussed above the composition of the samples of compliers and non-compliers differ in some dimensions. Therefore, the comparison between those actually treated and the control group may be misleading. We deal with this selection using an instrumental variable approach. If assignment to treatment and control groups is random, it is appropriate to use intention-to-treat as an instrument (Angrist and Pischke, 2009). In our case assignment to the treatment group is a valid instrument since as explained above students do not choose the assignment group. We thus define a variable Z_i, which equals one if a student i belongs to the treatment cohort and zero if she belongs to the control cohort. The treatment variable D_i is instrumented with Z_i. With a binary instrument the IV estimator of β is the Wald estimator.

The effect of the treatment is likely to be heterogeneous across individuals. For example, some individuals may be more responsive than others to the refutation text because of differences in unobserved cognitive abilities. In the presence of heterogeneous treatment effects, the Wald estimator identifies the LATE (local average treatment effect)

if the monotonicity assumption holds (Imbens and Angrist, 1994; Angrist, Imbens and Rubin, 1996). The LATE is the effect of the treatment on compliers. Monotonicity implies that there are no *defiers*, that is, people who receive treatment even if they have been assigned to the control group. In our case, the monotonicity assumption is satisfied since as discussed above no one from the 2015 cohort received treatment –i.e. no student enrolled in 2015 was also enrolled in 2017 (see Appendix C, Table C.3). When using intention-to-treat as instrument and under the monotonicity assumption, the LATE is equal to the average treatment-on-the-treated (ATT) effect as proved in Bloom (1984) and discussed in Angrist and Pischke (2009).

4.2. The laboratory experiment

The laboratory experiment was conducted at the LINEEX of Universitat de Valencia (Spain) in October 16 and 17 and November 8 and 9, 2018. Participants were recruited from a student pool of 40,000 students from Universitat de Valencia and Universitat Politècnica de Valencia in the fall 2018. Subjects are chosen from majors that do not include a principles of economics course.

We have three conditions: a control and two treatments, with different participants in each; in experimental terminology, we use a between-subjects design. In the control, participants read a text that is meant to reproduce as closely as possible the content of the lecture that students are exposed to in the control group of the field experiment (see the complete text in Appendix B.2 and instructions in Appendix B.4). In the first treatment, participants read the refutation text individually. In the second treatment, participants read and discuss the refutation text in small teams, as in the treatment of the field experiment.

In both treatments the refutation text is identical to the text used in the field experiment, except that we provide data on rents in Valencia, instead of Barcelona, since this is the city where participants live, with the purpose of using an environment close to them as we do in the field. We also add a short introduction to the concepts of supply and demand at the beginning of the text (see Appendix B.3). Laboratory participants are not familiar with these concepts since they have not had any exposure to economics courses, in contrast to participants in the field. This short introduction is also included in the control text. Thus, the only difference between the refutation and the control text in the laboratory is the approach used to explain price controls: textbook-based text in the control and the refutation text of the field in the treatment. Both texts include some questions. In the control text questions reproduce a typical textbook problem on market equilibrium and price controls; in the treatment text questions are identical to those in the field experiment. In the control and in the first treatment –i.e., no group discussion– participants have to individually answer these questions. In the second treatment –with group discussion– teams of three randomly chosen participants have to move to the desk of one of them to read and answer the questions. One of them answers the questions for the team and the other team members go back to their desks. Note that the decisions are only made in teams for the answers to the questions regarding the text. This is exactly parallel with the field.

As in the field experiment participants are asked to fill out a survey at different points during the sessions of the laboratory experiment. Some statements are the same as in the field (see Table A.2 in Appendix A). We introduce some changes in the survey in the laboratory experiment. First, the three statements concerning students' attitudes towards the content of the principles of economics course are excluded for obvious reasons.

Second, we include two new questions about participants' personal attitudes towards income distribution and equal opportunity to measure their correlation with the change in beliefs. Finally, participants are asked to answer the WT and the CRT. All the statements do not appear at every point of the laboratory experiment to introduce variation in the surveys in order to avoid memorization of answers as well as excessive survey length. Tables A.1 and A.2 in Appendix A show the statements and the point when they are presented to participants.

The sequence in the field and the laboratory experiments is the same; there is only a difference in the timing. In the field participants complete an initial survey at the beginning of the semester; around four weeks later they work with the text, and some weeks later they answer the final survey. In the laboratory, answering the initial survey and reading the text take place in the same session. We do it this way because exactly replicating the timing in the field experiment would have implied conducting three experimental sessions at different points in time with the same participants, which without an unaffordable budget would have led to a high attrition rate. In the laboratory the sequence of events is the same in all three conditions, the only variations pertaining to the type of the text about rent control and to the team discussion. The first session is planned to last 120 minutes, which is equivalent to the 90 minutes practice session plus the 30 minutes for completing the initial survey as in the field.

The experiment starts with the general instructions. Participants are told that the experiment consists of two sessions, the one they are attending at the moment and a second one that will take place three weeks later. They are told that after the first session they will be paid 15 euros (5 euros show-up fee and 10 euros for completing all the tasks) and after the second session they will be paid another 30 euros for completing the tasks.

Participants are also told that the second session will be shorter than the first. We pay twice as much for the second session to maximize the number of participants coming back.

The general instructions then go on to say, first, that participants will be asked to give their opinions about a number of social and economic issues, without there being correct or incorrect answers and without any influence on payments. Second, participants will have to read a short text and answer a number of questions. Lastly, participants will have to again answer a number of questions on economic and social issues.

After listening to the instructions, participants see on their screen the first survey, which includes the items indicated in column 4 in Table A.2. The statements and the WT appear in a particular order, but participants can move back and forth between them and all answers are not final until participants confirm their answers. Then, each participant is given a text (the refutation text or the control text) on paper, asked to read it and answer the questions individually or with her team depending on the treatment conditions.

In the final step of the first session, participants in all conditions are again asked to fill out a survey. Most of the statements are the same as those at the beginning of the session (see column 5 in Table A.2 in Appendix A). As before the questions appear on the screen in a particular order, but participants can answer them in any order since the answers are not final until the confirmation key is hit. After completing the tasks of the first session participants are paid 15 euros. They are then told on which dates they can come back to the laboratory for the second session. In the second session participants are first thanked for having come back and then are asked to answer the final survey, plus the WT and the CRT. After answering all questions participants are paid 30 euros and leave.

We use the estimated treatment effect in the field to select the sample size for the laboratory and to compute the statistical power. We set the total initial sample size for the laboratory equal to 180 individuals, 60 for each of the three conditions. With 180 individuals altogether in control and (each) treatment we get a power of around 72% to 75%.[13] While perhaps a little low, we face a trade-off between power and the cost of subject payments.

The final sample is 172, slightly smaller because of eight no-shows to the second session. These are practically equally distributed across conditions, so that the final subsamples are as follows: Control: N=57; first treatment, no group discussion: N=57; second treatment, group discussion, N=58. Table C.4 shows there are few significant differences in the distribution of demographic characteristics of control and treatment participants. The first session effectively took on average 114 minutes (maximum time is 126 minutes), and the second session took on average 50 minutes (maximum is 64 minutes).

5. Results

We first describe and compare beliefs and changes in beliefs in the field and the laboratory and present persuasion rates (section 5.1). We then report the results referring to the testing of hypothesis 1 to 3 (section 5.2). In section 5.3 we present evidence on the variation of treatment effects with respect to participants' initial beliefs. Section 5.4 shows results regarding the influence of cognitive factors (hypotheses 4 and 5), and section 5.5 reports the results of the follow-up experiment.

[13] We use the simulation method and STATA code proposed by Campos-Mercade (2018) to calculate the statistical power.

5.1 Beliefs, changes in beliefs and persuasion rates

We start with two preliminary issues. The first is whether the refutation text as such changes people's beliefs in the right direction. To study this we compare the response distribution before and after participants read the refutation text and test for statistical significance of opinion changes. This gives us a first indication of the persuasiveness of the refutation text. The second preliminary issue is how changes in beliefs in both treatment and control translate into the descriptive measure of persuasion proposed by Della Vigna and Gentzkow (2010).

Panel I in Table 1 reports the responses to the rent control statement in the field (assigned-to-treatment, compliers and control participants), at the beginning of the semester (initial beliefs) and at the end of the semester (delayed post-intervention beliefs).[14] Panel II reports the responses in the laboratory, for each of the three conditions and at three points in time: beginning of the first session (initial), end of the first session (immediate post-intervention) and second session (delayed post-intervention). The last two columns add up the two "agree", and the two "disagree" categories. For simplicity we focus on these two categories. Table 1 also shows changes in beliefs together with the results of t-tests of the null hypothesis of no change.

[Table 1 here]

In the field, at the beginning of the semester, more than two thirds of participants agree that rent controls will make housing available to more families: 69% in the control group, 77% in the treatment cohort, and 78% of compliers (Panel I.A). In the laboratory, the percentage of participants exhibiting the misconception is similar to that obtained in the field, if anything a little higher for all three conditions. Panel II.A shows that at the

[14] For the sake of brevity Table 1 does not report the distribution of responses of non-compliers but they are available upon request.

beginning of the first session 75% of participants believe that rent controls will make housing available to more families: 84% in the control group, 75% in the individual treatment and 83% in the group treatment.

Towards the end of the semester the percentage of participants in the control group in the field holding the misconception is even higher than at the beginning. Agreement with the statement increases by 6.2 percentage points (pp) while the percentage disagreeing falls by 3.8 pp. In contrast, among compliers the percentage agreeing drops by 16.9 pp while the percentage disagreeing increases by 17.3 pp. Hence among those exposed to treatment there is a response away from the misconception. The tests of the difference in means between final and initial surveys at the bottom of Panel I.A show that for the change of compliers we can strongly reject the null hypothesis at the 1% level, both for "sum agree" and "sum disagree". For changes in the control we cannot reject the null for "sum disagree" and for "sum agree" we can at the 10% level.

In the laboratory the distribution of beliefs immediately after the intervention – Panel II.B– shows that the percentages of participants disagreeing increases substantially in all three conditions, by 16 to 22 pp, and the percentages agreeing fall by around 21 pp. Three weeks after the intervention, figures show a small regression towards the misconception, but the percentage disagreeing with it is notably higher than the initial beliefs (Panel II.C). Comparing across individual and group conditions the impact of the treatment is stronger for the participants in the group discussion and it does not fade away to the same extent. In the control group, unlike in the field, we observe a change away from the misconception. We later explore potential explanations for this difference between the two control groups. The test results show that in the three groups changes are all in the right direction, significant at least at the 5% level, and stronger in the group treatment.

The delayed post-intervention change in the group condition in the laboratory is comparable to the delayed change in the field, where the refutation text is discussed in teams and the survey is conducted some weeks after the intervention. We can now state a preliminary result. In both environments the refutation text seems to be similarly effective. The percentage of participants disagreeing increases by a similar magnitude (17 pp) and the percentage of those agreeing decreases, somewhat more strongly so in the laboratory (22 pp) than in the field (17 pp). The refutation text *per se* appears to work in the field and in the laboratory.

Figures 1 and 2 complement the information provided in Table 1. They show the distribution of changes in beliefs by treatment condition in the field and in the laboratory. While the distribution for the field control group (panel A in Figure 1) is skewed to the left, meaning that beliefs change in the wrong direction, the distribution for the assigned-to-treatment group (panel B) is skewed to the right, reflecting the changes in the right direction among compliers (panel D). Figure 2 shows similar skewness to the right for both the individual and group treatment, although stronger in the latter.

[Figures 1 and 2 here]

The changes just described suggest that the refutation text is quite persuasive in the field. To get a feel of the magnitude of these changes we can compute the persuasion rate (Della Vigna and Gentzkow, 2010) and compare it with those obtained in previous studies on persuasion. The persuasion rate is defined as the percentage of receivers that change the behavior among those that receive a message and are not initially persuaded. It is computed as follows:

$$f = 100 * \frac{y_T - y_C}{e_T - e_C} * \frac{1}{1 - y_0}$$

where T and C refer to the treatment and control groups respectively; y_i is the share of group i adopting the behavior of interest, $i = \{T, C\}$; e_i is the share of group i receiving the message; y_0 is the share that would adopt it if there were no message. When y_0 is not observed it can be approximated by y_C. In our case, we define the behavior of interest as disagreeing with the rent control statement. We aggregate the totally disagree and disagree responses into one and obtain, in the field experiment, a persuasion rate of 14.7%. The studies reported in Della Vigna and Gentzkow (2010) yield persuasion rates between 0.01% and 20%, suggesting that our intervention is quite successful by this measure. In the laboratory, however, the immediate persuasion rate of both the individual and group treatment is close to zero. In the second session survey of the individual treatment the persuasion rate is zero; in the group treatment the persuasion rate is positive but low (1.7%).[15] The different persuasion rates in the field and in the laboratory derive from the good response of the control group to the non-refutational text in the laboratory: the number of participants in the control moving away from the misconception increases by 14 pp, while it falls by 3.8 pp in the field. The response of treated individuals is similar in the field and in the laboratory: the percentage disagreeing increases by 17 pp. This suggests that the different behavior of the control groups is driving the smaller persuasion rates in the laboratory. We will discuss the good response to the non-refutation text below.

5.2 Estimation results

The evidence presented in the previous section suggests a negative association between the refutation text and the prevalence of the misconception, especially in the

[15] In the field experiment, $y_T = 0.28$; $y_C = 0.156$; $e_T = 1$ and $e_C = 0$. In the laboratory the values are, for the immediate response of the individual treatment, $y_T = 0.28$ and $y_C = 0.31$; for the group treatment, $y_T = 0.31$ and $y_C = 0.31$. In the second session survey the values are $y_T = 0.24$ and $y_C = 0.24$ and $y_T = 0.26$ and $y_C = 0.24$, respectively for the individual and the group treatment.

field environment. Here we estimate the causal impact of the refutation text and thus test hypothesis 1, which states that the refutation text will have a positive impact on the misconception. To test it with the field data we estimate equation (1) using the OLS and the IV approaches described in section 4.1. Columns (1) to (4) in Table 2 show the results. Standard errors need to be cluster adjusted because of the design of the field experiment (Abadie et al., 2017). Clusters of units (complete classes) are assigned to the treatment or to the control group, and then all the students in the class either receive or do not receive the treatment. For OLS regressions we use the wild cluster bootstrap with the null imposed (or wild cluster restricted) proposed by Cameron, Gelbach and Miller (2008). For IV regressions we use wild restricted efficient bootstrap, which extends wild cluster bootstrap to IV linear models (Finlay and Magnusson, 2016). We have eleven clusters (classes), five in the assigned-to-treatment cohort and six in the control cohort (see Table C.2). [16] Wild cluster bootstrap can fail when there are few clusters in the sample. To assess whether the number of clusters is a problem we follow MacKinnon and Webb (2018) and compare the p-values obtained using both restricted (WCR) and unrestricted wild cluster (WCU) bootstraps. They yield very similar inferences, which indicates that the results are reliable. When the small number of clusters is an issue, WCR and WCU yield different inferences. Bootstrap p-values are reported in brackets in columns (1) to (4) in Table 2[17].

The OLS estimate of the treatment without control variables, 0.63, is significant and has a positive sign, indicating that in the treated group responses change in the right

[16] Average size of classes is similar across cohorts: 54.4 students (standard deviation 14.9) in the assigned to treatment group and 56.7 students (standard deviation 23.1) in the control group.

[17] We implement wild cluster bootstraps using the Stata post-estimation command "boottest" (Roodman et al., 2018). We use 999 bootstrap replicas and the Webb distribution for the auxiliary random variable as suggested by Roodman et al. (2018) when the number of clusters is small. P-values of the treatment coefficient in IV regressions are computed using the option "equaltail", which is appropriate when in a null test the assumption of symmetry around zero does not hold. This is the case with IV estimators, which are biased towards OLS estimators.

direction (from agreeing with the statement towards disagreeing), compared with the control group. Column (2) shows the results from the IV estimation of equation (1) using assignment to treatment as instrument. The IV estimate, interpreted as the ATT effect, is slightly higher (0.66). This indicates that not controlling for the bias from the non-compliance decision would underestimate the effect of the refutation text on changing the beliefs in the right direction. Since compliers are positively selected, as measured by the percentage of retakers but not by CAT grade (see section 4.1 and Table C.2), this would suggest that compliers might be harder to convince because they can think of better arguments to defend their initial view (Kahan et al., 2017). However, this bias has a limited role in driving the results since the difference between the IV and the OLS estimate is rather small.

The magnitude of the estimated coefficient falls only slightly after controlling for observed characteristics across cohorts (columns (3) and (4) in Table 2). For the IV estimates the ATT effect just falls from 0.66 to 0.65.[18] The estimated ATT is sizable: the treatment increases the change in beliefs in the correct direction by 0.65 points. This magnitude amounts to about 30% of the average response of compliers at the beginning of the semester (2.15), and to almost one half of the standard deviation of the change in beliefs (the standard deviation of y_i is 1.36). IV estimates in column (4) also show that higher CAT grades are positively correlated with the abandonment of the misconception, as are students in Law majors. Younger students are more resistant to doing so.[19] Thus,

[18] Results from the first-stage estimation show that the instrument is strong. The estimated coefficient is 0.89, significant at 1% level, with an F-test = 1138.5 (p-value = 0.00), and partial $R^2 = 0.826$ ($R^2 = 0.829$).

[19] We run the IV estimation of equation (1) separately for Law, Economics, and Business, the latter including the double major in Law and Business. In all cases the treatment effect is positive but it is only significant for Law and Business majors. The treatment effect is lowest for economics (0.35) and highest for business majors (0.74). For law students the treatment effect is in between, with a value of 0.63. Results are shown in Table D.1 in Appendix D. In brackets we report p-values obtained using wild bootstrap, without cluster, as suggested by MacKinnon and Webb (2018) because in the separate estimation the small

overall in the field environment the refutation text seems to trigger a significant change in beliefs.

[Table 2 here]

Columns (5) to (12) in Table 2 report the results of estimating equation (1) by OLS using the data generated in the laboratory environment. Standard errors are robust to heteroskedasticity. Columns (5) to (8) refer to the impact of the individual treatment. Surprisingly, the estimated immediate treatment effect is negative (columns (5) and (6)), although with a high standard error. Adding a similar set of control variables as in the field affects the absolute magnitude of the effect, but not its significance. Females, and students enrolled in engineering and science majors are more likely to move away from the misconception. The delayed impact (columns (7) and (8)) turns out to be positive, again with a large standard error, and the type of major becomes barely significant. These results suggest that in the laboratory environment the refutation text, when participants read it individually, does not do better than the non-refutational text we write for the control group, and that the change in beliefs may be mildly related to the subjects' choice of major.

Columns (9) to (12) in Table 2 show the impact of the group treatment. The immediate treatment effect is positive in this case, but again the standard error is large. As in the individual treatment, females and students in majors other than social sciences are more likely to pull out of the misconception, while non-Spanish participants are more likely to move towards it. Interestingly, the delayed effect is larger than the immediate effect, even after including control variables.[20]

number of clusters is an issue, unlike in the joint estimation. We use "boottest" with 999 bootstrap replicas, the Rademacher distribution for the auxiliary random variable and the option "equaltail".

[20] We run two regressions estimating jointly the individual and the group effects, one for the immediate change in beliefs and the second for the delayed change in beliefs as dependent variables. Results do not

In the field, and in line with results obtained in previous studies on misconceptions referring to psychology and to climate change (Tipett 2010; Nussbaum et al. 2017), our results indicate that the refutation text triggers a statistically significant change in beliefs. In the laboratory, the comparable results are the delayed effect of the group treatment (0.29). This magnitude amounts to about 15% of the initial average response of the laboratory participants in the group treatment (1.88), and to about one third of the standard deviation of the change in beliefs (the standard deviation of y_i is 0.9). Therefore, in the laboratory the size of the impact is about half as effective as in the field. The refutation text thus has a sizable positive impact both in the field and in the laboratory, although in the latter it is estimated imprecisely.

A potential concern regarding the estimated coefficient of the treatment dummy in the field is that it may capture the effect of external events that may have differed between the two cohorts. There are two types of external events. First, events related to the socio-economic environment, and second events related to the organization of majors and to our course. The first type of events might be a concern. We should note, however, that both 2015 and 2017 were similar in terms of the business cycle. Our guess is that external socio-economic events play a limited role. Regarding the second, there are no changes and in our course the only difference across cohorts is how we approach the topic of price ceilings.

To explore whether the different environments in the field and in the laboratory may influence the impact of the refutation text, we run two regressions of the form:

$$y_i = \gamma + \delta Lab_i + u_i \qquad (2)$$

change relative to the separate estimations shown in Table 2. We test the equality of individual and group effects in both regressions and cannot reject the null hypotheses. Results are available upon request.

The dependent variable, y_i, is the difference between final and initial beliefs of subject i (in the laboratory, final beliefs are the beliefs from the second session–delayed beliefs); Lab_i is a dummy variable indicating whether the subject participates in the laboratory experiment. In one regression the sample of observations consists of the complier participants in the field and in the laboratory group treatment. Column (1) in Table 3 shows that the change in beliefs of treated participants is not significantly correlated with the field versus laboratory setting. In a second regression, observations are the participants in the control groups of the field and the laboratory. Column (2) shows significantly different belief changes between control groups in the laboratory and in the field. This is consistent with the results of the t-tests shown in Table 1, and also with the lack of significant treatment effects in the laboratory shown in Table 2. Since the control in the laboratory does well, it is not possible to identify precisely the differential effect of the treatment.

[Table 3 here]

We believe that there are two potential explanations of why we cannot identify precisely the effect of the refutation text in the laboratory. First, it is possible that, as discussed above, the sample size used in the laboratory experiment lacks sufficient statistical power to measure precisely the effect of interest. Second, it is possible that in the laboratory participants in the control group allocate more attention to the control text than participants in the field attending a standard session on price controls. As a consequence the control text in the laboratory works rather well. Increased attention in the laboratory may stem from several sources. One is that participants find themselves in an unfamiliar physical environment that makes them be more attentive to any information. Two, the control text in the laboratory leads participants to focus only on

one topic, price controls, while the control group in the field is exposed to additional economic topics during the semester, hence their retentiveness may be affected. A third source of increased attention is the format information is delivered with. In writing the control text we tried to provide readers the same information that is provided viva voce in a lecture. Writing the text for the control group is indeed a delicate matter. We may have written a more complete explanation than what field participants in the control group really get. When writing the control text we decided to err on the side of excess rather than on the side of lack of information. We were worried that a positive impact of the intervention could be put into question because of a potential weakness of the control text.

As the specific content of the control text used in the laboratory already goes a long way towards changing the misconception, we below perform a follow-up treatment, where we test the effect of using the non-refutational text on the misconception in the field. We thus investigate the second and the third sources of increased attention explained above.

To sum up, our findings suggest that the refutation text changes an economic misconception in the field and in the equivalent treatment in the laboratory –the group treatment– supporting hypothesis 1. Compared to the control, the refutation text changes beliefs more strongly in the field, where the treatment effect is significant, than in the laboratory, where it is not significant. The good response of the control group in the laboratory, induced by the factors just mentioned above, may explain the lack of significance, hindering the identification of the effect of the refutation text. Finally, since the magnitude of the treatment effect is larger for the group treatment than for the individual, this may indicate that a team discussion induces a stronger departure from the

misconception. However, the estimated coefficients are not significant, therefore our findings do not reject hypothesis 2. Furthermore, our findings indicate that, if anything, the delayed effect is larger than the immediate effect, contrary to hypothesis 3.

5.3 Heterogeneity: Who changes her mind?

Results reported so far measure the average effect of the treatment. However, it is plausible to think that the effect may vary depending on the initial belief (agree, do not know, disagree). The issue is of interest since identifying who is harder to convince may give us insights into participants' thought processes and suggest where effort to convince should be exerted.

We analyze first the distribution of changes conditional on the initial answer –the transition matrices– in the field and in the laboratory. We split the sample into three groups according to the initial response –we aggregate totally agree and agree into one category, and totally disagree and disagree into another– so beliefs are measured on a three-point scale. Panel A in Table 4 reveals that, in the field, about twice as many students in the complier group as in the control switch from agreeing to disagreeing, or from not knowing to disagreeing. The reverse change, from disagreeing to agreeing –wrong change– is much smaller among compliers (36%) than in the control group (65%). Panel B shows the conditional distribution of changes immediately after the intervention in the laboratory. The persistence of the misconception among those initially agreeing in the control group is almost as high (69%) as in the control group in the field (79%), but the control text does better at reassuring those who initially disagree and preventing the wrong change (50%). The refutation text does even better in preventing the wrong change: in the group treatment the share of those who stick to their disagreement with the

misconception (60%) is somewhat higher than in the individual treatment (57%), and the latter is in turn higher than the share in the control group. Finally, Panel C shows that a few weeks after the intervention, the refutation text prevents more strongly than the control text the wrong change from disagreeing to agreeing (among control participants 67% change in the wrong direction, while the proportion among treated participants is just 14-20%). In addition, the refutation text does better than the control text at reassuring participants who initially disagree: among treated individuals 60-71% still disagree, while among control individuals the proportion falls to 17%.

All this shows that, although immediately after the intervention the control text works similarly to the refutation text, the latter promotes a more permanent change in beliefs. This suggests that the refutation text may induce a deeper understanding of the reasons why the misconception is wrong.

[Table 4 here]

We now estimate the causal effect of the treatment conditional on initial beliefs in the field and in the laboratory. Table 5 summarizes the results, by type of treatment (individual, group) and timing (immediate and delayed change). Columns refer to the initial belief, while rows refer to the type of experiment and treatment. The field regressions include control variables as in the baseline specification (Table 2). Given the small sample size in some of the laboratory regressions, we include only one control, CAT grade, as a proxy for individual ability. In the field the treatment has a positive and significant effect of 0.30 points on students who initially hold the misconception, moving them towards disagreeing with the statement, as intended. Thus the refutation text succeeds at inducing students who initially have the misconception to change their minds. The treatment does not have a significant effect on undecided. The effect on students who

initially disagree –column (3)– is positive and stronger (0.50) than for students who initially hold the misconception. This result suggests that the refutation text provides arguments that support these students' correct initial intuition, reassuring them and preventing them from changing over time towards the misconception. The smaller magnitude of the effect on students who initially hold the misconception, however, suggests that resistance to change is quite strong.

In the laboratory the estimated immediate effect for both the individual and group treatments on participants who initially agree with the statement is small and positive but not significant (0.03 and 0.02 respectively). The delayed impact on those initially agreeing is slightly negative in both cases, if not significant (-0.04 and -0.05 respectively). The impact of the estimated delayed effect on participants who initially disagree is positive, although only significant in the group treatment (1.29). This last result is in line with findings obtained in the field. Thus, both in the field and in the laboratory the refutation text has a differential effect according to initial beliefs. However, findings from the laboratory need to be taken with caution because of the small sample size in some regressions. Our results are consistent with Sunstein et al. (2017), who find that the effect of information about climate change in belief updating depends on the subjects' initial beliefs.

[Table 5 here]

5.4 Change in beliefs and cognitive performance tests

The results just described show that resistance to change the misconceived belief is quite strong. What factors might explain this resistance? As discussed in section 3 certain cognitive factors may provide an explanation: the propensity to experience confirmation

bias (hypothesis 4) and the propensity to engage in reflective versus intuitive thinking (hypothesis 5). To explore the role of these cognitive factors, we first look at their correlation with the change in beliefs in the laboratory, where we collect this information. Starting with confirmation bias, we compute WT scores as the percentage of correct responses to the three questions. Most participants (about 80%) did not provide a correct answer to any of them. The mean score at the beginning of the first session is 9% for the control and individual treatment conditions, and 6% for the group treatment.[21] Panel A in Table 6 shows the partial correlation between the Wason score measured in the first session and the immediate change in beliefs. Each column in the panel represents a separate regression, without and with control variables, for each treatment condition. Focusing on the regressions with controls, the WT only is significant and positive in the control group: higher scores are correlated with a change away from the misconception (column (2)). Estimates in Panel B show that the delayed change in beliefs is not significantly correlated with Wason scores measured in the second session for any treatment condition. The lack of significant results may be explained by the small variability of the WT scores since 80% of participants score zero points.

Panel C in Table 6 shows the correlation between the delayed change in beliefs and the CRT score; recall that this test was only used in the second session. We compute the score as the percentage of correct answers over the nine questions included in the CRT. The mean percentages are 40% for the control group, 43% for the individual treatment, and 45% for the group treatment. The CRT is positively correlated with the changes in beliefs for the control group (column (2), but not for the individual and group treatments.

[21] This is in line with the outcome reported in Wason (1977).

The pairwise correlation between CRT scores and Wason scores is very small (0.08) and not significant, suggesting that both measures capture different cognitive traits. When we include both measures jointly in the regression of the delayed change in beliefs, conclusions do not change (see Panel D in Table 6). CRT is positively correlated with the change in beliefs for the control group only and WT is not significantly correlated. Overall, these results suggest that a higher propensity to reflection is related to the change of beliefs, in the direction of decreasing the misconception, in the control group only.

[Table 6 here]

The fact that the CRT and WT are significant in some instances in the control but not in the treatments suggests that the refutation text may reduce the role of cognitive factors in the change in beliefs among the treated. We therefore estimate equation (1) for the delayed change in beliefs adding both cognitive measures and their interaction with the treatment:

$$y_i = \alpha + \beta_1 D_i + \beta_2 CRT_i + \beta_3 D_i CRT_i + \beta_4 WT_i + \beta_5 D_i WT_i + \varepsilon_i \quad (4)$$

Table 7 reports the results. For ease of comparison, columns (1) and (4) reproduce baseline results from columns (8) and (12) in Table 2. In all specifications we include the same set of control variables as in Table 2. In the case of the individual treatment, we do not find cognitive factors to be directly related to the change in beliefs (column (2)); the treatment effect is non-significant and close to the baseline. In column (3) we include the interaction terms of the treatment with the WT score and with the CRT score. The estimated coefficients of the CRT and its interaction are significant, with a positive effect on the control group (1.73), but a negative net effect on the treated (-0.64). WT remains non-significant and its interaction term is not significant either.

[Table 7 here]

In the model with interactions specified in equation (4) the marginal effect of the treatment for individual *i* varies with the values of CRT and WT scores as follows:

$$marginal\ effect_i = \beta_1 + \beta_3 CRT_i + \beta_5 WT_i \qquad (5)$$

At the means of CRT and WT, the marginal effect is 0.10, close to the baseline effect. Figures 3A and 4A show how the marginal effect of the individual treatment varies with the WT and CRT score, respectively (with the other score at its sample mean). First, the marginal effect does not change over the range of the WT score. Second, the marginal effect is higher the lower the level of the CRT, while it becomes negative for individuals that are more reflective, although the precision of the estimates falls quickly as the CRT score rises.

In the group treatment case, the estimated coefficient of the CRT is significant and positive, while the Wason score is not (column (5) in Table 7). Interactions of these variables with the treatment are not significant, suggesting that the treatment effect does not vary across treated participants with different cognitive scores (column (6)). These findings indicate that a higher CRT score is related with the change away from the misconception for both the control and the treated participants. At the sample means of CRT and Wason scores, the treatment effect is 0.18, slightly smaller than the baseline estimate in column (4). The marginal effect of the group treatment seems to become negative for higher WT scores; however the precision of the effect is very low as shown by the wide confidence interval (Figure 3B). The marginal effect of the treatment over the range of CRT shows a similar but less pronounced pattern as in the individual treatment (Figure 4B).

The bottom line is that to the extent that the Wason score captures the propensity to experience confirmation bias, our results in the laboratory do not support hypothesis 4.

We do not find evidence that confirmation bias is directly related to the change in beliefs nor that it interacts with the treatment, either in the individual or in the group treatments.[22]

The CRT, however, directly affects the ability to change beliefs in the three conditions. In the individual treatment the refutation text has on average a positive effect of 0.10 but this hides that the effect is higher for more intuitive participants and lower for more reflective individuals. This suggests that the individual treatment would offset the lower ability to change of less reflective individuals. In the group treatment the refutation text has a larger overall effect of 0.18 and does not vary as much with the CRT levels. This suggests that in the group treatment the refutation text would help everybody change, independently of his or her CRT score. These findings provide some support in favor of the refutation text being more effective when combined with the team discussion. Our results do not support the positive correlation posited in hypothesis 5, but challenge it.

A potential explanation for the finding that the refutation text is less persuasive for participants with high CRT scores is that they may be less likely to hold the misconception to start with, so they do not change their view. To explore this possibility, we regress initial beliefs on CRT and control variables by treatment condition (results are available upon request). We do not find evidence that the CRT score is significantly related to initial beliefs, suggesting that this explanation is not plausible in our case. An alternative explanation may be that more reflective participants are more resistant to change because they can think of better arguments in support of the misconception (Kahan et al., 2017). Unfortunately we cannot test this possibility with our data.

In addition to the influence of cognitive factors, recent evidence shows that emotions may affect a range of economic decisions (Guiso, Sapienza and Zingales, 2018; Breaban

[22] Findings are robust to using Wason scores measured at the beginning of the first session.

and Noussair, 2018). In Appendix D, we explore whether different emotions and attitudes interact with the treatment. We find that the refutation text is more persuasive than the control text for participants who feel more outraged about the high rents, especially when they discuss the text in teams.

5.5 The follow-up field treatment

In this follow-up treatment we explore whether the success of the non-refutational text in the laboratory may be attributed to increased attention. Participants in the field are now exposed to the same non-refutational text about price controls as the control group participants in the laboratory (see Appendix B.2). The experiment was conducted in the fall of 2019, i.e., two years after the field experiment with the refutation text. We thus maintain the two-year gap between all the cohorts involved in this research (2015, 2017 and 2019). This minimizes the presence of students previously exposed to the refutation text because of failing to pass the course in 2017. Like in the control and refutation-text cohorts, participants in the follow-up treatment are first-year college students enrolled in a principles of economics course in Business, Business and Law, Economics, and Labor Relations majors at UAB.

The course content is equal in the three cohorts. The procedure of the follow-up treatment is parallel to that of the refutation-text cohort. Participants fill out the same beginning and end of semester surveys to elicit their beliefs about rent controls. With the purpose of minimizing attrition, students who take both the beginning and the end of semester surveys receive a small bonus in their final course mark (0.2 points; on a 0-10 points scale). We ask students to report their true opinion, explicitly stating that they

would receive the bonus just for their participation, regardless of the answers to the surveys.

Around the fourth week of the semester, the topic on price controls is introduced in a lecture, like in previous cohorts. Then students have a practice session where they work with the non-refutational text. Like in the 2017 cohort, students are randomly allocated to teams of three and they are given one copy of the text each. They have to read this text individually and solve in teams two numerical problems about price controls. The first problem is equal to the problem that the control group in the laboratory solved. The second problem is a standard numerical problem on price controls very similar to the problems solved by the field 2015 control cohort. The session is graded with a weight equal to that of the refutation session in 2017.

In 2019 the response rate to both surveys is very high. Of the 344 students who respond to the beginning of the semester survey, around 91.9% (316 students) fill out the end of the semester survey. This response rate is larger than the response rates in the control and treatment cohorts (both around 67%). Table C.5 in Appendix C compares the distribution of the observed characteristics in 2019 with the characteristics in the 2015 and 2017 cohorts (columns (5) to (7)). We highlight that compared with previous cohorts, participants in the 2019 cohort exhibit lower academic ability as proxied by the lower average CAT grade and the higher proportion of retakers. As before we take these differences into account by including the vector of observed characteristics in the specifications below. Like in the field experiment in 2017, in the follow-up in 2019 some subjects do not comply with the treatment, i.e, they do not participate in the non-refutation session. Column (8) in Table C.5 shows that, like in 2017, compliers and non-compliers

in 2019 have different characteristics. As before we deal with this selection using IV techniques.

Table 8 shows the distribution of initial and final beliefs of 2019 compliers. At the beginning of the semester the prevalence of the misconception is similar to the previous cohorts, around 77%. By the end of the semester, students move towards the misconception: the percentage agreeing with the statement increases more (2.3 pp) than the percentage disagreeing (0.7 pp). The test of difference in means between the final and initial surveys shows that the changes are not statistically significant. Overall, the change of beliefs in 2019 is only slightly better than the one in 2015 and clearly worse than the change in 2017.

[Table 8 here]

Using the data from the follow-up experiment we estimate two specifications. We follow the empirical strategy explained in section 4.1. First, we estimate the effect of the non-refutational text (2019 cohort) compared with the effect of standard teaching (2015 cohort), that is, doing nothing special to deal with the misconception. The purpose of this regression is to understand whether students devote more attention to written material than to listening to a lecture, regardless of the refutational content of the written text. To this end we estimate equation (1) pooling the data from the 2015 and the 2019 cohorts. The dependent variable is the student's change of beliefs and the variable of interest is a dummy variable equal to one if a student is in the 2019 cohort and complies with the treatment –i.e. participates in the non-refutational session–, and zero otherwise. We deal with the potential selection from non-compliers by using intention-to-treat as instrument. The monotonicity assumption is satisfied, none from the 2017 cohort in enrolled in 2019.

We include the same set of observed variables explained in section 4.1 to control for differences in the characteristics of the students across the two cohorts.

Columns (1) to (4) in Table 9 show the results. In the specification with all the controls the IV estimate is positive and significant. This means that the non-refutational text has a positive impact on moving away from the misconception compared with a standard lecture. This finding suggests that organizing a specific session on price controls focuses students' attention on this issue leading to a decrease in the prevalence of the misconception. Thus, part of the positive effect of the refutation text in the field (column (4) in Table 2) may be attributed to an increased attention effect. However, since the estimate of the non-refutational text is 0.21, this suggests that the attention effect does not fully explain the estimated impact of the refutation text (0.64).

Second, we estimate the effect of the refutation text compared with the effect of the non-refutation text. The purpose of this regression is to assess the additional impact of the refutational content of the written text on dispelling the misconception. We estimate equation (1) pooling the data from the compliers of the 2017 and the 2019 cohorts –the students who actually participate in the refutational and in the non-refutational session, respectively. We regress the change in beliefs on a dummy variable equal to one if a participant complies with the refutational treatment, and zero if the participant complies with the non-refutational treatment.

Columns (5) and (6) in Table 9 show the OLS results. In the specification with all the control variables we find a positive and significant effect of the refutation text. This result suggests that the refutational content of the text has an additional impact on dispelling the misconception. The magnitude of the effect (0.27) is similar to the magnitude of the effect of discussing the refutation text in groups in the laboratory

(column (12) in Table 2). Recall that in the laboratory the control benchmark against which we estimate that effect is the change in beliefs induced by the non-refutational text.

[Table 9 here]

When interpreting results we should take into account that attrition is lower in the follow-up cohort than in the two previous cohorts. The differences in observed characteristics shown in Table C.5 may suggest also differences in unobserved variables, such as academic ability, motivation, conscientiousness, etc. Differences in unobserved characteristics across cohorts thus may affect the results although it is difficult to identify the direction of the potential bias, overestimating the effect of the refutation text in Table 2 (0.65) or underestimating the effects shown in Table 9.

Findings from the follow-up treatment suggest that the magnitude of the effect of the refutation text in Table 2 (0.65) reflects a combination of an attention effect and an additional refutation effect. In order to gauge the potential bias from unobserved variables, note that in the laboratory, where there is no attrition, the estimated effect of the refutation text versus the non-refutation text is 0.29, very close to the 0.27 found in the field when comparing 2017 and 2019 cohorts. This makes us think that the selection in unobservables has a limited role in driving the main results and that hence, refutation content has an additional impact on reducing the misconception.

6. Conclusions

Misconceptions about socio-economic issues, some of them bearing on economic policies of current interest, are widespread. They may foster the implementation of policies that have detrimental effects on well-being. Dispelling any popular misconception, whether on economic, health or other issues, is a challenge because

people may stick to previous beliefs even if they are provided well-grounded scientific information that contradicts them. In other scientific fields the concern about effectively communicating scientific evidence and, in particular, dispelling misconceptions, has become an active area of research.

Our research contributes new empirical evidence on the effectiveness of a particular communication tool, the refutation text, to help people replace intuitive false beliefs about a prominent economic topic by slow, analytical thinking. This approach, although previously used to debunk misconceptions in other socially relevant fields, has not been studied in the context of popular misconceptions about economic policies. In economics, awareness of the importance of cognitive biases when communicating economic knowledge is present but research is still embryonic. Our refutation text works in the laboratory and in the field, although in the former the effect is estimated imprecisely. Measured against a common benchmark, the non-refutational text, the effects of the refutation text are of a similar magnitude.

While tentative, our findings provide some insights for rethinking how to communicate economic evidence about socially relevant economic issues to the public. They also suggest that misconceptions should be explicitly addressed in economic principles courses in order to counter misconceptions based on cognitive biases. Delivering information with a refutational emphasis, may help in dispelling false beliefs more effectively than if presented in a standard non-refutational way.

In this research we do not investigate what specific elements of the refutation text are effective, i.e. stating the misconception, providing real data from a close example, empathy with inequality concerns about access to housing, providing alternative policies, etc. We test whether the refutation text as a whole leads to a decrease in the percentage

of students holding the misconception compared with a benchmark where students are not provided with these refutational arguments.

Further research paths open up. Some would be exploring how the refutation text for this particular misconception could be improved upon, which implies exploring the effect of specific elements. The proportion of those who stick to the misconception after the refutation text is still high in both the field and the laboratory, suggesting that there is room for identifying other formats that improve on it. Another question is whether our results would replicate elsewhere.

A broader issue is how to communicate social science research results when researchers' consensus is weaker or even consensus is non-existing, in the sense of experts' opinion being strongly divided. In some other cases, the majority of experts are uncertain about the effect of a policy, that is, there is consensus about uncertainty. In both of these cases, some questions that remain open are what should be communicated and how.

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017), "When Should You Adjust Standard Errors for Clustering?", NBER Working Paper No. 24003.

Allgood, S., W.B. Walstad and J. J. Siegfried (2015), "Research on Teaching Economics to Undergraduates", *Journal of Economic Literature*, 53(2), pp. 285- 325.

Andersson, R., and B. Söderberg (2012), "Elimination of rent control in the Swedish rental housing market: Why and how?", *Journal of Housing Research*, 21(2), 159-181.

Andrews, D., A. Caldera Sánchez and A. Johansson (2011), "Housing Markets and Structural Policies in OECD Countries", OECD Economics Department Working Papers, No. 836, OECD Publishing, Paris.

Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996), "Identification of Causal Effects Using Instrumental Variables," *Journal of the American Statistical Association*, 91, 444–72.

Angrist, J. D. and J. Pischke (2009), Mostly Harmless Econometrics: An Empiricist's Companion. Princeton: Princeton University Press.

Bensley, D. A., and S.O. Lilienfeld (2017), "Psychological misconceptions: Recent scientific advances and unresolved issues", *Current Directions in Psychological Science*, 26(4), 377-382.

Bloom, H. S. (1984), "Accounting for no-shows in experimental evaluation designs", *Evaluation review*, 8(2), 225-246.

Breaban, A., and C.N. Noussair (2017), "Emotional state and market behavior", *Review of Finance*, 22(1), 279-309.

Braasch, B., J. L., S.R. Goldman and J. Wiley (2013), "The Influences of Text and Reader Characteristics on Learning from Refutations in Science Texts", *Journal of Educational Psychology*, 105(3), 561-578.

Brañas-Garza, P., P. Kujal and B. Lenkei (2015), "Cognitive Reflection Test: whom, how, when", MPRA_paper_68049.

Broughton, S. H., G.M. Sinatra and R.E. Reynolds (2010), "The nature of the refutation text effect: An investigation of attention allocation", *The Journal of Educational Research*, 103(6), 407-423.

Busom, I.; C. Lopez-Mayan and J. Panadés (2017), "Students' persistent preconceptions and learning economic principles", *The Journal of Economic Education*, *48*(2), 74-92.

Byrnes, J. P., and K. N. Dunbar (2014), "The nature and development of critical-analytic thinking", *Educational Psychology Review* 26.4: 477-493.

Camerer, C. (2015), "The promise and Success of Lab-Field Generalizability in Experimental Economics: A critical Reply to Levitt and List", in Fréchette, G. R., and Schotter, A. (Eds.), Handbook of experimental economic methodology, Oxford University Press, USA.

Cameron, A. C., J. Gelbach, and D. Miller (2008), "Bootstrap-based improvements for inference with clustered errors", *Review of Economics and Statistics*, 90(3), 414-427.

Campos-Mercade, P. (2018), "Power analysis through simulations in Stata: a guide for dummies", unpublished manuscript. Downloaded from: https://sites.google.com/site/polcamposmercade/resources

Caplan, B. (2002), "Systematically biased beliefs about economics: robust evidence of judgemental anomalies from the Survey of americans and economists on the economy", *The Economic Journal*, 112: 433-458.

Cardoso, A. R., A. Loviglio and L. Piemontese (2016), "Misperceptions of unemployment and individual labor market outcomes", *IZA Journal of Labor Policy*, 5:13

Charness, G., and M. Sutter (2012). "Groups Make Better Self-Interested Decisions." *Journal of Economic Perspectives*, 26 (3): 157-76.

Charness, G., R. Oprea and S. Yuksel (2018), "How do people choose between biased information sources? Evidence from a laboratory experiment", manuscript, downloaded from: http://econ.ucsb.edu/~sevgi/CharnessOpreaYuksel_Dec2018.pdf.

Coibion, O., Y. Gorodnichenko and M. Weber (2019), "Monetary policy communications and their effects on household inflation expectations", *NBER Working Paper* No. 25482.

Cooper, D.J. and J.H. Kagel (2005), "Are two heads better than one? Team versus individual play in signaling games", *American Economic Review*, 95 (3), 477-509

CUSE (1997). Science teaching reconsidered: A handbook, Chapter 4. Committee on Undergraduate Science Education National Academy of Sciences (NAP), Washington, DC: National Academy Press.

Dal Bo, E., P. Dal Bo, and E. Eyster (2017), "The demand for bad policy when voters underappreciate equilibrium effects. *The Review of Economic Studies*, 85(2), 964-998.

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

Della Vigna, S., and M. Gentzkow (2010), "Persuasion: empirical evidence", *Annual Review of Economics 2*(1), 643-669.

Diamond, R. and T. McQuade (2019), "Who wants affordable housing in their backyard? An equilibrium analysis of low-income property development", *Journal of Political Economy*, 127(3), 1063-1117.

Druckman, J. (2015), "Communicating Policy-Relevant Science", PS: Political Science & Politics, 48(S1), 58-69.

Dunbar, Kevin, J. Fugelsang, and C. Stein (2007). "Do naïve theories ever go away? Using brain and behavior to understand changes in concepts", Chapter 8:193-2016, in M. C. Lovett and P. Shah (Eds.), Thinking with data. Carnegie Mellon symposia on cognition. Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Finlay, K., and L. M. Magnusson (2016), "Bootstrap methods for inference with cluster sample IV models", Working paper, University of Western Australia.

Frederick, S. (2005), "Cognitive Reflection and Decision Making", *Journal of Economic Perspectives*, 19 (4): 25-42.

Glaeser, E. L., and Luttmer, E. F. (2003), "The misallocation of housing under rent control", *American Economic Review*, 93(4), 1027-1046.

Golman, R., D. Hagmann, and G. Loewenstein (2017), "Information Avoidance", *Journal of Economic Literature*, 55 (1): 96-135.

Guiso, L., P. Sapienza, and L. Zingales (2018), "Time varying risk aversion", *Journal of Financial Economics*, 128(3), 403-421.

Gyourko, J., A. Saiz, and A. Summers (2008), "A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index". *Urban Studies* 45(3): 693-729.

Gyourko, J.E, and R. Molloy (2015), "Regulation and Housing Supply," in Duranton, Gilles, J. Vernon Henderson, and William C. Strange eds., Handbook of Regional and Urban Economics. Volume 5B: Elsevier, 1289-1337.

Haldane, A., and M. McMahon (2018), "Central bank communications and the general public", *AEA Papers and Proceedings,* 108, 578-83.

Herbst, D., and A. Mas (2015), "Peer effects on worker output in the laboratory generalize to the field", *Science*, 350(6260), 545-549.

Hilber, C A L and W. Vermeulen (2016), "The Impact of Supply Constraints on House Prices in England", *Economic Journal* 126(591): 358–405

Imbens, G., and J. Angrist (1994), "Identification and Estimation of Local Average Treatment Effects", *Econometrica, 62*(2), 467-475. doi:10.2307/2951620

Jacob, R., F. Christandl, and D. Fetchenhauer (2011), "Economic experts or laypeople? How teachers and journalists judge trade and immigration policies", *Journal of Economic Psychology*, 32 (5): 662–71.

Jamieson, K. H., D. Kahan, and D.A. Scheufele (Eds.), (2017), *The Oxford handbook of the science of science communication*. Oxford University Press

Johnston, C. D., and A.O. Ballard (2016), "Economists and public opinion: Expert consensus and economic policy judgments", *The Journal of Politics*, 78(2), 443-456.

Kahan, D. M., E. Peters, E.C. Dawson, and P. Slovic (2017), "Motivated numeracy and enlightened self-government", *Behavioural Public Policy*, 1(1), 54-86.

Kahneman, Daniel (2011). Thinking, fast and slow. New York: Farrar, Straus and Giroux. ISBN 9780374275631.

Kahneman, D., and A. Tversky (Eds.) (2000), Choices, values, and frames. Cambridge University Press.

Kendeou, P., J.L.G. Braasch, and I. Bråten (2016), "Optimizing conditions for learning: Situating refutations in epistemic cognition", *The Journal of Experimental Education* 84.2: 245-263.

Kowalski, P., and A.K. Taylor (2009), "The effect of refuting misconceptions in the introductory psychology class", *Teaching of Psychology*, 36, 153–159.

Kowalski, P., and A.K. Taylor (2017), "Reducing students' misconceptions with refutational teaching: For long-term retention, comprehension matters", *Scholarship of Teaching and Learning in Psychology*, 3(2), 90.

Krugman, P., R. Wells and K. Graddy (2015), Fundamentos de Economía, Editorial Reverté, third edition.

Lassonde, K. A., P. Kendeou, and E.J. O'Brien (2016), "Refutation texts: Overcoming psychology misconceptions that are resistant to change", *Scholarship of Teaching and Learning in Psychology,* 2(1), 62-74.

Lavecchia, A. M., Liu, H. and P. Oreopoulos (2016), "Behavioral Economics of Education: Progress and Possibilities", Handbook of the Economics of Education, Elsevier, 5, 1-74.

Levitt, S. D., J. A. List, S. Neckermann, and S. Sadoff (2016), "The behavioralist goes to school: Leveraging behavioral economics to improve educational performance", *American Economic Journal: Economic Policy*, 8(4), 183-219.

Lewandowsky, S., U.K. Ecker, C.M. Seifert, N. Schwarz and J. Cook (2012), "Misinformation and its correction: Continued influence and successful debiasing", *Psychological Science in the Public Interest*, 13(3), 106-131.

Levendusky, M.S., J.N. Druckman, and A. McLain (2016), "How group discussions create strong attitudes and strong partisans", *Research and Politics*, 3(2), 1–6.

Lilienfeld, S. O. (2010), "Confronting psychological misconceptions in the classroom", *Observer* 23 (7).

Lilienfeld, S. O., S. J. Lynn, J. Ruscio, and B. L. Beyerstein (2009), Fifty great myths of popular psychology: Shattering widespread misconceptions about human behavior. Chichester, UK: Wiley-Blackwell.

List, J.A. (2014), "Using Field Experiments to Change the Template of How We Teach Economics", *Journal of Economic Education*, 45(2), pp. 81-89.

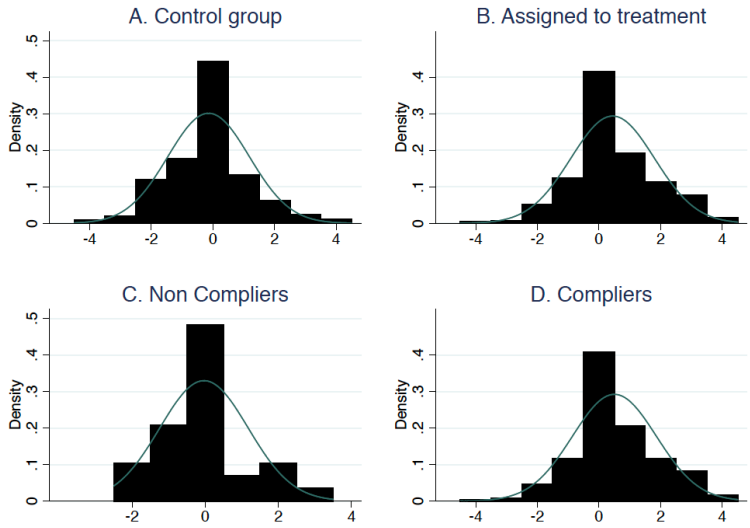
Lucariello, J., M. T. Tine, and C. M. Ganley (2014), "A formative assessment of students' algebraic variable misconceptions", *Journal of Mathematical Behavior* 33, pp. 30–41.

Lusardi, A and O.S. Mitchell (2014), "The Economic Importance of Financial Literacy: Theory and Evidence", *Journal of Economic Literature*, 52(1), pp. 5–44.

MacKinnon, J. G. and M. D. Webb (2018), "The wild bootstrap for few (treated) clusters", *Econometrics Journal,* 21, 114-135.
Mankiw, G. (2015), Principles of Economics, 8th Edition, Cengage Learning.
Masson, S., P. Potvin, M. Riopel, and L.M.B. Foisy (2014), "Differences in brain activation between novices and experts in science during a task involving a common misconception in electricity", *Mind, Brain, and Education*, 8(1), 44-55.
Molloy, R. (2020), "The effect of housing supply regulation on housing affordability: A review", *Regional Science and Urban Economics,* 80(C).
Mora-Sanguinetti, J.S. (2011), "The Regulation of Residential Tenancy Markets in Post-War Western Europe: An Economic Analysis", *European Journal of Comparative Economics*, 8(1), pp. 47-75.
Nakhleh, M.B. (1992), "Why some students don't learn chemistry: Chemical misconceptions", *Journal of Chemical Education*, 69(3), 191-196.
Nyhan, B., and J. Reifler (2010), "When corrections fail: The persistence of political misperceptions", *Political Behavior*, 32(2), 303-330.
Nussbaum, E.M., J. R. Cordova. and A. P. Rehmat (2017), "Refutation texts for effective climate change education", *Journal of Geoscience Education*, 65(1), pp. 23-34.
Pennycook, G., and Rand, D. G. (2019), "Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning", *Cognition, 188, 39-50.* https://doi.org/ 10.1016/j.cognition2018.06.011
Roodman, D., J. G. MacKinnon, M. O. Nielsen, and M. D. Webb (2018), "Fast and Wild: Bootstrap inference in Stata using boottest", Queen's Economic Department Working Paper No. 1406.
Sapienza, P., and L. Zingales (2013), "Economic experts versus average Americans", *American Economic Review*, 103(3), 636-42.
Stanovich, K.E., and R.F. West (2000), "Individual differences in reasoning: Implications for the rationality debate?", *Behavioral and brain sciences* 23.5 pp. 645-665.
Sunstein, C. R., S. Bobadilla-Suarez, S.C. Lazzaro, and T. Sharot (2017), "How people update beliefs about climate change: Good news and bad news", *Cornell Law Review*, 102, 1431.
Thaler, R. (2015), Misbehaving: The Making of Behavioral Economics. New York: W. W. Norton & Company. ISBN 978-0-393-08094-0.
Tippett, D. (2010), "Refutation text in science education: a review of two decades of research", *International Journal of Science and Mathematics Education*, 8 (6), pp 951–970.
Tirole, J. (2017). Economics for the Common Good. Princeton University Press.
Thomson, K. S., and D.M. Oppenheimer (2016), "Investigating an alternate form of the cognitive reflection test", *Judgment and Decision making*, 11(1), 99.
Toplak, M. E., R.F. West, and K.E. Stanovich (2014), "Assessing miserly information processing: An expansion of the Cognitive Reflection Test", *Thinking & Reasoning*, 20(2), 147-168.
Wason, P.C. (1960), "On the failure to eliminate hypotheses in a conceptual task", *Quarterly Journal of Experimental Psychology*, 12(3), 129-140.
Wason, P. C. (1977), "Self-contradictions", in Johnson-Laird, P. N.; Wason, P. C. (eds.). Thinking: Readings in cognitive science. Cambridge: Cambridge University Press. ISBN 978-0521217569.

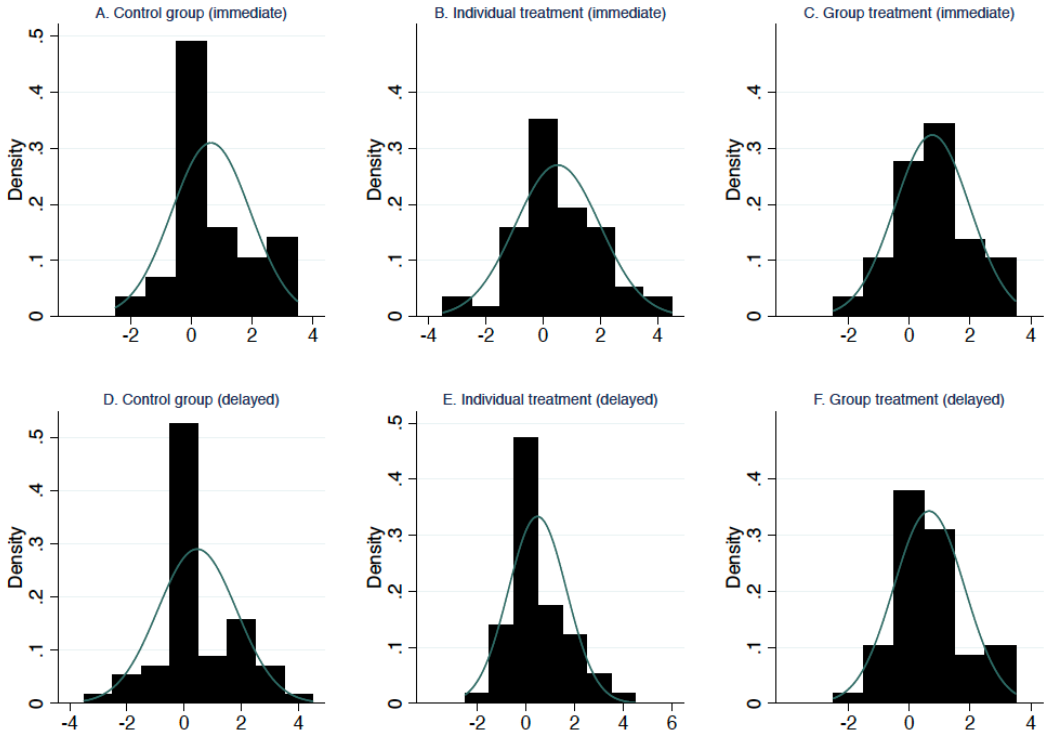
Wood, A. K., R.K. Galloway, and J. Hardy (2016), “Can dual processing theory explain physics students’ performance on the Force Concept Inventory?”, *Physical Review Physics Education Research*, 12(2), 023101.
World Health Organization (WHO), (2017), “How to respond to vocal vaccine deniers in public”.
Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. Experimental Economics, 13(1), 75-98.

Figure 1. Changes in beliefs on rent controls in the field experiment



Note: Positive values indicate a change away from the misconception

Figure 2. Changes in beliefs on rent controls in the laboratory experiment



Note: Positive values indicate a change away from the misconception. Immediate: Change in beliefs between the end and the beginning of the first session. Delayed: Change in beliefs between the beginning of the first session and the second session.

Figure 3. Marginal treatment effect at Wason Task values

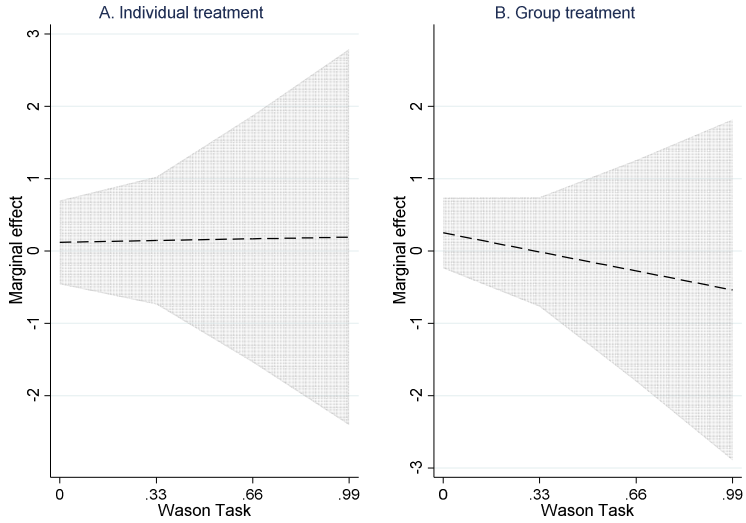
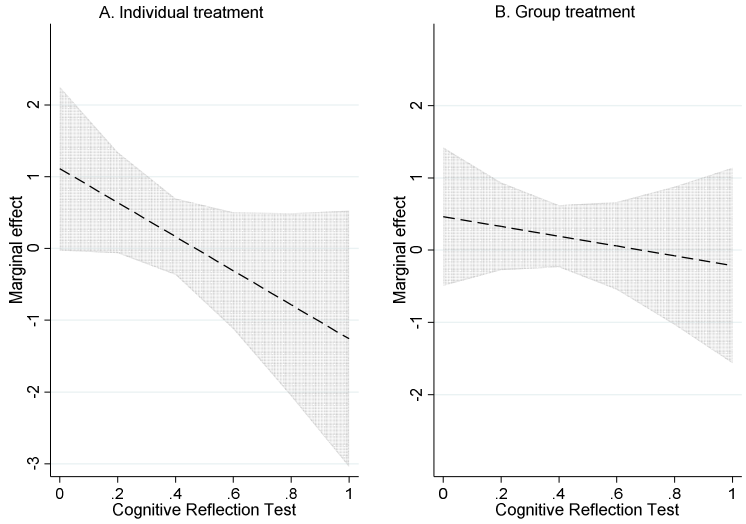


Figure 4. Marginal treatment effects at Cognitive Reflection Test values



Tables

Table 1. Beliefs about rent controls and change in beliefs (%)

I. Field experiment

	A. Initial						
	Totally disagree	Disagree	Agree	Totally agree	Do not know	Sum disagree	Sum agree
Control	3.2	16.2	56.2	12.9	11.5	19.4	69.1
Assigned-to-treatment	1.5	10.7	55.9	20.6	11.4	12.1	76.5
Compliers	1.7	9.9	57.2	20.6	10.7	11.6	77.8
	B. Delayed post-intervention						
	Totally disagree	Disagree	Agree	Totally agree	Do not know	Sum disagree	Sum agree
Control	5	10.6	56.5	18.8	9.1	15.6	75.3
Assigned-to-treatment	7.0	21.0	49.6	12.1	10.3	28.0	61.8
Compliers	7.8	21.0	50.2	10.7	10.3	28.8	60.9
Change Control					*-2.4*	*-3.8*	*6.2**
Change Compliers					*-0.4*	*17.3****	*-16.9****

II. Laboratory experiment

	A. Initial						
	Totally disagree	Disagree	Agree	Totally agree	Do not know	Sum disagree	Sum agree
Control	0.0	10.5	50.9	33.3	5.3	10.5	84.2
Individual treatment	1.8	10.5	52.6	22.8	12.3	12.3	75.4
Group treatment	0.0	8.6	44.8	37.9	8.6	8.6	82.8
	B. Immediate post-intervention						
	Totally disagree	Disagree	Agree	Totally agree	Do not know	Sum disagree	Sum agree
Control	7.0	24.6	45.6	17.5	5.3	31.6	63.2
Individual treatment	5.3	22.8	42.1	12.3	17.5	28.1	54.4
Group treatment	5.2	25.9	48.3	12.1	8.6	31.0	60.4
Change Control					*0.0*	*21.1****	*-21.1***
Change Individual treat.					*5.3*	*15.8***	*-21.1***
Change Group treat.					*0.0*	*22.4****	*-22.4****
	C. Delayed post-intervention						
	Totally disagree	Disagree	Agree	Totally agree	Do not know	Sum disagree	Sum agree
Control	3.5	21.1	47.4	21.1	7.0	24.6	68.4
Individual treatment	1.8	22.8	49.1	7.0	19.3	24.6	56.1
Group treatment	0.0	25.9	48.3	12.1	13.8	25.9	60.4
Change Control					*1.8*	*14.0***	*-15.8***
Change Individual treat.					*7.0*	*12.3**	*-19.3***
Change Group treat.					*5.2*	*17.2***	*-22.4****

Note: Sample sizes in Panel I: 340 in Control, 272 in Assigned-to-treatment, 243 in Compliers. Sample sizes in Panel II: 57 in Control; 57 in Individual treatment; 58 in Group treatment. Immediate post-intervention: answers immediately after receiving treatment (end of first session). Delayed post-intervention: answers some weeks after treatment. "Change" rows: Difference between % of students answering a given level of agreement in the corresponding final and initial surveys. Significance levels of t-tests of the difference in means between final and initial surveys: *10%, **5%, ***1%.

Table 2. Estimation Results

	Field				Laboratory							
					Individual Treatment				Group Treatment			
	OLS	IV	OLS	IV	Immediate	Immediate	Delayed	Delayed	Immediate	Immediate	Delayed	Delayed
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treatment	0.63***	0.66***	0.61***	0.64***	-0.14	-0.32	0.02	0.11	0.11	0.09	0.20	0.29
	[0.00]	[0.00]	[0.00]	[0.00]	(0.26)	(0.31)	(0.24)	(0.28)	(0.24)	(0.24)	(0.24)	(0.23)
Age[a]			-0.24*	-0.24*		0.12		0.17		0.10		0.19*
			[0.08]	[0.08]		(0.12)		(0.12)		(0.09)		(0.10)
Female			0.06	0.07		0.60**		0.18		0.56**		0.21
			[0.55]	[0.55]		(0.29)		(0.28)		(0.25)		(0.26)
CAT grade			0.13	0.14		0.09*		-0.01		0.08*		0.03
			[0.17]	[0.18]		(0.05)		(0.06)		(0.05)		(0.05)
Scholarship			0.13**	0.13**		0.37		0.16		0.12		0.10
			[0.01]	[0.02]		(0.28)		(0.28)		(0.26)		(0.27)
Nonspanish			-0.07	-0.07		0.18		-0.72		-1.77***		-1.01
			[0.47]	[0.45]		(0.94)		(0.63)		(0.65)		(0.79)
Retaker			-0.23	-0.23								
			[0.35]	[0.30]								
Law			0.47***	0.46**								
			[0.01]	[0.01]								
Economics			0.29	0.29								
			[0.12]	[0.11]								
HS track			-0.02	-0.03								
			[0.87]	[0.87]								
Health						0.37		0.09		0.59**		0.55*
						(0.36)		(0.28)		(0.28)		(0.28)
Engineering						1.27***		0.80*		0.92*		1.15***
						(0.47)		(0.44)		(0.49)		(0.42)
Science						1.29**		0.83*		1.27***		0.68**
						(0.56)		(0.49)		(0.36)		(0.29)
Humanities						0.54		0.40		0.97**		1.04*
						(0.40)		(0.41)		(0.38)		(0.54)
Constant	-0.13	-0.14	-1.29**	-1.30**	0.65***	-3.46	0.46**	-3.25	0.65***	-3.02	0.46**	-4.19**
	[0.31]	[0.30]	[0.01]	[0.02]	(0.17)	(2.46)	(0.18)	(2.25)	(0.17)	(1.95)	(0.18)	(1.95)
N	612	612	610	610	114	114	114	114	115	115	115	115
R^2	0.05	0.05	0.08	0.08	0.00	0.14	0.00	0.09	0.00	0.23	0.01	0.15

Note: Dependent variable is belief change and takes values between -4 and 4; positive values indicate a change away from the misconception. In field regressions the sample includes control and assigned-to-treatment observations. In brackets we report p-values obtained with wild cluster bootstrap restricted (with Webb weights for the auxiliary random variable). In laboratory regressions, the sample includes control and corresponding treatment observations. Robust standard errors are reported in parentheses. Significance levels: *10%, **5%, ***1%. [a]In field regressions "age" is a dummy variable that takes value 1 if the student is 18 years old and 0 otherwise.

Table 3. Field vs laboratory setting

	(1) Treated	(2) Control
Laboratory	0.15	0.60***
	(0.18)	(0.19)
Constant	0.50***	-0.14*
	(0.09)	(0.07)
N	301	397
R^2	0.00	0.02

Note: Dependent variable is the change between beliefs at the second session (laboratory) or end of semester (field) and initial beliefs. Robust standard errors in parentheses. Significance levels: *10%, **5%, ***1%.

Table 4. Transition matrices

	A. Field				
Initial belief	Agree	Don't know	Disagree	Total	
Agree	79.1	8.5	12.3	100	Control
	65.1	9.5	25.4	100	Compliers
Don't know	69.2	20.5	10.3	100	Control
	57.7	15.4	26.9	100	Compliers
Disagree	65.1	4.6	30.3	100	Control
	35.7	10.7	53.6	100	Compliers
	B. Laboratory: Immediate change				
Initial belief	Agree	Don't know	Disagree	Total	
Agree	68.8	2.1	29.2	100	Control
	55.8	20.9	23.3	100	Individual
	62.5	10.4	27.1	100	Group
Don't know	33.3	33.3	33.3	100	Control
	57.1	14.3	28.6	100	Individual
	60.0	0.0	40.0	100	Group
Disagree	33.3	16.7	50.0	100	Control
	42.9	0.0	57.1	100	Individual
	40.0	0.0	60.0	100	Group
	C. Laboratory: Delayed change				
Initial belief	Agree	Don't know	Disagree	Total	
Agree	70.8	2.1	27.1	100	Control
	65.1	16.3	18.6	100	Individual
	66.7	14.6	18.8	100	Group
Don't know	33.3	66.7	0.0	100	Control
	42.9	42.9	14.3	100	Individual
	40.0	0.0	60.0	100	Group
Disagree	66.7	16.6	16.7	100	Control
	14.3	14.3	71.4	100	Individual
	20.0	20.0	60.0	100	Group

Note: In the field, the sample size of the control group is 340; the sample size of the complier group is 243. In the laboratory, sample sizes are 57 in control group; 57 in individual treatment, and 58 in group treatment.

Table 5. Treatment effect conditional on initial belief

	(1) Agree	(2) Do not know	(3) Disagree
	A. Field: IV estimates		
Treatment	0.30**	0.19	0.50*
	[0.02]	[0.54]	[0.08]
N	441	70	99
	B. Laboratory: Individual treatment		
Immediate effect	0.03	-0.46	-0.59
	(0.18)	(0.45)	(0.50)
Delayed effect	-0.04	0.09	0.71
	(0.18)	(0.42)	(0.72)
N	91	10	13
	C. Laboratory: Group treatment		
Immediate effect	0.02	0.28	-0.05
	(0.18)	(0.68)	(0.86)
Delayed effect	-0.05	0.28	1.29**
	(0.17)	(0.79)	(0.45)
N	96	8	11

Note: Dependent variable is belief change, taking values between -2 and 2. We aggregate totally disagree and disagree into one category, totally agree and agree into another. Positive values indicate a change away from the misconception. Field estimates: Regressions include control variables as in Table 2. In brackets we report p-values obtained using wild bootstrap restricted (with Rademacher weights for the auxiliary random variable). Laboratory estimates: regressions include only CAT grade as control because of the small sample sizes. Robust standard errors are in parentheses. Significance levels: *10%, **5%, ***1%.

Table 6. Change in beliefs and cognitive performance in the laboratory

	Control group		Individual treatment		Group treatment	
	(1)	(2)	(3)	(4)	(5)	(6)
	A. Immediate change					
WT	1.32	1.01*	-1.34	-1.44	-0.08	-0.34
	(0.87)	(0.60)	(1.08)	(1.21)	(1.09)	(0.87)
R^2	0.04	0.56	0.03	0.17	0.00	0.26
	B. Delayed change					
WT	-0.69	-0.50	-0.69	-0.49	-0.75	-0.82
	(1.11)	(1.25)	(0.80)	(0.82)	(0.52)	(0.58)
R^2	0.01	0.24	0.01	0.18	0.01	0.25
	C. Delayed change					
CRT	1.56*	1.79**	0.23	-0.80	1.13*	1.17
	(0.88)	(0.89)	(0.91)	(1.02)	(0.63)	(0.85)
R^2	0.05	0.27	0.00	0.18	0.05	0.27
	D. Delayed change					
WT	-0.94	-0.60	-0.67	-0.71	-1.21*	-0.91
	(0.95)	(1.15)	(0.80)	(0.78)	(0.61)	(0.68)
CRT	1.68*	1.83*	0.12	-1.05	1.41**	1.23
	(0.89)	(0.94)	(0.90)	(1.04)	(0.62)	(0.84)
Control variables	No	Yes	No	Yes	No	Yes
N	57	57	57	57	58	58
R^2	0.07	0.28	0.01	0.20	0.08	0.28

Note: Each column in each panel represents a separate regression. Dependent variable is belief change and takes values between -4 and 4; positive values indicate a change away from the misconception. Results with control variables include the same set of explanatory variables as in Table 2. CRT: Cognitive Reflection Test. WT: Wason Task. Robust standard errors are in parentheses. Significance levels: *10%, **5%, ***1%.

Table 7. Change in beliefs, treatments and cognitive scores (delayed effect)

	Individual Treatment			GroupTreatment		
	(1) Baseline	(2)	(3)	(4) Baseline	(5)	(6)
Treatment	0.11	0.08	1.11*	0.29	0.17	0.54
	(0.28)	(0.28)	(0.61)	(0.23)	(0.22)	(0.48)
CRT		0.57	1.73*		1.47**	1.87**
		(0.73)	(0.89)		(0.60)	(0.85)
WT		-0.73	-0.96		-1.07	-0.67
		(0.71)	(1.13)		(0.73)	(1.09)
Treat.*CRT			-2.37*			-0.68
			(1.37)			(1.08)
Treat.*WT			0.07			-0.80
			(1.41)			(1.30)
Age	0.17	0.18	0.21*	0.19*	0.19*	0.20**
	(0.12)	(0.11)	(0.11)	(0.10)	(0.10)	(0.10)
Female	0.18	0.25	0.21	0.21	0.32	0.34
	(0.28)	(0.30)	(0.30)	(0.26)	(0.25)	(0.25)
CAT grade	-0.01	-0.01	-0.03	0.03	-0.00	-0.01
	(0.06)	(0.06)	(0.06)	(0.05)	(0.06)	(0.06)
Scholarship	0.16	0.19	0.11	0.10	0.10	0.05
	(0.28)	(0.30)	(0.31)	(0.27)	(0.27)	(0.29)
Nonspanish	-0.72	-0.57	-0.71	-1.01	-0.65	-0.70
	(0.63)	(0.59)	(0.48)	(0.79)	(0.77)	(0.75)
Health	0.09	0.06	-0.07	0.55*	0.39	0.39
	(0.28)	(0.28)	(0.29)	(0.28)	(0.29)	(0.29)
Engineering	0.80*	0.73*	0.78*	1.15***	1.03**	1.08**
	(0.44)	(0.44)	(0.43)	(0.42)	(0.43)	(0.43)
Sciences	0.83*	0.74	0.61	0.68**	0.49	0.50
	(0.49)	(0.48)	(0.56)	(0.29)	(0.35)	(0.33)
Humanities	0.40	0.31	0.20	1.04*	0.86	0.87
	(0.41)	(0.42)	(0.43)	(0.54)	(0.53)	(0.55)
Constant	-3.25	-3.39	-4.27**	-4.19**	-4.39**	-4.68**
	(2.25)	(2.20)	(2.06)	(1.95)	(1.99)	(1.94)
N	114	114	114	115	115	115
R^2	0.09	0.11	0.14	0.15	0.21	0.21

Note: Dependent variable is belief change and takes values between -4 and 4; positive values indicate a change away from the misconception. Columns (1) and (4) show baseline estimates from Table 2. CRT: Cognitive Reflection Test. WT: Wason Task. Robust standard errors in parentheses. Significance levels: *10%, **5%, ***1%.

Table 8. Beliefs about rent controls and change in beliefs in the follow-up treatment (%)

	A. Initial						
	Totally disagree	Disagree	Agree	Totally agree	Do not know	Sum disagree	Sum agree
Compliers 2019	1.97	7.21	42.95	34.43	13.44	9.2	77.4
	B. Delayed post-intervention						
	Totally disagree	Disagree	Agree	Totally agree	Do not know	Sum disagree	Sum agree
Compliers 2019	2.3	7.54	48.85	30.82	10.49	9.8	79.7
Change Control					*-2.4*	*-3.8*	*6.2**
Change Compliers 2017					*-0.4*	*17.3****	*-16.9****
Change Compliers 2019					*-3.0*	*0.7*	*2.3*

Note: Sample sizes: 340 in Control, 243 in Compliers 2017, 305 in Compliers 2019. Delayed post-intervention: answers some weeks after treatment. "Change" rows: Difference between % of students answering a given level of agreement in the corresponding final and initial surveys. Significance levels of t-tests of the difference in means between final and initial surveys: *10%, **5%, ***1%.

Table 9. Comparison with follow-up field treatment

	Control (2015 cohort) vs Follow-up (2019 cohort)				Compliers (2017 cohort) vs Compliers Follow-up (2019 cohort)	
	(1) OLS	(2) OLS	(3) IV	(4) IV	(5) OLS	(6) OLS
Non-refutation text	0.14	0.14	0.19	0.21*		
	[0.44]	[0.24]	[0.23]	[0.06]		
Refutation text					0.48***	0.27*
					[0.00]	[0.07]
Age		-0.12		-0.12		-0.24*
		[0.37]		[0.36]		[0.09]
Female		0.13		0.13		-0.10
		[0.34]		[0.35]		[0.32]
Retaker		-0.24		-0.25		-0.64
		[0.34]		[0.34]		[0.14]
CAT grade		-0.04		-0.03		0.10*
		[0.32]		[0.43]		[0.09]
Non-Spanish		0.07		0.07		0.10
		[0.66]		[0.66]		[0.39]
Law		0.43		0.42*		-0.05
		[0.11]		[0.08]		[0.85]
Economics		0.02		0.03		0.29**
		[0.92]		[0.84]		[0.04]
HS track		-0.14		-0.15		0.07
		[0.29]		[0.25]		[0.66]
Scholarship		-0.14*		-0.14*		0.18
		[0.09]		[0.07]		[0.26]
Constant	-0.12	0.37	-0.14	0.24	0.02	-0.74
	[0.38]	[0.40]	[0.29]	[0.57]	[0.83]	[0.17]
N	656	653	656	653	548	547
R^2	0.00	0.04	0.00	0.04	0.04	0.07

Note: Dependent variable is belief change and takes values between -4 and 4; positive values indicate a change away from the misconception. In brackets we report p-values obtained with wild cluster bootstrap restricted (with Webb weights for the auxiliary random variable). Significance levels: *10%, **5%, ***1%.

Appendices (for on-line publication)

Appendix A. The statements

Appendix B. The texts

Appendix C. Data description

Appendix D. Additional estimation results

Appendix A. The statements

Table A.1 contains all the statements used in the surveys. Subjects had to indicate their degree of agreement with the statement according to a 5-point scale: totally disagree; disagree; do not know; agree; totally agree. There is only one exception (*Innate ability vs Effort*), where the responses are: only effort, more on effort than on innate ability; equally on both; less on effort than on innate ability; only innate ability.

Table A.1. Statements used in the surveys

Statement
Rent control Establishing rent controls, such that rents did not exceed a certain amount of money, would increase the number of people who have access to housing facilities.
Minimum wage Raising the minimum wage from 650€ to 800€ per month would increase employment.
Mistrust Statistics Economic statistics do not reflect, in general, the true economic situation.
Fare evasion Skipping public transportation fares may be justified in some cases.
Tax evasion Prosecution of tax evasion should be increased, since taxes finance health, education, pensions and other social expenditures.
Buying home country If the home government bought only home products, employment would increase.
Innate ability vs Effort To understand and master a subject, how much depends on effort and how much on innate ability?
Value of personal experiences I believe my personal experiences allow me to understand many economic issues of our society.
Methods in Social Sciences Social sciences are not based on the same scientific method as natural sciences.
Concern use of math I am worried about the weight math may have in my degree
Outrage with respect to housing I feel outraged at the high price of rental housing in metropolitan areas
Lectures touch on real issues I think that lectures touch on real issues about our society
Learned unexpected content This course (principles of economics) has taught me unexpected things
Difficult course due to math It is hard for me to understand this subject (economics) because of the use of math.
Differences in income Nowadays in Spain the differences in people's incomes are too great Source: European Eurobarometer 2017
Equal opportunities Nowadays in Spain I have equal opportunities for getting ahead in life, like everyone else Source: European Eurobarometer 2017

Table A.2 indicates whether and when the statement is introduced in the field and in the laboratory experiment.

Table A.2. Statements included in each survey

	Field experiment			Laboratory experiment		
	Control	Treatment, beginning	Treatment, end	Session 1, beginning	Session 1, end	Session 2
Economic statements						
Rent Control	x	x	x	x	x	x
Minimum Wage	x	x	x	x	x	x
Buying home country	x	x	x	x	x	x
Emotions and attitudes:						
Mistrust Statistics		x	x	x	x	x
Fare evasion		x	x	x	x	x
Tax evasion		x	x	x	x	x
Innate ability vs effort		x			x	
Value of personal experiences		x		x		x
Methods in Social Sciences		x		x		x
Concern use of math		x			x	
Outrage with respect to housing		x		x	x	x
Lectures touch on real issues			x			
Learned unexpected content			x			
Difficult course due to math			x			
Differences in income				x	x	x
Equal opportunities				x	x	x
Wason Task				x		x
Cognitive Reflection Test						x

Appendix B. The texts

B.1 The refutation text

THE MARKET FOR RENTAL HOUSING
Or how things are often more complex than they seem

1. Introduction

Government intervention in the market usually aims at improving the society's well-being. However, government regulation sometimes produces outcomes that differ from the initially planned effects.

When the price of a good or service is too high or too low, harming either consumers or sellers, state and local governments frequently decide to impose a price control in that market.

There are two types of price controls: **price floors**, which set a minimum legal price to sell/buy the good (it is illegal to sell at a lower price); and **price ceilings**, which set a maximum legal price to sell/buy the good (it is illegal to sell at a higher price).

By imposing a binding price floor or price ceiling, the purchase/sale price in general will be different from the equilibrium price (the price resulting from the interaction between buyers and sellers). The reason is that the legal restrictions on the price are actually intended to reduce the purchase/sale price (when a price ceiling is imposed) and to increase the purchase/sale price (when a price floor is imposed).

2. Thinking slowly...

Price controls are highly controversial in public debates. Many people believe that regulating prices is a simple way of guaranteeing a certain level of income for everybody, or of guaranteeing the ability to consume certain goods and services.

Let us think more deeply about it and ask ourselves the following questions:

1. Are we sure that everybody will benefit from imposing a price floor or a price ceiling?
2. If not everybody is better off, who are the winners and who are the losers?
3. Is it possible that undesired and negative effects arise from a price ceiling?
4. Are there alternative policies that enable people to have access to the good or service?

These are some important questions that we need to consider when we think about the consequences of the price controls in competitive markets.

In practice, despite the good intentions of price controls, in many countries this type of policy did not solve the problem of lack of access to certain goods and services because this problem is after all related to how income is distributed across the population. Price controls can, on the contrary, create more serious problems. The supply and demand analysis of competitive markets is a useful tool to understand and anticipate the effects of price controls.

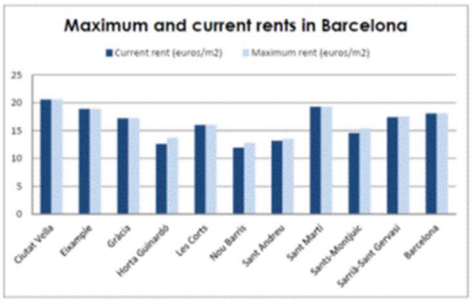
Does this mean that we cannot do anything? Not at all. What alternative policies can we use? Below we will see some examples.

Along the course we will learn more about income distribution and inequality and the different policies that deal with them.

3. The housing rental market: facts and reasoning

In April 2017, the newspaper "El Confidencial" published that rents had strongly increased, by more than 10%, in twelve Spanish capital cities. People who offer a house in exchange for a rent (i.e. the supply) see that the revenue from renting their properties rises. However, people who search for a house for rent (i.e. the demand) see that finding a house at an affordable price becomes increasingly hard.

Let us have a look at the data from Barcelona. In the following figure, we can observe the rental price (euros/m2) at different neighborhoods of Barcelona. That price ranges from 20.6 euros/m2 at Ciutat Vella to 11.9 euros/m2 at Nou Barris (average prices). This means that renting an apartment of 80 m2 will cost on average 1648€ per month at Ciutat Vella, and 952€ per month at Nou Barris.



Source: El Confidencial

On the other hand, according to the Tax Office, the annual gross average income per capita in Barcelona in 2016 was 32954€. This means that the average tenant would expend on renting 60% of the annual gross average income per capita if she lives in the most expensive neighborhood. However, she would spend 34.6% of the annual income on rent if she lives in Nou Barris.

Cities, such as Stockholm, Paris, Marseille, New York, San Francisco, Toronto …have experienced this problem before. In some of them local governments decided to impose a rent ceiling, in some cases since 1940 (New York).

What do we know about the real effects of this policy in these cities?

1. Research shows that in Stockholm and other cities there are long waiting lists of people that want an apartment for rent; that is, lines have appeared. People in Stockholm have to wait around 10 years to have an apartment (see, among others, the work by researchers Andersson and Soderberg, 2012). This means that many people cannot access an apartment for rent when they need it, since apartments are allocated first to people who have been in the waiting list for a longer time. These people are not always the ones who need the apartment the most at a given time.
2. Research also shows that black markets for rent housing arise. This implies that tenants end up paying a rent above the price ceiling: either they pay more to advance positions in the waiting list (which is a type of corruption), or they pay more (illegally) to get the apartment.

3. Finally, apartments initially intended for rent are often sold or remain empty, because real estate owners consider that the price they can charge (the price ceiling) is not high enough to offer their apartments for rent.

As time goes by, we observe that the quantity of apartments available for rent decreases (there is a supply shortage) and the housing problem gets worse. Low-income families are the most harmed by a policy that was intended to help them. The richest families however gain from the policy because they can pay more money to get a house: either they can pay money to advance positions in the waiting list or they can (illegally) pay an extra rent (black money) to the owner of the apartment.

These types of problems frequently appear in all the cities that impose rent ceilings. For this reason researchers study and propose policies other than price ceilings that avoid the previous negative effects and that really benefit the lowest income families. Some examples are:

1. Building and renting public housing, where access to this housing would be determined by a family's income (families with income above certain level would not be able to apply). This policy is aimed at increasing the supply of housing for rent, and thus it affects the rent.
2. Imposing taxes to the owners of non-used land. Thus, owning the land will cost money, and some owners will decide to sell or rent that land to build public or private housing. This policy is also aimed at increasing the supply and therefore it will decrease the rent.
3. Providing direct subsidies to the families whose income is below a given threshold.

Note that policies 1 and 2 increase the supply and so they will move down the price. This will imply that more low-income families could rent a house.

4. Discussion (answer the questions in the template provided)

Using the information provided in the previous sections and your own knowledge on the topic, answer the following questions. Reason your responses.

1. Mention two possible causes that can explain the observed increase in rents. Illustrate your answer with a graphical analysis that uses the supply and demand model.

2. Suppose that the Barcelona's City Council considers introducing a rent control in order to stop the increase in rents. This new regulation will set a price ceiling of 15€ in all of Barcelona's neighborhoods. What are the effects this rent control will have on the quantity of apartments for rent supplied, on the final quantity of rented apartments and on the rents paid in Ciutat Vella? And in Nou Barris? Use the supply and demand model to explain these effects. Illustrate your answer with a graphical analysis.

3. Write a list with the negative and positive consequences that can potentially arise from this rent control in Barcelona. Discuss what consequences may be more serious for Barcelona. Do you think that this rent control will allow more people to have access to houses for rent?

B.2 The non-refutational text for the laboratory experiment and for the follow-up field treatment.

Public intervention: price controls.
Market regulation by public institutions in principle has the objective of improving the welfare of society. Sometimes, however, the effects that these decisions have may not be those initially desired.

Normally price controls are imposed when governments or city halls consider that the market price of a good or service is too high or too low, hurting either buyers or sellers. There are two types of price controls: price floors, which impose a legal minimum for the purchase/sale of a good (it is not legal to sell at a lower price), and price ceilings, that establish a legal maximum for the purchase/sale of a good (it is not legal to sell at a higher price).

When a price floor or ceiling is established, the purchase/sale price will in general not be equal to the equilibrium price that would emerge from the interaction between buyers and sellers. The government regulates the market precisely with the purpose of making the product or service cheaper (when it establishes a price ceiling because it believes that the market price is too high) or more expensive (when it establishes a price floor because it believes that the market price is too low).

To understand the workings of price controls and how they influence market results let us start by, for example the market for bread. If bread is sold on a competitive market, free of government regulation, its price will adjust to balance supply and demand: at the equilibrium price, the quantity of bread that buyers are willing to buy is exactly equal to the quantity that sellers are willing to sell. To be concrete, suppose that the equilibrium price reaches 1.5 euros per loaf of bread. It is possible that not everybody is satisfied with the results of the free market process. Suppose that the Association of Bread Consumers complain, since it believes that the price of 1.5 euros per loaf is too high to allow everybody to buy one loaf of bread per day. At the same time, the National Organization of Bread Sellers disapproves of the price of 1.5 euros, because it is too low and reduces the earnings of sellers.

Each of these groups pressures the government or the pertinent public institution to pass laws that modify the market results through the direct control of the price of a loaf of bread. Due to the fact that buyers of any good always want the lowest price, whereas sellers always look for the highest price, the interests of the other groups are in conflict. If the Association of Bread Consumers is successful in their pressure, the government will impose a maximum legal price at which to sell each loaf of bread. This means that selling a loaf of bread above this price will not be allowed, therefore this maximum authorized bound will be called price ceiling.

This bound is typically fixed in a way that the price ceiling is lower than the equilibrium price in the market for bread (remember what has been explained above about supply,

demand and the equilibrium price of a market. The objective of this policy is to make a loaf of bread more affordable so that everybody can buy this product.

What are the effects of fixing a price ceiling?

On the supply side some sellers consider that the price ceiling is too low and react by bringing less bread to the market. They can also decide to produce bread of lower quality (with worse quality flour for example). On the demand side the price ceiling encourages more people to buy bread every day (instead of buying it from time to time). Hence, as time goes by, the quantity of bread loafs supplied tends to fall and the quantity of bread loafs demanded tends to increase. The result is scarcity of bread loafs.

Rent control is another common example of a price ceiling. In many cities local government imposes a cap on rents that the lessor (housing owners) can charge the tenants. This cap is typically fixed such that the price ceiling is below the equilibrium price of the rental housing market. Just like in the case of bread, fixing a price ceiling for rents generates scarcity of rental housing.

Therefore, the main effect of a price ceiling is to generate scarcity of the product. Demand will be larger than supply and this will happen in any market, not only in the case of breads or housing, but also in others like taxis or gasoline. As a consequence of the scarcity of the product the price of which has been regulated, in many cases a black market arises in which the product is bought and sold outside of the official market to circumvent the price ceiling regulation.

The negative effects of the price ceiling are not evident for the general population, especially at the beginning, since these effects manifest themselves as time goes by. For these reason, economists frequently criticize the use of a price ceiling, since they consider that it is a very inefficient way of helping that everybody can acquire a product.

Once you have read the text answer the questions you will find in the template we have handed you. You can go back to read the text as many times as you want.

Problem 1 (included both in the control group of the laboratory experiment and in the follow-up field treatment):

A recent study shows that the supply and demand of ham sandwiches in cafeterías in Barcelona are the following:

Price per unit (in thousands)	Quantity demanded (in thousands)	Quantity supplied (in thousands)
6	1	15
5	2	12
4	4	9
3	6	6
2	8	3
1	10	1

a) Graphically represent supply and demand. What is the equilibrium price and what is the equilibrium quantity?

b) Some college students start a campaign directed at the Barcelona city government to require a reduction it the price of sandwiches. The city government accepts to set a price ceiling of 2€ per sandwich. What will the new market price be and how many ham sandwiches will be consumed? Represent this situation in the graph of part a).

Problem 2 (included in the follow-up experiment in the field, but not in the laboratory)
The demand function and the supply function of a certain good are given by the following equations:
qd = 26-2p
qs = 4p-10

a) Do these demand and supply functions follow the law of demand and supply? Why?
b) Define the concept of market equilibrium.
c) Calculate numerically the market equilibrium (the price and the quantity) and represent it graphically.
d) Suppose that the government decides to set a price ceiling of 5 euros in this market.
- How many units will be exchanged and at what price?
- Is there a market equilibrium?

Reason your answer numerically and intuitively. Represent the solution graphically.
e) Suppose that the government changes its economics minister and that the new minister decides to change the price ceiling and fix it at 7 euros per unit.
- How many units will be exchanged and at what price?
- Is there a market equilibrium?

Reason your answer numerically and intuitively. Represent the solution graphically.

f) After a new change of ministers the new economics minister decides to set a price floor of 7 euros in this market.
- How many units will be exchanged and at what price?
- Is there a market equilibrium?

Reason your answer numerically and intuitively. Represent the solution graphically.

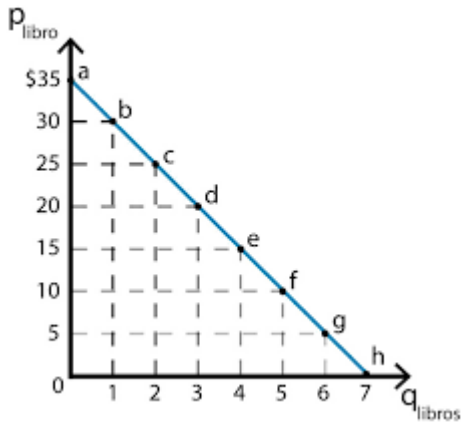
B.3. Introduction of supply and demand for the laboratory experiment

1. Introduction: buyers, sellers and markets.
Economists use the supply and demand model to analyze the workings of markets. In every market of a good or a service there are always two sides: buyers (the people who wish to buy the product) and sellers or producers (the people or companies that supply the product).

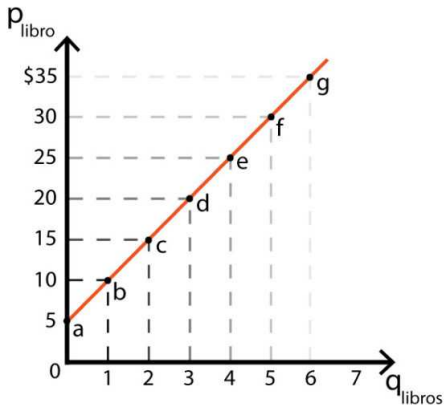
On the buyers side, and given the income of each of them, as the price of a good goes down, the quantity demanded increases. This relation is known as the market demand. One can visualize graphically the relation between price and quantity. For example, in Panel A of the figure below illustrates what would happen, in a given town, with the quantity of books people wish to buy as a function of the price per book. If the price of a book is 25€, 2 books will be demanded (point c in the figure). And if the price of the book falls to 10€, then the quantity of books demanded will rise to 5 books (point f in the same figure).

On the sellers side, as the price of a good goes up the quantity supplied increases, either because more sellers appear for whom it is worth selling, or because each of them wants to sell a larger quantity of the good. This relation between price and quantity supplied is known as market supply. One can visualize graphically the relation between, price and the quantity supplied. Panel B in the figure below shows the quantity of books sellers are willing to sell as a function of the price per book. For instance, if the price of a book is 10€, only 1 book will be supplied (point b). If the price increases to 25€, we see that the quantity of books supplied will rise to 4 (point e).

Panel A



Panel B

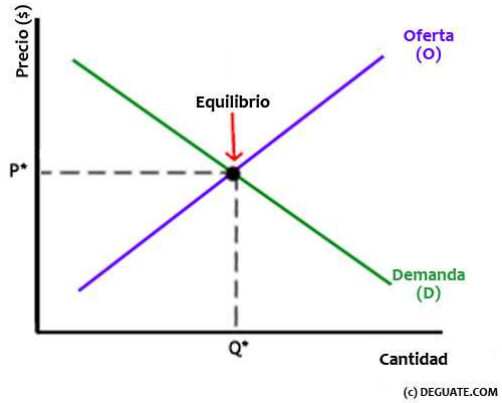


How is the price per book paid by buyers and received by sellers determined? If the price of a good or service is very high such that there are more sellers (or quantity supplied) than buyers (or quantity demanded), sellers will have to lower the price if they want to sell. If instead the price of the good or service is very low, there will be few sellers willing to supply the good, while there will be many buyers willing to buy it but they will be unable to do so because only a smaller quantity will be supplied. That is, there will be

product scarcity. In this case, the price will tend to increase, therefore attracting more sellers (or to a higher quantity being supplied).

In a market without public regulation, prices will go up if there is scarcity and will go down if there is too much of the product, and a final price will be reached at which the number of buyers and sellers will be the same. This price is called the equilibrium price and at this price the quantity that buyers want to buy and the quantity that sellers want to sell coincides exactly. Panel C in the figure below places the demand and supply curves on the same graph and illustrates the equilibrium price: that is, the price where supply and demand coincide (that is, they intersect).

Panel C



B.4 Instructions given to participants in each condition.

Field participants, 2015 cohort: Practice session, where students solve in class standard supply and demand problems, including price controls.

Field Participants, 2017 and 2019 cohorts. Just preceding the respective refutational and non-refutational texts, the following paragraph is included:

- *This practice session takes place in the class-room. The text and questions will be discussed and solved in groups. Each group will hand in the solution to the case in the template prepared for this purpose, at the end of the time indicated at the beginning of the session. The maximum time for solution is 50 minutes. You are not allowed to discuss with members of other groups. All members of one group will obtain the same grade.*
- *Read the text and answer all the questions in the template we have given you.*

Laboratory Participants, Control and Treatment groups: Just preceding the respective non-refutational and refutational texts, the following paragraph is included:

In what follows we present a text to you that explains how a market works and what happens when prices are controlled. We ask you to read it carefully and then answer some questions on the computer to verify the comprehension of the text. Both tasks can take you approximately 1 hour.

Appendix C. Data description

Table C.1. Field. Students taking the survey

	Cohort 2015		Cohort 2017	
	Students		Students	
Only beginning of semester	128	25.2	103	25.8
Only end of semester	40	7.9	24	6.0
Both surveys	340	66.9	272	68.2
Total sample size	508	100	399	100

Table C.2. Field. Demographic characteristics of students who answered both surveys (%)

	Cohort 2017			Cohort 2015	t-test	
	(1) Assigned to treatment	(2) Treated (compliers)	(3) Non treated (non-compliers)	(4) Control	(5) Difference (4) – (1)	(6) Difference (3) – (2)
Female	47.1	49.4	27.6	47.9	0.8	-21.8**
Age: 18 when enrolled	70.2	70.7	65.5	62.1	-8.1**	-5.2
Average CAT grade	9.2	9.2	9.0	9.2	0.0	-0.2
	(1.5)	(1.5)	(1.2)	(1.3)		
Non-Spanish citizen	9.6	8.6	17.2	6.2	-3.4	8.6
Retaker	2.6	1.6	10.3	4.1	1.5	8.7**
Receives a scholarship	30.5	30.9	27.6	39.4	8.9**	-3.3
Average course grade	5.4	5.5	4.2	5.7	0.3**	-1.3***
	(1.8)	(1.7)	(2.1)	(1.6)		
Track of origin:						
High school	81.9	82.6	75.9	74.7	-7.2**	-6.7
Vocational school	12.6	12.4	13.8	15.9	3.3	1.4
Other	5.5	5.0	10.3	9.4	3.9*	5.3
Major:						
Business + Law	21.7	20.6	31.0	19.1	-2.6	10.4
Business	26.1	27.1	17.3	37.7	11.6***	-9.8
Economics	22.1	19.8	41.4	23.8	1.7	21.6***
Law/Labor Relations	30.1	32.5	10.3	19.4	-10.7***	-22.2**
N	272	243	29	340		
Number of classes	5			6		

Note: CAT refers to the College Admission Test. Standard deviation in parentheses. Significance levels: *10%, **5%, ***1%.

Table C.3. Field. Assignment to treatment and Treatment compliance

	Assigned to control (Z = 0)	Assigned to treatment (Z = 1)	Total number of students
Compliers (D = 1)	0	243	243
Non-compliers (D = 0)	340	29	369
Total students	340	272	612

Table C.4. Laboratory. Demographic characteristics of students who attended both sessions

				t-test	
	(1)	(2)	(3)	(4)	(5)
	Control	Individual	Group	Difference (1) – (2)	Difference (1) – (3)
Female	57.9	71.9	70.7	-14.0	-12.8
Age	19.5	18.8	19.3	0.8***	0.3
	(1.6)	(0.8)	(1.1)		
CAT grade	9.8	10.5	10.1	-0.7	-0.4
	(2.5)	(2.4)	(2.1)		
Non-Spanish	1.8	5.3	5.2	-3.5	-3.4
Receives a scholarship	12.3	17.5	24.1	-5.3	-11.9
Track of origin					
High School	91.2	93.0	93.1	-1.8	-1.9
Vocational school	8.8	7.0	5.2	1.8	3.6
Other	0.0	0.0	1.7	0.0	-1.7
Major					
Health	33.3	45.6	34.5	-12.3	-1.2
Engineering	12.3	22.8	12.1	-10.5	0.2
Science	12.3	7.0	19.0	5.3	-6.7
Social sciences	26.3	12.3	29.3	14.0*	-3.0
Humanities	15.8	12.3	5.2	3.5	10.6*
N	57	57	58		

Note: CAT refers to the College Admission Test. Standard deviation in parentheses. Significance levels: *10%, **5%, ***1%.

Table C.5. Follow-up field. Demographic characteristics of students who answered both surveys (%)

	Cohort 2019			t-test			
	(1) Assigned to treatment	(2) Treated (compliers)	(3) Non treated (non-compliers)	(5) Difference 2015 – 2019	(6) Difference 2017 – 2019	(7) Difference compliers 2017 – 2019	(8) Difference 2019 (2) – (3)
Female	48.4	48.5	45.5	-0.5	-1.4	0.9	-3
Age: 18 when enrolled	67.7	68.9	36.4	-5.6	2.5	1.8	-32.5**
Average CAT grade	7.7	7.7	7.9	1.5***	1.5***	1.5***	0.2
	(1.3)	(1.3)	(1.4)				
Non-Spanish citizen	13	12.1	36.4	-6.8***	-3.4	-3.5	24.2**
Retaker	10.4	9.5	36.4	-6.3***	-7.9***	-7.9***	26.9***
Receives a scholarship	47.8	48.5	27.3	-8.4**	-17.3***	-17.7***	-21.3
Average course grade	5.5	5.5	4.8	0.2*	-0.1	0.0	-0.7
	(1.4)	(1.4)	(1.6)				
Track of origin:	88.3	89.2	63.6				
High school	7.9	7.5	18.2	-13.6***	-6.4**	-6.5**	-25.5***
Vocational school	3.8	3.3	18.2	8***	4.7*	4.9*	10.6
Other				5.6***	1.7	1.7	14.9**
Major:	27.2	27.5	18.2				
Business + Law	30.4	30.5	27.3	-8.1**	-5.5	-7.1*	-9.4
Business	27.5	26.6	54.5	7.3**	-4.3	-3.4	-3.2
Economics	14.9	15.4	0	-3.7	-5.5	-6.8*	27.9**
Law/Labor Relations	88.3	89.2	63.6	4.5	15.3***	17.1***	-15.4
N	316	305	11				
Number of classes	5						

Note: CAT refers to the College Admission Test. Standard deviation in parentheses. Significance levels: *10%, **5%, ***1%. Column (5) shows the difference with the control cohort in the field (column (4) in Table C.2). Column (6) shows the difference with the 2017 assigned-to-treatment cohort (column (1) in Table C.2). Column (7) shows the difference with compliers from the 2017 cohort (column (2) in Table C.2). Column (8) shows the difference between compliers and non-compliers in the 2019 cohort.

Appendix D. Additional estimation results

Table D.1 Field. Estimation results by major

	IV-Law	IV-Econ	IV-Bus
Treatment	0.63	0.35	0.74***
	[0.14]	[0.18]	[0.01]
Age 18	-0.38	-0.28	-0.21
	[0.48]	[0.13]	[0.18]
Female	0.08	-0.24	0.15
	[0.85]	[0.56]	[0.12]
Retaker	0.02	-0.11	0.23
	[0.96]	[0.47]	[0.45]
CAT grade	0.16	0.28	0.01
	[0.20]	[0.16]	[0.89]
Scholarship	-0.02	-0.04	-0.13
	[0.84]	[0.93]	[0.41]
Nonspanish	-0.65	-0.12	-0.05
	[0.26]	[0.47]	[0.92]
HS track	0.16	-0.13	-0.02
	[0.74]	[0.42]	[0.95]
Constant	-1.16	-2.07	-0.21
	[0.32]	[0.19]	[0.74]
N	147	141	322
Number of classes	4	2	5
R^2	0.06	0.09	0.08

Note: Dependent variable is belief change and takes values between -4 and 4; positive values indicate a change away from the misconception. In brackets we report p-values obtained using wild bootstrap restricted (with Rademacher weights for the auxiliary random variable). We use wild bootstrap and not wild cluster bootstrap, like in the joint regression, because by field the small number of clusters is an issue (see MacKinnon and Webb, 2018). Significance levels: *10%, **5%, ***1%.

Emotions

Using the data from the laboratory we study whether the observed change in beliefs is correlated with emotions, in particular, with how strongly individuals feel about the housing issue and about income distribution.

We first analyze whether the initial belief on rent controls is correlated with the intensity of anger felt by participants about high rents at the beginning of the experiment. The statement is "I feel outraged at the high price of rental housing in metropolitan areas" (Table A.1). We focus on the intensity of emotion felt at the beginning because it is exogenous to the treatment, which could potentially affect the emotion in subsequent surveys. At the beginning of the first session in the laboratory 77% of participants agree with the statement, while 12.8% disagree. Controlling for the same variables as in Table 2 and the CRT, we regress initial beliefs on the intensity of outrage separately for each condition. We find that individuals who feel more outraged tend to hold the misconception. The estimated coefficients are, respectively, for control, individual and group conditions, 0.26 (significant at 5%), 0.23, and 0.40 (significant at 1%).[23]

Given this positive correlation we then explore whether the treatment effect varies with the intensity of outrage at the beginning of the experiment. We estimate equation (1) including an interaction term between the treatment and the intensity of outrage, the same control variables as in Table 2, and the CRT.[24] Figure D.1 shows how the marginal effect of the individual and group treatments varies with the intensity of outrage (table with results is available upon request). In the individual treatment the refutation text appears

[23] Full results are available upon request.

[24] We cannot estimate an equivalent specification for the field data because the statement about the emotion was not included in the survey for the control cohort.

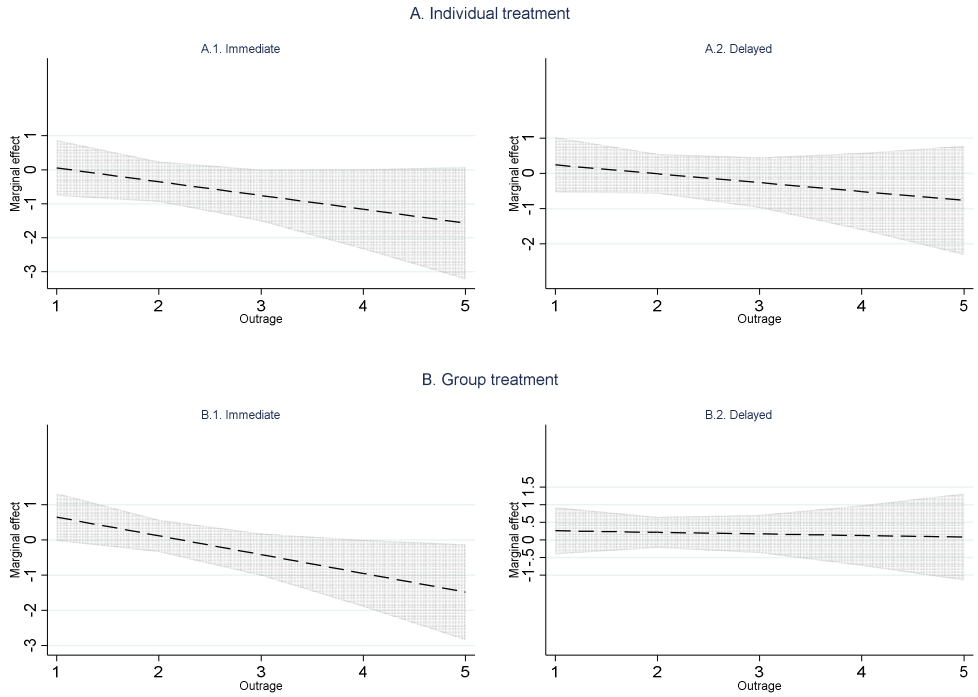
to be more effective for participants who feel more outraged by the high rents as shown by the positive marginal effects, although they are measured imprecisely. In contrast, the marginal effect becomes negative as participants feel less outraged (Panel A.1). In the group treatment this pattern is stronger (Panel B.1). This pattern however fades away for the delayed response, indicating that a few weeks after the treatment, the correlation between the outrage emotion and the change in beliefs is weaker.

We also explore whether the treatment effect varies with participants' initial opinions on income differences and on equality of opportunities. At the beginning of the first session, 92.5% of participants believe that income differences in Spain are too high; this percentage is close to the response of the Spanish sample in Eurobarometer 2017.[25] At the same time, 76.2% of participants disagree that they have equal opportunities for getting ahead in life like everyone else. In the Eurobarometer the percentage is lower, 32%, which possibly reflects that our sample includes only young people. For each of these two statements on fairness perceptions, we estimate equation (1) including its interaction with the treatment. Estimation results do not show significant marginal effects over the range of values of the responses to the statements (for the sake of brevity, results are not shown here but are available upon request). In addition, in the field, we examine the correlation between the change in beliefs of compliers and the set of statements regarding other emotions and attitudes included in the initial survey (see Table A.2). In a regression including all these statements and control variables we do not find a significant correlation with the change in beliefs in any of the cases, except for the opinion on tax

[25] Special Eurobarometer 471, Fairness, inequality and intergenerational mobility, December 2017.

evasion. Participants who support fighting against tax evasion are more likely to move away from the misconception.

Figure D.1. Marginal treatment effects at outrage values



Note: Outrage statement: “I feel outraged at the high price of rental housing in metropolitan areas”. Values: 1 (totally agree), 2 (agree), 3 (do not know), 4 (disagree), 5 (totally disagree).