



Civility and trust in social media

Angelo Antoci^a, Laura Bonelli^{b,c,d}, Fabio Paglieri^d, Tommaso Reggiani^{e,f},
Fabio Sabatini^{f,g,*}

^a University of Sassari, Department of Economics and Business, Sassari, Italy

^b Sapienza University of Rome, Department of Psychology, Rome, Italy

^c USI University, Institute of Argumentation, Linguistics and Semiotics, Lugano, Switzerland

^d National Research Council, Institute of Cognitive Sciences and Technology, Rome, Italy

^e Masaryk University, Department of Public Economics, Brno, Czech Republic

^f IZA, Bonn, Germany

^g Sapienza University of Rome, Department of Economics and Law, Rome, Italy



ARTICLE INFO

Article history:

Received 20 February 2018

Revised 28 January 2019

Accepted 21 February 2019

JEL classification:

C91

D9

D91

Z1

Keywords:

Social media

Facebook

Online incivility

Trust

Social networks

Cooperation

Trust game

ABSTRACT

Social media have been credited with the potential of reinvigorating trust by offering new opportunities for social and political participation. This view has been recently challenged by the rising phenomenon of online incivility, which has made the environment of social networking sites hostile to many users. We conduct a novel experiment in a Facebook setting to study how the effect of social media on trust varies depending on the civility or incivility of online interaction. We find that participants exposed to civil Facebook interaction are significantly more trusting. In contrast, when the use of Facebook is accompanied by the experience of online incivility, no significant changes occur in users' behavior. These results are robust to alternative configurations of the treatments.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

There is wide consensus that trust and trustworthiness are fundamental assets for economic development. In the words of Arrow (1972): "Virtually every commercial transaction has within itself an element of trust, certainly any transaction conducted over a period of time. It can be plausibly argued that much of the economic backwardness in the world can be explained by the lack of mutual confidence" (p. 357). Yet, we still have limited knowledge of where trust comes from. The social capital literature posits that trust crucially depends, whether positively or negatively, on social interactions (e.g. Banfield, 1958; Fukuyama, 1995; Putnam et al., 1993; Knack and Keefer, 1997). As documented by the Pew Research Center

* Corresponding author.

E-mail addresses: antoci@uniss.it (A. Antoci), laura.bonelli@istc.cnr.it (L. Bonelli), fabio.paglieri@istc.cnr.it (F. Paglieri), tommaso.reggiani@econ.muni.cz (T. Reggiani), fabio.sabatini@uniroma1.it (F. Sabatini).

(PRC), social interactions now increasingly take place online, especially in the context of social networking sites (SNS)¹ (Anderson and Perrin, 2017; Antoci et al., 2011; Conrads and Reggiani, 2017; Greenwood et al., 2016).

Facebook studies in the fields of psychology and sociology credited the use of SNS with the potential of reinvigorating trust by offering new opportunities for social interaction, civic engagement and political participation (e.g. Hampton and Wellman, 2003; Ellison et al., 2007; Steinfield et al., 2008; Valenzuela et al., 2009). These works, however, focused on specific local communities and did not account for the rising phenomenon of online incivility that is progressively spreading across social media (Duggan, 2017).

Online incivility is a manner of harassing behavior that can range from aggressive commenting in threads, incensed discussion and rude critiques, to outrageous claims, hate speech, and more severe forms of harassment such as purposeful embarrassment and physical treats (Anderson et al., 2014; Antoci et al., 2016). Overall the evidence available so far suggests that incivility is the prevailing mode of interaction in the presentation of opposing views and opinions in the online discussions of controversial topics (Rainie et al., 2012; Duggan, 2014; 2017; Antoci et al., 2016). This may seem in contrast with the often friendly vibe social networks strive to promote: taking Facebook as an example, it is no happenstance that disliking a post is not allowed by the platform, whereas a vast array of positive emoticons is made available to users; similarly, the technology constantly encourages us to celebrate each other achievements, reminds us of our friends' birthdays, automatically creates celebratory videos of "online friendships", and so on. While these efforts do create an abundance of cheerful and amicable posts among friends, they do not prevent things to turn nasty, as soon as a controversial topic enters the discussion, especially in the interaction with strangers. When users are faced with sensitive and divisive issues (e.g., political views, cultural diversity, gender discrimination, and more generally any belief on matters of great importance, including health, education, and ethics), online discussions show a systematic tendency to take a turn for the worse, in two possible ways: either with an escalation of inflammatory comments against something or someone who is not part of the discussion, when most participants share the same general standpoint (a sort of online mob behavior, that leads to rabble-rousing); or with a rapid and aggressive back-and-forth of mutual accusations and insults among people with different views, when no consensus is apparent among the discussants (Siegel et al., 1986; Rost et al., 2016).

Broadly speaking, two main factors make toxic discussion of controversial topics the norm rather than the exception, on social media: on the one hand, people are acutely aware that their debate is taking place in front of a (potentially very large) audience, and this creates huge reputational concerns and a tendency never to publicly reconsider one's positions, once stated; on the other hand, most social networks are not designed to encourage argumentation, but rather to promote the straightforward assertion of one's position with little or no justification (Facebook is a good example of that, as argued elsewhere; see Kirschner, 2015; Paglieri, 2017).

The prevalence of incivility in the online discussion of controversial topics is an established finding in the literature on computer-mediated communication. Early psychological experiments suggest that, when it comes to the presentation of opposing opinions, there is a fundamental difference between face-to-face and online interaction. In SNS-mediated interactions, interlocutors are basically 'invisible' and their feelings can hardly be perceived. Kiesler et al. (1984) observed that "The overall weakening of self or normative regulation might be similar to what happens when people become less self-aware and submerged in a group, that is, deindividuated" (p. 1126). Deindividuation has in turn been found to be conducive to disinhibition and lack of restraint (Diener, 1979). In computer-mediated interaction, "Communicators must imagine their audience, for at a terminal it almost seems as though the computer itself is the audience. Messages are depersonalized, inviting stronger or more uninhibited text and more assertiveness in return" (Kiesler et al., 1984, 1126). As a result, while in physical interactions people are generally reluctant to offend a person who expresses an opposing view, SNS users are generally more inclined to behave aggressively in online conversations. Pioneering experiments comparing face-to-face and online interaction found that people in online groups were more aggressive, more responsive to immediate textual cues, more impulsive and assertive, and less bound by societal norms of how to behave in public than they were in face-to-face groups. Overall, online discussions were characterized by reduced self-awareness, deindividuation, disinhibition, and lack of restraint, resulting in a propensity for uncivil behavior that had no equal in face-to-face conversations (Kiesler et al., 1984; Siegel et al., 1986; Sproull and Kiesler, 1986; Lea et al., 1992). The Facebook "Pages" and the Twitter accounts of actors of public interest such as political parties, magazines, and celebrities provide a typical setting for the discussion of sensitive topics resulting in online incivility in that they allow users to randomly read comments written by strangers who subscribed to the same feed (i.e. who "liked" the same page or "followed" the same account).² Descriptive evidence supports the view that incivility is the prevailing mode of interaction in the online discussion of divisive issues (Rainie et al., 2012; Duggan, 2014; 2017). A nationally representative PRC survey found that 41% of U.S. citizens have been personally subjected to harassing or abusive behavior on social media. 66% have witnessed others being targeted with online incivility. Anonymity is often blamed as a key enabler of cruelty and abuse in online discussions. 89% of U.S. Internet users say anonymity allows

¹ Hereafter we will use the terms social media, social networking sites and online social networks as synonyms for the sake of brevity.

² Even if subscribers to a specific feed of public interest (e.g. the Facebook page of a celebrity or of a newsmagazine) may have specific interests in common, they are likely to be heterogeneous in terms of personal traits, preferences, opinions, and modes of interaction (Barberà et al., 2015; Barberà and Rivero, 2015; Boxell et al., 2017). In this context, incivility is widespread and usually takes the form of aggressive or outrageous commenting in threads (Rainie et al., 2012; Duggan and Smith, 2016).

people to be cruel or harass one another and 83% think it makes it hard to trust what people share.³ Political views and gender are the most common reasons for aggressive and/or offensive behavior. Survey data indicate that a large majority of Americans view the tone of online political discussions as uniquely angry and uncivil, believe that social media are places where people say things while discussing politics that they would never say in person and describe their online interactions with those they disagree with as stressful and frustrating (Duggan, 2017; Rainie, 2018). The direct or indirect experience of incivility seems to leave an impact that weakens the positive potential of social media. According to a recent report, large shares of users say that “social media interactions related to politics are less respectful, less conclusive, less civil and less informative than offline interactions” (Rainie, 2018). Around one-quarter of Americans (27%) say they have decided not to post something online after witnessing the harassment of others, while more than one-in-ten (13%) say they have stopped using a social network after witnessing other users engage in uncivil behaviors (Duggan, 2017).⁴ Even if women and men have been logging on in equal numbers since 2000, the vilest and most harsh communications disproportionately target women, which are more likely to report being stalked and harassed on social media (Anderson, 2018; Duggan, 2017).

In a keynote speech at the New Media and Society conference in Toronto on July, 2017, Lee Rainie, director of internet research at PRC, highlighted how the pervasiveness of online incivility is making the social media ecosystem increasingly hostile to users. According to Rainie, PRC statistics suggest that the specific characteristics of SNS-mediated interaction incentivize disinhibition, aggressiveness, outrage, and extremism, resulting in a hyper-partisan climate that has invaded personal interactions and in the spreading of confusion and mistrust (Rainie, 2017).

There are reasons to suspect that the infringement of social norms for the polite expression of opinions can affect the trust and trustworthiness of SNS users. In an experimental study about televised incivility, Mutz and Reeves (2005) conclude that witnessing a civil interaction about a topic of public interest might strengthen the belief that most people respect the rules and, therefore, can be trusted. In contrast, the experience of incivility could erode trust in others. This argument is consistent with earlier findings of a negative correlation between television watching and social trust in survey data (Putnam, 2000; Glaeser et al., 2002). As television, a unidirectional mass medium, was found to significantly affect trust, it stands to reason that social media, which allow for interactive communication, might induce an even more powerful effect that could neutralize the supposedly positive impact of social media and prime distrustful attitudes in their users.

Given the penetration of social media and the importance of trust in the economic activity, we believe that the role of SNS-mediated social interaction should be put under scrutiny by economic research. Studying the relationship between SNS-mediated interaction and trust, however, poses several methodological problems. The use of survey data entails endogenous sample selection and endogenous treatment assignment that may make it impossible to ascertain causality. Active membership of SNS, trust and trustworthiness may be codetermined by unobservable personality traits. Reverse causality also is at stake, as more trusting people may have a stronger propensity for online socialization. In addition, the available surveys do not reveal the kind of interaction users actually experience in SNS. Economic studies relying on survey data basically analyze how certain behaviors correlate to binary variables measuring broadband access (Bauernschuster et al., 2014; Falck et al., 2014; Campante et al., 2017) and Internet or SNS use (Pénard et al., 2013; Castellacci and Schwabe, 2017; Sabatini and Sarracino, 2017; Castellacci and Tveito, 2018).

To overcome selection issues, endogeneity problems and the lack of information about users' activities in SNS, we design the first randomized controlled experiment aimed at studying the effect of online incivility on trust. We compare the trust and trustworthiness of three different samples of participants randomly involved in two antithetic Facebook-mediated interaction (civil vs. uncivil) or in an interaction-neutral environment. One group is exposed to four, authentic, threads of uncivil discussion. According to the survey data just reviewed, this treatment is likely to approximate the predominant style of interaction in the online discussion of divisive issues. Another group is exposed to the same threads in which the uncivil discussion has been replaced with polite interactions by experimenters. As a robustness check, a subsample of each group is randomly exposed to an alternative configuration of the treatments in which the same Facebook-mediated interactions, whether uncivil or polite, occur with anonymous users. The third group is used as a control condition: participants are exposed to the same thematics used in the other treatments, but in the form of short news excerpts and without any kind of social interaction. To assess trust and trustworthiness, we use a slightly modified version of the trust game introduced by Berg et al. (1995).

The results of the experiment indicate that the improvement in the tone of discussion (from incivility and neutrality up to civility) leads to higher trust. As the interaction between users deviates from the “uncivil status quo” and results in a civil discussion, it significantly raises participants' trust with respect to both the uncivil treatment and the control condition, whereas it has no effect on trustworthiness. In contrast, if Facebook use is associated with the experience of online incivility in line with the status quo, we observe no significant change in participants' trust and trustworthiness with respect to the control condition. These results are compatible with the view that participants feel no difference between the priming of uncivil discussions possibly associated with previous Facebook experience and the incivility-based treatment they were targeted with in the laboratory; therefore they do not manifest any significant reaction in terms of trust and trustworthiness. However, when participants are primed with civil interaction, a significant increase in their trust emerges. Results are robust

³ Views on anonymity, however, are conflicting, as 85% of Americans report it favors discussion of sensitive topics and 77% think it makes people feel more private and secure.

⁴ These figures are consistent with evolutionary models suggesting that the dissatisfaction with online interaction prompted by incivility may lead to a drastic decline in the use of social media (Antoci et al., 2015; Antoci and Sabatini, 2018).

to a different configuration of the treatments in which participants are exposed to Facebook threads among anonymous users, whose identity has been camouflaged by experimenters: thus anonymity does not seem to mediate the effects of online civility and incivility on trust attitudes.

Our paper bridges three strands of literature. The first studies how broadband access and the use of Facebook impact various dimensions of social capital by exploiting survey data, whether collected in representative samples (e.g. Bauernschuster et al., 2014; Falck et al., 2014; Geraci et al., 2018) or in limited groups of college students (e.g. Ellison et al., 2007; Steinfield et al., 2008) from a cross-disciplinary perspective. The second includes economic studies more broadly investigating the roots of trust using experimental (e.g. Glaeser et al., 2000; Carpenter et al., 2004; Guerra and Zizzo, 2004; Fehr et al., 2005; Ben-Ber and Putterman, 2009; Bigoni et al., 2016; Burdin et al., 2018; Gupta et al., 2018) or survey-based methods (e.g. Alesina and La Ferrara, 2002; Yamamura, 2009; Berggren and Bjørnskov, 2011; Nunn and Wantchekon, 2011; Accetturo et al., 2014; Ljunge, 2014). The third comprises earlier psychological studies examining how the specific features of online communication reduce self-regulation and self-awareness, resulting in stronger disinhibition and aggressiveness in respect to face-to-face interaction (e.g. Diener, 1979; Kiesler et al., 1984; Siegel et al., 1986; Sproull and Kiesler, 1986).

We contribute to the first two strands by providing the first controlled experiment comparing the effects of civil and uncivil online interaction on trust. Our results also adds to psychological studies by showing one possible outcome of the violation of the social norms for the polite expression of opposing views.

The rest of the paper is organized as follows. Section 2 describes our experimental design and procedures. Results are presented in Section 3. Section 4 discusses our findings and their possible policy implications. Section 5 concludes.

2. Experimental design

2.1. The trust game

To measure trust and trustworthiness, we use a slightly modified version of the trust game introduced by Berg et al. (1995). The trustor has an initial endowment of $X = 10$ experimental currency units (ECU), and can choose to send an amount x to player B, the trustee, with $0 \leq x \leq X$. Player B receives $3x$ and chooses an amount y to return to player A, with $0 \leq y \leq 3x$. The final payoff for player A is $X - x + y$ and for player B is $3x - y$. In this paper we adopt a “strategy method” version of the trust game, in which the trustee decides which amount to return for every possible transfer received from the trustor. The amount x is used as a measure for the trustor’s trust in an anonymous partner. The returned amount y , in relation to the received amount $3x$, is considered as an indicator for a subject’s trustworthiness. The unique subgame perfect Nash equilibrium (SPNE) is for Player B to return no money (zero trustworthiness) and, thus, for Player A to send none (zero trust). This equilibrium is socially inefficient in that it is Pareto-dominated by other outcomes. Data from experiments, however, are regularly inconsistent with the predictions of zero trust and trustworthiness.

2.2. Experimental manipulations

In order to circumvent the selection and endogeneity problems that arise in the analysis of naturally occurring data, we employed a framed laboratory experiment (Harrison and List, 2004) in which we targeted two randomly selected groups of participants with two treatments of civility and incivility in a Facebook environment. A third group received neutral treatment characterized by the absence of any salient civil or uncivil feature. Following this, we asked them to play a trust game involving an anonymous stranger. Adopting a between-subjects design, we compared the behavior of individuals exposed to civility with that of the groups that experienced the incivility priming, which represented the social network’s status quo, and the neutral treatment.

In our study, the civility and incivility primings were then administered and framed in the context of Facebook Pages. The incivility treatment was as follows. We asked a randomly selected group of participants to carefully read four *authentic* threads drawn from the Facebook pages of the news websites *Today.it*, *Leggo*, and *Ansa* and of the popular pro-conspiracy theories blog *Su la testa* (literally “Raise your head”). The first thread was about the spreading of a new form of female body decoration called “vajazzling”. The thread was made of a series of sexist comments characterized by prejudice and stereotyping against women, mostly written by women. The second thread reported the decision of some parents to withdraw their children from school to protest against the enrollment of two six years old Roma students.⁵ Comments expressed racist beliefs, ethnocentrism, and a strong hostility against minorities. In a third thread, an op-ed article by a famous conspiracy theorist claimed that the Human Immunodeficiency Virus (HIV) did not exist and that the Acquired Immune Deficiency Syndrome (Aids) had no viral origins. The thread was filled with comments expressing hate and distrust towards researchers, considered responsible of manipulating data to favor pharmaceutical companies. The last thread regarded an accident occurred to a teenager while trespassing a fence nearby the Colosseum in Rome. Users’ comments manifested delight for the incident and rudely blamed the victim for her imprudence. The four threads were collected and republished as they were originally by a specialized Facebook page monitoring hate speech on the social network.

⁵ Roma are a traditionally nomadic ethnic group, living mostly in Europe and the Americas and originating from the northern regions of the Indian subcontinent. A 2016 Pew Research poll found that Italians, in particular, hold strong anti-Roma views, with 82% of Italians expressing negative opinions about Roma.

Table 1
Summary of the treatments.

Treatment	Obs.	Configuration share: Anonymous/Public	Gender share: Male/Female
Incivility	128	0.50/0.50	0.48/0.52
Civility	126	0.50/0.50	0.57/0.43
Baseline	158	—	0.49/0.51

In light of the public debate on the tone of political conversation in social media, we operationalize the main experimental variations under two alternative configurations to improve our understanding of the roles of anonymity and recognizability of users. Anonymity is often blamed as an enabler of incivility in online discussions and is considered by some experts as a source of distrust among Internet users (Rainie et al., 2017). Attempts have been made to ban anonymity from online interaction and to require sites to have online commenters identify themselves. In Italy, for example, a law bill was presented to ban the anonymous use of the Internet on October 2017⁶. From the point of view of civil rights and privacy protection, however, the emphasis that is given to anonymity and recognizability in the public debate is misleading and conducive to potentially harmful side effects. The right to anonymity is crucial in the political and social discourse, as it allows the free expression of ideas and protects unpopular individuals from retaliation. The Supreme Court of the United States has ruled repeatedly that the right to anonymous free speech is protected by the First Amendment. Anonymity has long been considered the norm, rather than the exception, in online networks until the advent of Facebook, with most SNS still allowing users to remain anonymous (e.g. Twitter and Instagram). PRC statistics suggest that users see anonymity as a standard tool to protect their privacy and freedom of expression (Madden et al., 2013; Rainie et al., 2013). In addition, there is evidence that recognizable users are more aggressive than anonymous ones (Rost et al., 2016). To test whether the trust and trustworthiness of participants change depending on the recognizability of commenters, we prepared two alternative configurations of the treatments in which the identity of Facebook users was (i) camouflaged by replacing their actual names and profile pictures with clearly counterfeit ones or (ii) revealed⁷. We call the two configurations “anonymous” and “public” respectively for the sake of brevity.

After reading the threads, participants were required to: 1) briefly summarize the prevailing opinion in the thread; 2) state the extent to which they agree with that prevailing opinion on a 5-point Likert scale ranging from 0 (“not at all”) to 4 (“totally”); 3) assess the aggressiveness and/or offensiveness of the thread on the same five point-scale; 4) select the most aggressive and/or offensive comment in the thread; 5) write a short (max 300 characters) comment in reply to the thread.

To prime with online civility another randomly selected group, uncivil comments were replaced with polite ones created by experimenters. Finally, subjects exposed to the control condition were required to perform the same 5 tasks after having read neutral presentation of the same topics (vajazzling, Roma students, HIV complot theories, teenager injury after trespassing) in the form of short news excerpts, without any further comments and thus without any social interaction, civil or uncivil. Given that there were no comments in the pieces of news participants were required to read in the control condition, they were asked to assess the extent to which they perceived the text as aggressive and/or offensive on a 5-points scale, and select the most aggressive/offensive sentence in the text.⁸

A summary of the treatments is described in Table 1. The main test materials, in Italian, along with their English translation, are reported in the Appendix A.

Following this, subjects participated in a trust game in which we used the strategy method to maximize the data acquired from each participant. We then compared the trust and trustworthiness of participants primed with different types of Facebook interaction, i.e. civil vs. uncivil and anonymous vs. non-anonymous.

The trust game requires players to take two decisions. Trustors must choose the amount to send and trustees choose how much to return. The first choice is affected by one's expectations on the other player (which is also the reason why it serves to measure trust), whereas the second choice is based on one's own willingness to reciprocate the other's trusting attitude, with no strategic consideration on what the other will do, since the trustee returns the money (or not) only after having received whatever the trustor decided to transfer. Moreover, the trustor has a strategic incentive to transfer some money even in one-shot versions of the game, like the one used here, because the more it is transferred, the higher the opportunity for further gain if the trustee proves to be trustworthy. The trustee, in contrast, is strategically incentivized to return something to the trustor only in repeated versions of the game, in which such behavior may allow to establish one's reputation as a trustworthy agent, thus eliciting higher transfers of money from the trustor. But in the one-shot version of the game used here the trustee has no strategic incentive to reciprocate, and social expectations on the other player are irrelevant for the trustee's choice. Based on these asymmetries in strategic incentives, as well as on the assumption that the status quo in online social interaction is incivility rather than civility, we formulate the following predictions on the effects of our experimental manipulations:

⁶ The bill can be accessed at the url (only in Italian): http://www.camera.it/_dati/leg17/lavori/stampati/pdf/17PDL0055510.pdf.

⁷ See the online Appendix A for further details.

⁸ Full instructions are reported in the online Appendix B.

Hypothesis 1. Senders (trustors) primed with civility send a larger amount on average than those in both the incivility treatment and the control condition. The improvement in the tone of the online discussion (from incivility and neutrality up to civility) leads to higher trust.

Hypothesis 2. Senders (trustors) primed with incivility send the same amount as those in the control condition, thus confirming that incivility is perceived as the status quo in online interaction on controversial topics.

Hypothesis 3. Neither civility nor incivility primes affect the behavior of receivers (trustees), with respect to the control condition, because these primes target social expectations on the counterpart, yet as discussed in the one-shot version of the trust game trustees have no strategic reason to care about these social expectations: hence we predict our primes will modulate only the trustors' behavior.

As experienced Internet users are accustomed to anonymity, and generally perceive it as a standard choice to protect privacy and freedom of expression (Madden et al., 2013; Rainie et al., 2013), we expect to observe no systematic consequences from allowing commenters to be recognizable or anonymous in our treatments.

2.3. Procedures

Experimental subjects were recruited via ORSEE (Greiner, 2015) among students who had previously expressed a general willingness to take part in experiments. All sessions were conducted at the Bologna Laboratory for Experiments in Social Science (BLESS) of the University of Bologna (Italy). The experiment was programmed and implemented using the software z-Tree (Fischbacher, 2007). The ECU to € conversion rate was 1:1. Experimental sessions lasted 55 min and the average individual payoff was about 18 € (including a 5 € show-up fee).

Participants were not informed of the goal of the priming tasks. After showing up at pre-scheduled session times, subject were seated at individual cubicles equipped with their own computers. Seats were randomly assigned.

A total of 412 subjects participated in the experiment. We ran 13 sessions with 32 participants per session (with the exception of sessions #1 and #12 in which only 30 subjects took part due to no-shows). Treatments were randomized across sessions. The mean age of participants was 24 years, and they were almost equally split between males (51%) and females (49%). Participants were exposed to multiple batteries of survey tasks in order not to put great emphasis on the trust game. This feature aims to prevent experimental demand effect (Zizzo, 2010).

All participants had a remarkable Facebook experience and almost 70% of them declared to be familiar with the action of reporting offensive posts or comments, as measured through the five point-scale answers to the questions: "How familiar you are with the possibility to report offensive contents on Facebook?", 0 meaning "not at all" and 4 meaning "very much".

The first survey task consisted in responding to a simplified "Yamagishi test of trust", a 6-items questionnaire that uses general statements to measure participants' beliefs about honesty and trustworthiness of others, and that has been validated in previous studies (Yamagishi, 1986; Yamagishi and Yamagishi, 1994). In this questionnaire, participants are asked to value their agreement to the following statements on a five point-scale in which 1 means "strongly disagree" and 5 means "strongly agree": "Most people are basically honest", "Most people can be trusted", "Most people are basically good and polite", "Most people trust others", "I am a trusting person", and "Most people respond trustworthily when they are trusted by others". Participants' trust was further assessed through two additional questions. We used binary responses to the question: "Do you think that most people can be trusted, or that you can't be too careful in dealing with people?" developed by Rosenberg (1956) as a measure of social trust.

Trust in other specific entities was then assessed through the so-called "wallet question": "Imagine you lost your wallet with your money, identification or address in your city/area and it was found by someone else. How likely do you think your wallet would be returned to you if it were found by a neighbor/the police/a stranger?" Possible responses were: "Very likely", "Fairly likely", "Not very likely", and "Not likely at all". This measure of trust has been validated by experimental evidence (Knack and Keefer, 1997; Knack, 2001).

The second task was the Big Five 10-items inventory (BFI-10, Rammstedt and John, 2007), a shortened version of the longer Big Five Inventory (BFI, 44 items, John and Srivastava, 1999), which is a self-report instrument designed to assess five main personality traits: openness, conscientiousness, extraversion, agreeableness, and neuroticism, based on Goldberg's (1993) theory of personality.⁹

A third survey evaluated participants' familiarity with SNS and their experience of civility and incivility in social media. Subjects were asked to report the frequency of use of a series of SNS, the amount of time they daily devote to them, the number of their Facebook friends, their familiarity with specific Facebook actions and functions, and the frequency of specific Facebook-mediated social interactions. Finally, they were asked to assess how their previous Facebook experience matched 12 adjectives on a five point-scale, the adjectives being friendly, dangerous, productive, unpredictable, superficial, tender, offensive, interesting, variable, aggressive, banal, and safe.

Participants were then engaged in the priming task described in Section 2.2. Immediately after the priming, we assessed their emotions through a Visual Analogue Scale (VAS), a tool that tries to measure a characteristic or attitude across a continuum of values (Aitken, 1969). In this study, participants received VAS on each of 9 negative emotions (sadness, shame,

⁹ The BFI-10 was developed to provide a reliable measure of these five dimensions for research situations with severe time constraints. Data gathered on English and German respondents showed that the psychometric properties of the BFI-10 are comparable to those of the full-scale BFI, both in size and structure, and similar results were also obtained on the Italian version of the BFI-10 (Guido et al., 2015), which is the one we used here.

Table 2

Descriptive statistics and randomization checks.

variable	scale	Incivility(<i>i</i>)	Baseline (<i>b</i>)	Civility (<i>c</i>)	$\Delta(i - b)$	$\Delta(c - b)$	$\Delta(i - c)$	Anonym. (<i>a</i>)	Public (<i>p</i>)	$\Delta(a - p)$
– (1) Individual characteristics – mean values										
Female	binary	0.516 (0.502)	0.506 (0.502)	0.429 (0.497)	0.010 [0.876]	–0.078 [0.193]	0.087 [0.166]	0.532 (0.501)	0.414 (0.494)	0.118* [0.061]
Age		25.289 (4.499)	24.323 (5.268)	24.746 (2.942)	0.966 [0.101]	0.423 [0.420]	0.543 [0.257]	25.262 (4.113)	24.781 (3.484)	0.481 [0.316]
B5: Openness to experience	1–5	3.586 (0.764)	3.506 (0.705)	3.702 (0.762)	0.080 [0.361]	0.196** [0.025]	–0.116 [0.225]	3.623 (0.784)	3.664 (0.745)	–0.041 [0.669]
B5: Conscientiousness	1–5	3.488 (0.879)	3.462 (0.797)	3.659 (0.766)	0.026 [0.792]	0.197** [0.036]	–0.170 [0.101]	3.587 (0.895)	3.559 (0.760)	0.029 [0.783]
B5: Extroversion	1–5	3.199 (0.971)	3.016 (1.001)	3.155 (0.889)	0.183 [0.119]	0.139 [0.223]	0.044 [0.704]	3.127 (0.951)	3.227 (0.909)	–0.100 [0.394]
B5: Agreeableness	1–5	3.332 (0.761)	3.411 (0.853)	3.460 (0.786)	–0.079 [0.413]	0.049 [0.620]	–0.128 [0.188]	3.452 (0.765)	3.340 (0.784)	0.112 [0.248]
B5: Neuroticism	1–5	3.203 (0.891)	3.066 (0.944)	3.214 (0.975)	0.137 [0.213]	0.148 [0.197]	–0.011 [0.924]	3.278 (0.931)	3.141 (0.931)	0.137 [0.241]
Yamagishi test of trust	1–5	2.827 (0.610)	2.885 (0.652)	2.915 (0.598)	–0.058 [0.440]	0.030 [0.687]	–0.089 [0.244]	2.903 (0.620)	2.839 (0.589)	0.064 [0.393]
– (2) familiarity and experience with social media – mean values										
Number of social media		4.35 (1.89)	4.22 (2.04)	4.05 (1.89)	0.13 [0.58]	–0.17 [0.46]	0.30 [0.20]	4.00 (1.74)	4.40 (2.02)	–0.40* [0.09]
Facebook user	binary	0.961 (0.195)	0.949 (0.220)	0.944 (0.230)	0.012 [0.642]	–0.005 [0.854]	0.016 [0.537]	0.952 (0.214)	0.953 (0.212)	–0.001 [0.978]
Number of Facebook friends		515.12 (439.02)	546.94 (462.47)	510.15 (429.96)	–31.82 [0.55]	–36.79 [0.49]	4.97 [0.93]	475.91 (388.31)	548.83 (472.90)	–72.91 [0.18]
Frequency Facebook use	0–4	3.641 (0.954)	3.525 (1.045)	3.413 (1.053)	0.115 [0.335]	–0.113 [0.369]	0.228* [0.072]	3.524 (1.002)	3.531 (1.019)	–0.007 [0.953]
Positive Facebook experience	0–4	1.89 (0.713)	1.951 (0.805)	1.712 (0.733)	–0.058 [0.523]	–0.24*** [0.010]	0.182** [0.047]	1.847 (0.731)	1.760 (0.724)	0.086 [0.346]
Negative Facebook experience	0–4	2.138 (0.843)	1.964 (1.092)	2.138 (0.972)	0.174 [0.140]	0.173 [0.164]	0.000 [0.997]	2.116 (0.967)	2.159 (0.848)	–0.042 [0.710]
– (3) Manipulation check – mean values										
Priming score	0–4	3.105 (0.673)	1.331 (0.736)	1.165 (0.782)	1.775*** [0.001]	–0.17* [0.067]	1.941*** [0.001]	2.155 (1.277)	2.131 (1.154)	0.024 [0.876]
Visual analogue scales (VAS)	0–14	5.994 (3.796)	2.899 (2.763)	3.025 (3.015)	3.095*** [0.001]	0.126 [0.714]	2.969*** [0.001]	4.378 (3.656)	4.661 (3.817)	–0.283 [0.547]

Note: Standard deviations reported in parenthesis. Balance checks *p*-values according to Chiapello (2018) reported in squared brackets.

anger, guilt, embarrassment, sorrow, pain, disgust, fear), to assess the emotional impact of the treatments. The final task was the modified trust game. Players of both roles (A and B) were initially endowed with 10 ECU (ECU to € exchange rate, 1:1) to avoid inequity concerns. The final amount paid to each player depended on their strategic choices in the trust game.

Subjects were randomized into the role of either the trustor or the trustee and were anonymously matched with a unique, unknown partner for a one-shot interaction. After the structure and the payoff of the game were explained to them, participants were not allowed to proceed in the experiment until they correctly answered three comprehension questions on the game. Then each trustor decided how much to transfer to the other player (from 0 to 10 ECU), while each trustee indicated how much s/he would be willing to return for each possible amount transferred by the trustor and multiplied by the experimenter (strategy method), from 3 to 30 ECU. At the end of the experiment participants were immediately paid in cash based on the actual outcome of the interaction, in addition to the show-up fee.

3. Results

In this section we first present descriptive statistics describing our sample along several dimensions and report about their balancedness across experimental groups. We also report evidence about priming manipulation checks in Section 3.1. We then document how the exposure to online civility affects participants' trust in Section 3.2. Section 3.3 presents results regarding trustworthiness.

3.1. Descriptives and randomization/manipulation checks

Table 2 reports the mean values as well as randomization checks (*p*-values according to Chiapello, 2018) of several individual characteristics – obtained from the subjects' answers to the initial questionnaire. The first section of Table (1) displays general individual demographics such as gender, age, nationality as well as psychological traits (Yamagishi test of trust, Big Five). The second block of Table (2) focuses on individuals' familiarity with SNS and their experience of civility and incivility in the social network. At the bottom of Table (3), it is reported a third section in which the effectiveness of the two different priming conditions, incivility vs civilly, is documented. In general, the overall sample is balanced across treatments

Table 3
Average levels of Trust, by treatments.

	Incivility	Baseline	Civility	Jonckheere – Terpstra test
Average Trust level	4.59 (3.41) [64]	4.44 (3.23) [79]	5.52 (3.20) [63]	$p\text{-value}=0.05$

Standard deviations reported in parenthesis.

Number of independent observations in squared brackets.

(uncivil, neutral, civil) and configurations (anonymous, public). For the three main treatments, minor unbalances can be detected along two dimensions of the Big Five inventory (openness and conscientiousness) and in two variables associated to Facebook use (use frequency and positive experience).¹⁰

Almost the totality of the sample uses Facebook. On average, participants use approximately four social networking sites. The frequency of Facebook use is measured on a 5 points scale, 0 meaning never and 4 meaning on regular daily basis. Most of the sample uses Facebook on regular daily basis.

The positiveness of Facebook experience is computed as the average score given to the adjectives “friendly”, “tender” and “productive” to describe the participant’s experience on Facebook, with 0 meaning “never the case” and 4 meaning “always the case”. The negativeness of Facebook experience is computed through the average score given to the adjectives “aggressive”, “offensive” and “banal” to describe the participant’s experience, as measured on the same scale as above. The average score participants gave to the adjectives “aggressive”, “offensive” and “banal” to describe their Facebook experience is equal to 2.07 and systematically higher than the average score given to “friendly”, “tender” and “productive”, which is equal to 1.86. The negative balance between positive and negative experiences proved to be statistically significant not only on average (mean-test, $p\text{-value} < 0.01$) but also at intra-individual level (within-subject) according to a Wilcoxon signed-rank test ($p\text{-value} < 0.01$). Overall, participants seem to share a prevalently negative experience of Facebook along the considered dimensions suggesting that, in our sample, incivility is perceived as the norm of online interaction rather than the exception. This negative balance still holds if we compare the average values of all the negative and positive dimensions accounted for in the assessment of participants’ past Facebook experience given through the questionnaire.

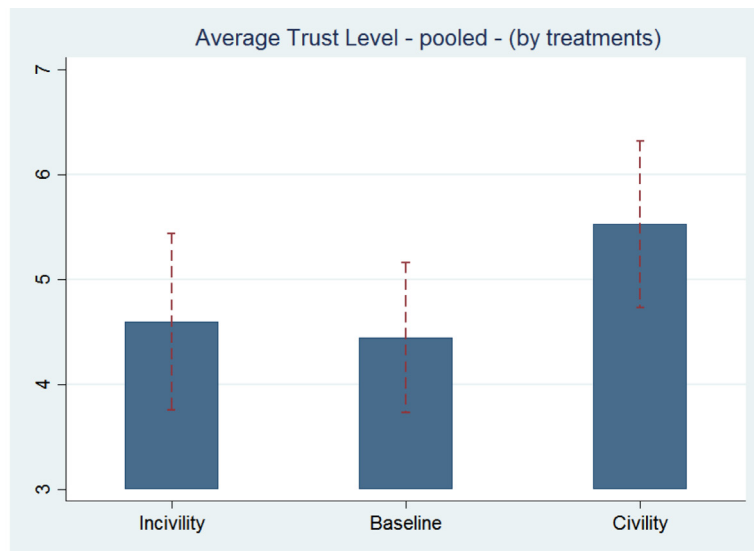
As far it concerns the effectiveness of the main priming manipulations, it is possible to test their impact comparing the average evaluations reported after each Facebook thread. The priming score is computed as the average score given to the aggressiveness and/or offensiveness of comments in the four threads on a 5 points scale, 0 meaning “non aggressive” and 4 meaning “extremely aggressive”. The observed average score under the incivility condition turned out to be quite high and equal to 3.11 while it was only 1.16 under the civility condition. The difference in terms of perceived aggressiveness under the two different conditions is strongly statistically significant at the 1% level (MWU-test, $p\text{-value} < 0.01$).

The reaction to the visual analogue scales (VAS) is computed as the average score given to each of the nine items described above on a 15 points scale ranging from 0 to 14. The score given by participants exposed to the incivility treatment is significantly higher than the one given by those in the civility group (MWU-test, $p\text{-value} < 0.01$). On the other hand, both the VAS score and the priming score given by participants in the control group are very close to the one emerged from the civility treatment (see Table 2: VAS $\Delta(c - b) = 0.13$; Priming score $\Delta(c - b) = -0.17^*$). While it is confirmed the salience of the uncivil and civil treatments, which prompted remarkably different reactions in participants, the similar values of the average responses to the civility treatment and to the baseline condition may cast some doubts on what is actually perceived as the norm in the online discussion of controversial topics. This might be due to the different expectations of participants regarding the materials they were exposed to within the manipulation. On the one hand, the neutral tone of the news excerpt may be perceived as similar to the climate of a civil conversation, thereby prompting similar emotional reactions. The different response in terms of trust, on the other hand, might be related to the different benchmarks of participants regarding the materials they were exposed to within the manipulations. When reading a piece of news, in fact, subjects expect a neutral tone, as much as they expect incivility when they read an online thread on a sensitive topic. The deviation from the expected aggressiveness might be responsible of the positive response in terms of trust. In addition, it must be reminded that the average response to VAS is hardly comparable to the priming score. VAS, in fact, are measured on a remarkably different scale (0–14 vs. 0–4), they refer to participants’ own feelings and emotions instead of to the evaluation of the tone of the comments in the thread and, perhaps most importantly, they encompass a much wider range of dimensions. While the priming score is computed as the average score with which participants critically assess the aggressiveness and/or offensiveness of the comments they read in the four threads, VAS serve to express the reaction that the treatment prompted in them regarding the nine dimensions of sadness, shame, anger, guilt, embarrassment, sorrow, pain, disgust, and fear.

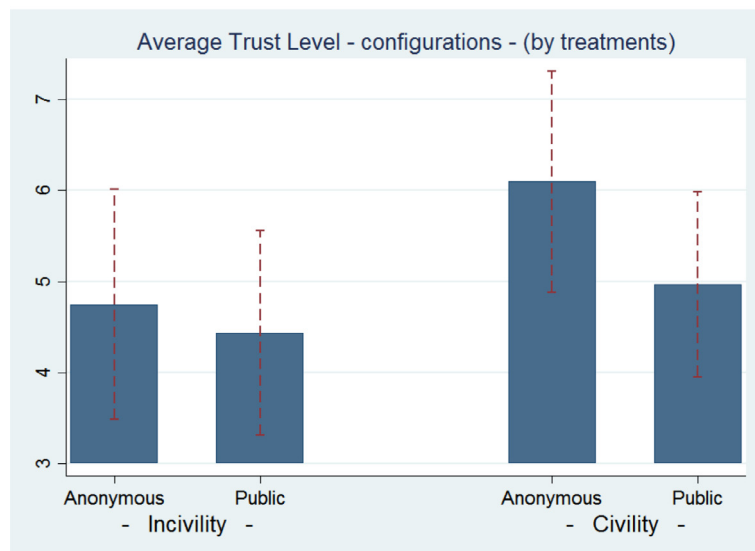
3.2. Trust

Using exogenous manipulations, we find evidence that exposure to online civility can make people more trusting. Visual inspection of Fig. 1. A shows how the average level of trust exerted by trustors under the civility treatment resulted to be

¹⁰ In the parametric analysis we will also control for this specific array of unbalanced characteristics.



(a)



(b)

Fig. 1. Trust. (a) Average levels of trust (solid bars) for each experimental treatment. Error bars based on the standard deviations of the means (dashed bars) for each experimental treatment. (b) Average levels of trust (solid bars) for each experimental configuration, by treatment. Error bars based on the standard deviations of the means (dashed bars) for each experimental condition.

approximately 1 ECU higher (corresponding to a +22% in relative terms) than under the incivility and the neutral treatments. Average values and correspondent standard deviations are reported in Table 3. We perform a Jonckheere–Terpstra test, a non-parametric test designed to detect alternatives of ordered class differences for more than two independent samples. It tests for ordered differences between treatments and hence it requires an ordinal ranking of the test variable. In our case the ordering is given by the intensity of the “positive” tone of interaction: *Incivility* < *Baseline* < *Civility*.¹¹ The non parametric test rejects the null-hypothesis ($p\text{-value} = 0.05$) that the average level of trust is constant across the three treatments, considering ascending ordered alternatives. The effect is driven by the higher average trust level under the civility condition.

¹¹ See for example Bigoni et al. (2015) and Eckel and Füllbrunn (2015). For a detailed description of the test, see Hollander and Wolfe (1999).

Table 4

Average levels of Trust - by configuration/treatment.

	Incivility - Anonymous	Incivility - Public	Mann-Whitney <i>U</i> test
Trust	4.75 (3.65) [32]	4.43 (3.24) [32]	<i>p</i> -value=0.81
	Civility - Anonymous	Civility - Public	Mann-Whitney <i>U</i> test
Trust	6.09 (3.55) [31]	4.97 (2.92) [32]	<i>p</i> -value=0.14

Standard deviations reported in parenthesis.

Number of independent observations in squared brackets.

Table 5

Treatments effects on trust (Tobit).

Outcome: Trust level (0– 10)	(1)	(2)	(3)	(4)	(5)	(6)
Civility	1.992** (0.91)	2.107** (0.91)	2.041** (0.87)	1.958** (0.88)	1.920** (0.89)	
Incivility	0.597 (0.9)	0.612 (0.89)	0.955 (0.85)	1.096 (0.86)		0.862 (1.03)
Public	−0.945 (0.84)	−1.034 (0.84)	−1.24 (0.80)	−1.231 (0.80)	−1.387 (1.04)	−0.865 (1.24)
Female			−2.578*** (0.64)	−2.685*** (0.67)	−2.892*** (0.74)	−2.132** (0.87)
Age (ctr)			−0.058 (0.06)	−0.073 (0.06)	−0.082 (0.07)	−0.067 (0.07)
Yamagishi test of trust			0.645 (0.54)	0.9 (0.62)	1.04 (0.69)	1.02 (0.76)
Frequency Facebook use		0.353 (0.34)	0.350 (0.32)	0.342 (0.32)	0.108 (0.31)	0.16 (0.46)
Positive Facebook experience		−0.151 (0.46)				
Openness to experience		−0.541 (0.45)		−0.268 (0.59)	−0.849 (0.71)	0.376 (0.74)
Conscientiousness		0.218 (0.4)		0.502 (0.39)	0.464 (0.47)	0.205 (0.48)
Extroversion				−0.441 (0.34)	−0.336 (0.38)	−0.331 (0.43)
Agreeableness				−0.194 (0.45)	−0.331 (0.49)	−0.637 (0.55)
Neuroticism				0.168 (0.49)	0.556 (0.57)	0.01 (0.65)
Constant	4.442*** (0.53)	5.701** (2.41)	2.711 (2.02)	2.696 (2.96)	4.379 (3.19)	3.119 (3.97)
N	206	206	206	206	142	143
pseudo – <i>R</i> ²	0.006	0.007	0.026	0.03	0.042	0.016

Significance levels: * *p* < 0.10; ** *p* < 0.05; *** *p* < 0.01. Standard deviations reported in parenthesis.Age variable is centered around the mean ($Age_i - \bar{Age}$).

Fig. 1.B and Table 4 report about the average level of trust when the configurations anonymous vs. public are considered. Both visual inspection of Fig. 1.B and MWU-tests reported in Table 4 show how anonymity and recognizability do not affect systematically the outcomes under civility and incivility.

We complement the non-parametric analysis with a further set of regression analyses.

Table 5 reports Tobit regression analyses. The outcome of interest (*x*) is represented by the level of trust exerted by trustors. The dependent variable ranges between 0 and 10 ECU. Exploiting the between-subjects design, the level of trust is regressed against the main treatment dummies for civility and incivility, while the baseline is included in the constant term.

$$x = \beta_0 + \beta_1 \text{Civility} + \beta_2 \text{Incivility} + [\beta_n \text{controls}] + \varepsilon$$

In column (1) the trust level is simply regressed against the two treatment dummies *Civility* and *Incivility* (*Baseline* is captured in the constant term). In addition to the treatment covariates, we control for the configuration of the treatments in which users were *Public* through a flag variable. The positive effect generated by the exposure to civil discussion detected in the non-parametric analysis is corroborated by regression results. The coefficient associated to the civility treatment dummy is sizable in its magnitude and statistically significant at the 5% level while the coefficient of the incivility treatment does not prove to be statistically different from the baseline. Following the same line, also the experimental configuration featured by publicity turned out to be neutral to the outcome. In column (2) the basic model incorporates the additional individual

Table 6

Average levels of Trustworthiness, by treatments.

	Incivility	Baseline	Civility	Jonckheere – Terpstra test
Trustworthiness	6.16 (3.64) [64]	6.12 (3.54) [79]	5.91 (4.10) [63]	$p\text{-value}=0.80$

Standard deviations reported in parenthesis.

Number of independent observations in squared brackets.

Table 7

Average levels of Trustworthiness, by configuration/treatment.

	Incivility - Anonymous	Incivility - Public	Mann–Whitney U test
Trustworthiness	6.41 (4.05) [32]	5.91 (3.21) [32]	$p\text{-value}=0.42$
Trustworthiness	Civility - Anonymous 6.66 (3.86) [32]	Civility - Public 5.19 (4.25) [31]	Mann–Whitney U test $p\text{-value}=0.18$

Standard deviations reported in parenthesis.

Number of independent observations in squared brackets.

characteristics that proved to be not totally balanced in the randomization checks (Table 2). Their inclusion does not affect the main result delivered by the baseline model (1).

In column (3) the model controls for a conventional array of covariates including gender (equal to 1 if female and 0 if male), age (centered around the mean), the average score of the Yamagishi test, as well as for the self-reported intensity of Facebook use. Also in this case, the coefficient associated to the civility treatment dummy is sizable in its magnitude and statistically significant at the 5% level while the coefficient of the incivility treatment does not prove to be statistically different from the baseline. Following the same line, also the configuration featured by publicity turned out to be neutral to the outcome. Females exhibit a significantly lower level of trust. All the other controls do not show any systematical effect on the individual trust levels.

In column (4), we refine the analysis adding a supplementary battery of individual-specific psychological traits taken from the Big Five test. Accounting for these further personal characteristics does not affect the estimates. Also in this case, the coefficient associated to the civility treatment dummy is stable in its magnitude and statistically significant at the 5% level while the coefficient of the incivility treatment is not distinguishable from the baseline.

Specifications reported in last two columns focus on models in which the civility and incivility variables are included separately. Again, the general result finds additional support. While in column (5) the coefficient of the civility treatment dummy proved to be economically sizable and statistically significant at the 5% level, in column (6) the incivility dummy variable is not statistically significant at any conventional level. Gender's coefficients in columns (3)–(6) are significantly negative, indicating that female participants systematically trust less independently of the tone of the online conversations witnessed within our manipulations. Keeping in mind that this result must be handled with caution as the experiment was not designed ad-hoc to investigate gender issues (Sent and van Staveren, 2018), the sign and significance of the coefficients seem to be consistent with PRC descriptive statistics depicting social media as an hostile environment for women (Duggan, 2017; Anderson, 2018).

3.3. Trustworthiness

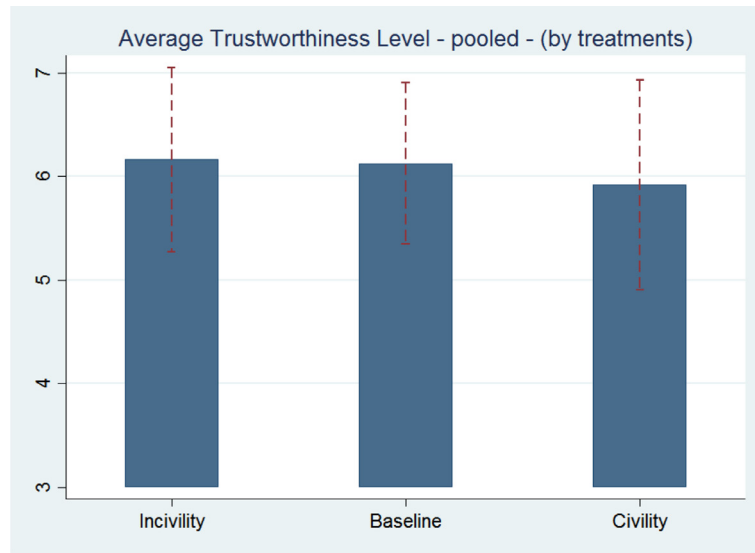
Visual inspection of Fig. 2.A shows how the average level of trustworthiness exerted by trustees resulted to be rather constant across the three experimental treatments.

Average values and correspondent standard deviations are reported in Table 6. As for the trustors' case, we perform a Jonckheere–Terpstra test. The non-parametric test does not reject the null hypothesis ($p\text{-value}=0.80$) that the average level of trust is constant across the three treatments considering ascending ordered alternatives.

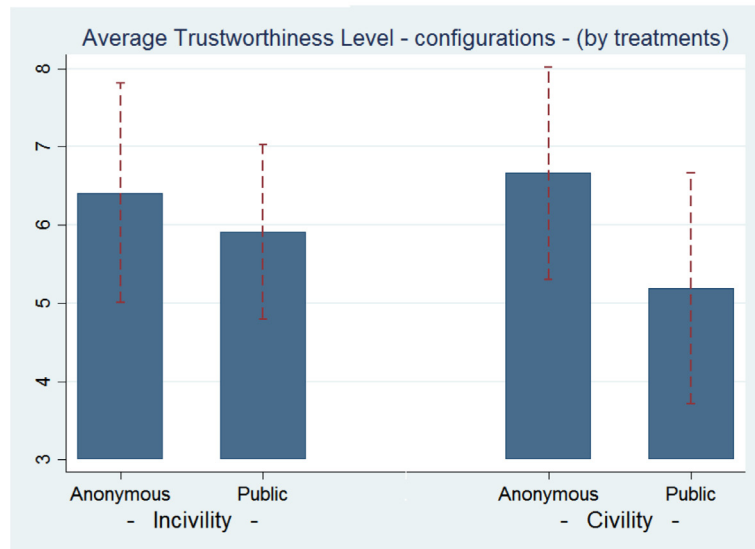
Fig. 2.B and Table 7 report about the average level of trustworthiness when the configurations anonymous vs. public are considered. Both visual inspection of Fig. 2.B and MWU-tests reported in Table 7 show how anonymity and recognizability do not systematically affect trustees' responses under civility and incivility.

We complement the non-parametric analysis with a further set of regressions. Table 8 focuses on trustees' choices and replicates the regression analysis designed to study trustors' decisions (Table 5).

In this section the outcome of interest (\bar{y}) is represented by the level of trustworthiness exerted by trustees. Individual trustworthiness is elicited with the strategy method. The outcome is represented by the individual mean amount returned (\bar{y}) by the trustee to the trustor. The dependent variable is computed at the individual level averaging the multiple trustworthy responses under the different trust profiles. This elaboration allows targeting independent observational units at the individual level. This outcome, by construction, ranges between 0 and 16.5 ECU. As for the analysis of trust level, treatment



(a)



(b)

Fig. 2. Trustworthiness. (a) Average levels of trustworthiness (solid bars) for each experimental treatment. Error bars based on the standard deviations of the means (dashed bars) for each experimental treatment. (b) Average levels of trustworthiness (solid bars) for each experimental configuration, by treatment. Error bars based on the standard deviations of the means (dashed bars) for each experimental condition.

effects are addressed exploiting between-subjects design adopting an analogous Tobit models:

$$\hat{y} = \beta_0 + \beta_1 \text{Civility} + \beta_2 \text{Incivility} + [\beta_n \text{controls}] + \varepsilon$$

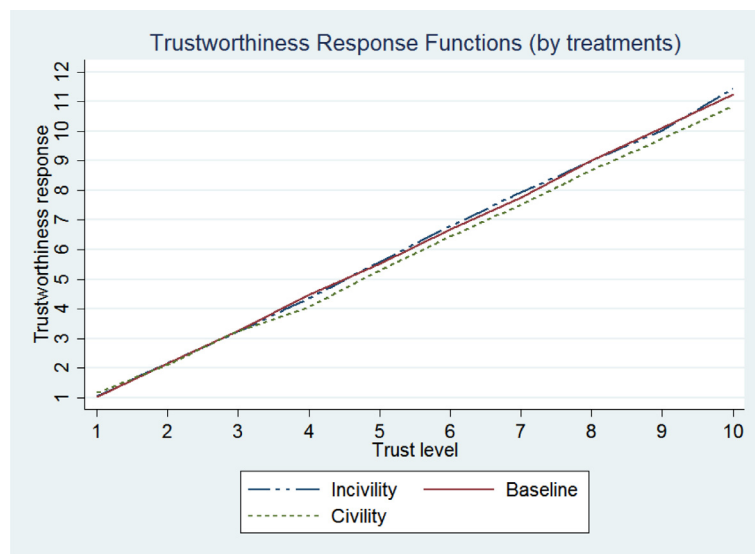
As far as it concerns trustworthiness, there is no evidence for differential treatment effects. In all the different specifications both coefficients for *Civility* and *Incivility* result to be systematically negligible in their magnitudes and never statistically significant at any conventional level. A weakly negative marginally significant effect on trustworthiness it is found to be associated to the public-based configuration. Among the battery of control variables, only Age and the Yamagishi trust score appear to have a mildly systematic positive influence on trustworthiness. In this third block of analysis focusing on trustees' behavior, we study if treatments have an impact on the degree of positive reciprocity, which can be measured (thanks to the "strategy method" elicitation protocol) as the correlation between the amount received from the trustor and the correspondent amount returned by the trustee. By construction, the average level of trustworthiness can be the same

Table 8

Treatment effects on trustworthiness (Tobit).

outcome: Trustworthiness (0 – 16.5)	(1)	(2)	(3)	(4)	(5)	(6)
Civility	0.04 (0.68)	0.048 (0.69)	–0.079 (0.67)	–0.092 (0.67)	0.209 (0.73)	
Incivility	0.51 (0.68)	0.583 (0.67)	0.227 (0.68)	0.274 (0.67)		–0.121 (0.73)
Public	–1.075* (0.63)	–1.184* (0.63)	–0.885 (0.62)	–1.032* (0.62)	–1.794* (0.87)	–0.109 (0.86)
Female			0.227 (0.50)	–0.299 (0.54)	0.026 (0.65)	–0.893 (0.69)
Age (ctr)			0.162** (0.07)	0.141** (0.07)	0.056 (0.09)	0.143* (0.08)
Yamagishi test of trust			0.796** (0.39)	1.012** (0.43)	0.787 (0.50)	0.855* (0.48)
Frequency Facebook use		–0.016 (0.24)	0.017 (0.23)	0.068 (0.23)	0.328 (0.29)	–0.01 (0.26)
Positive Facebook experience		0.116 (0.33)				
Openness to experience		0.564* (0.33)		0.019 (0.48)	0.487 (0.58)	0.031 (0.54)
Conscientiousness		–0.567* (0.31)		–0.415 (0.31)	–0.549 (0.38)	–0.397 (0.36)
Extroversion				0.05 (0.27)	0.601* (0.32)	–0.184 (0.3)
Agreeableness				(0.099)	–0.343 (0.38)	0.341 (0.39)
Neuroticism				0.661* (0.38)	0.321 (0.45)	0.903** (0.44)
Constant	5.954*** (0.4)	5.719*** (1.67)	3.599** (1.48)	2.556 (2.36)	1.089 (2.7)	1.918 (2.70)
N	206	206	206	206	142	143
pseudo – R ²	0.004	0.01	0.015	0.023	0.028	0.032

Significance levels: * $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$. Standard deviations reported in parenthesis.
Age variable is centered around the mean ($Age_i - \bar{Age}$).

**Fig. 3.** Trustees' Response Functions.

across treatments, even if reciprocity patterns differ. In principle, it could be the case that subjects under a given treatment return less in response to low transfers, but relatively more in response to high levels of trust, that is they may have an higher propensity to reciprocate trust. We explore these dimensions in Fig. 3. This graph pictures average response functions of the trustees for the three main treatments.

Visual inspection of the graph suggests that reciprocity patterns are quite stable and consistent across treatments since their plots are essentially overlapped. This visual result is confirmed by the very similar magnitude of the Spearman

correlation coefficients – σ , Trustworthiness|Trust – across treatments (incivility: $\sigma = 0.63$, $p\text{-value} < 0.01$; baseline: $\sigma = 0.65$, $p\text{-value} < 0.01$; civility: $\sigma = 0.57$, $p\text{-value} < 0.01$).

4. Discussion

This study provides the first experimental evidence of a positive effect of online civility on social trust. As predicted, what induced a significant change in people's trust was the exposure to civil online interaction, whereas the opposite condition, i.e. online incivility, seemed to be considered “business as usual” and thus did not produce any effect on trust, with respect to the control treatment. This lack of effect of online incivility is striking, whether it is interpreted as a form of expectation matching (incivility is what people routinely expect online, thus being exposed to it does not change their expectations on others' behaviors) or as a sort of immunization effect (people are used to such a high level of online incivility that the experimental manipulation failed to elicit a response): in either case, incivility seems to be perceived as the norm of online interaction, rather than the exception. This is a rather depressing finding, but also one in full accordance with the survey data provided by Pew Research Center that we reviewed at the onset of this paper (e.g. Duggan and Smith, 2016; Duggan, 2017; Rainie, 2018) and with previous findings from early psychological experiments on the distinctive features of computer-mediated communication respect to offline interactions (e.g. Kiesler et al., 1984; Siegel et al., 1986; Sproull and Kiesler, 1986). As reported in Section 3.1, this result is also consistent with information on the Facebook experience of participants drawn from the questionnaire.

What is much less depressing, and indeed encouraging, is the positive effect on trust of even a brief exposure to online civility: contrary to intuition, according to which a quarrel is much more salient than a polite discussion, a simple lack of aggression in expressing a difference of opinions online acts as a powerful determinant of higher levels of trust towards other people. This result was not only significant, but also robust to a further manipulation in terms of anonymous vs. non-anonymous users: this suggests that what matters for social trust is the nature (civil vs. uncivil) of the online interaction, and not the degree of identity disclosure of the people taking part in the debate. As long as the interaction remains civil, the anonymity of our online interlocutors does not undermine our trust in them.

The second relevant finding was the lack of effect of our manipulation on trustworthiness level. This is a negative result, thus it should be treated with caution. Nonetheless, this finding was in line with our hypotheses, based on the fact that exposure to online civility or incivility affects one's expectations on the behavior of other people, thus modulating trust attitudes: but trustworthiness, or lack thereof, is a feature of our own behavior, thus there is no reason why it should be affected by the civil or uncivil behavior we witness in others. This is especially true with respect to the one-shot trust game we used in this study: whereas the trustor has a strategic incentive to increase the amount transferred, under the assumption that the other is trustworthy, the trustee has no reason to increase the amount returned based on her expectations about the other player, especially since the effect of that choice will happen *ex post*, based on how much the trustor actually transferred in the first place. The strategy method does not mitigate this asymmetry, since it simply forces the trustee to indicate how much would be returned under all possible allocations of resources chosen by the first player: a trustee with positive expectations on the trustor may consider certain allocations more likely to occur than others, but there is no reason to expect an effect on how she will respond to such allocations. In fact, strategic considerations are likely to affect the behavior of the trustee, and thus trustworthiness levels, only when the trust game is repeated among players, due to reputation concerns: only when the trustee expects to interact again with the same trustor (or with trustors who may be aware of the trustee's past behavior), it makes sense to adapt one's choice to the expectations about the other player.

Another worth mentioning finding is the lack of any significant effect generated by allowing the identity of commenters being recognizable. Anonymity is probably considered as a typical feature of Internet-mediated interaction that does not particularly impress users. Revealing the identity of commenters in threads does not affect participants' trust. In other words, what seems to matter is not the identity of users but the tone of the online discussion. Overall, this null result does not provide support to the view that anonymity should not be allowed in social media.

The fact that the subjects were recruited among students of the University of Bologna may raise question as to whether they may share progressive political views and cultural values that are at odds with the topics used in the experiment, leading them to react to the content of the topics, rather than the way in which users' opinions about the topics were expressed (civility vs. incivility). If subjects would have reacted to the topics and not to the civility or incivility of the related comments, then we should have seen no treatment effect, because the topics were the very same both under the civility and incivility treatments. This is not what we observe in our data, where we find a positive difference when we contrast the average trust level under civility against the average trust level under incivility. It is possible, however, that students of the University of Bologna are on average more trusting. If that was true, our experiment would overestimate the trust level of the three groups. We cannot rule this possibility out completely. Still, results obtained in the BLESS Lab have been proved to be fully consistent with results generated in comparable studies conducted in other Italian venues regarding outcomes such as trust and cooperation (see Bigoni et al., 2018 and Casari et al., 2018 for experiments conducted in Bologna and Calabria region), tax morale (see Andrighetto et al., 2016; Guerra and Harrington, 2018 for experiments conducted in Bologna, Milan, and Rome), and reaction to feedback in strategic interaction (Bigoni and Suetens, 2012 and Altavilla et al., 2006 for experiments conducted in Bologna and Siena). In addition, our outcome measures under the control treatment are well in line with the standard result described in the literature for the trust game (see for example the meta-analysis in Johnson and Mislin, 2011). In some specifications of our regressions we have also included individual psychological traits

as further control variables. This set of evidence, looking at the issues from complementary angles, can reassure about the external validity of the data generated in the BLESS Lab.

5. Conclusions

Social media are facing increasing criticism and scrutiny, since recent analyses of key political events (such as the 2016 US presidential election and the Brexit referendum) have suggested a link between the extreme polarization of public opinion and the relatively small number of platforms that monopolize online discourse – most notably, Facebook and Twitter. Public discussion on such platforms has also been shown to create and maintain so called “echo chambers” (Del Vicario et al., 2016; Quattrocioni et al., 2016), thus leading to increased polarization and partisanship. Therefore it is not entirely surprising to find confirmation, in our data, of a rather bleak outlook on public discourse on SNS, where uncivil debate seems to be considered as normal.

However, the striking result of our study is that even minimal exposure to the opposite trend, i.e. civil online interaction, has a significant effect on social trust, even (or maybe especially) when it takes place on the same SNS where incivility is considered to be the standard. This suggests that what is at stake in moderating online discussion may not be simply the prevention of negative phenomena (hate speech, cyberbullying, digital harassment, etc.), but also the achievement of significant social benefits, most notably a measurable increase in social capital. At present, it is unclear whether the positive effect on trust is just a transient result of being exposed to a different norm of behavior (in this case, civility rather than incivility) or a more durable benefit that would persist even after typical online attitudes have shifted towards more open-minded and respectful interaction. But our data indicates that this issue is definitely worth pursuing – and besides, promoting better norms of online discussion would arguably be in and by itself a net gain in social capital.

The take-home message for policy makers is rather straightforward: instead of focusing exclusively on fighting against noxious online behavior, we should also create the preconditions to promote civil discussion on online platforms. While the former goal needs to be pursued via strict regulations, the latter is best fostered by carefully designing (or tweaking) the platforms themselves, bringing a wide variety of competences to bear on such a task: most notably, psychological insight on users' attitudes and profiles, interaction design principles from ergonomics, nudging strategies and incentives planning from economics. In doing so, freedom of expression needs to be preserved for all users, which is why we are skeptical that legal prohibitions alone can act as a panacea against online incivility. Instead, we should strive to effect a paradigm shift in what kind of interactions SNSs afford to their users: today uncivil and shallow confrontation is the norm, with civil debate relegated to being the exception; tomorrow it should be the other way around, thanks to social networking platforms designed to encourage reflective discourse, unbiased assessment of evidence, and open-minded belief change – while still leaving people free to verbally attack each other, should they feel so inclined.

Existing online platforms could and should be used as valuable case studies, leveraging both negative and positive aspects. For instance, whereas violent confrontation seems rather standard in many SNS (which is why we used Facebook threads as our primes), Wikipedia represents a platform where a high number of users, the editors, freely interact in a more constructive fashion and with a higher degree of mutual respect and even solidarity: in fact, so called “edit wars” are relatively rare phenomena on Wikipedia, typically perceived as malicious attempts to thwart the legitimate mission of the platform (Yasseri et al., 2012). Moreover, Wikipedia provides a positive climate for civil interaction not by obsessively policing users' behavior, but rather by building a shared ethos of respect and collaboration in view of a collective mission, that is, to create a universal encyclopedia that anyone can access and modify. Indeed, an appeal to mutual respect and civility is part and parcel of the core of the Wikipedia philosophy, and the fourth of their “Five Pillars”¹² is the recommendation that “Wikipedia's editors should treat each other with respect and civility”, which is further specified as follows: “Respect your fellow Wikipedians, even when you disagree. Apply Wikipedia etiquette, and don't engage in personal attacks. Seek consensus, avoid edit wars, and never disrupt Wikipedia to illustrate a point. Act in good faith, and assume good faith on the part of others. Be open and welcoming to newcomers. Should conflicts arise, discuss them calmly on the appropriate talk pages, follow dispute resolution procedures, and consider that there are 5,524,387 other articles on the English Wikipedia to improve and discuss”. The success of Wikipedia in spreading a climate of mutual respect and trust among its users is proof that similar platforms can prosper online: the key challenge thus is to find ways of adapting this success story to different contexts of online interaction, such as social media.

Acknowledgments

We are indebted to two anonymous reviewers whose comments allowed a substantial improvement of the work. We thank Diya Elizabeth Abraham, Stefania Bortolotti, Martijn J. Burger, Fulvio Castellacci, Giuseppe Ciccarone, Miloš Fišar, Alessandra Gualtieri, Monica D'Agruma, Giacomo Degli Antoni, Giovanni Di Bartolomeo, Antonio Montañés Bernal, Vittorio Pelligra, Matteo Rizzolli, Jiří Špalek, Luca Stanca, and Luca Zarri for comments and suggestions. The paper also benefited from comments by participants at the Workshops “Social capital, institutions and economic development” in Valencia, “ICTs and well-being” in Oslo, “Regional Growth Differentials and Social Capital” in Cagliari, at the Conferences “Social and Cultural

¹² https://en.wikipedia.org/wiki/Wikipedia:Five_pillars.

Changes in Cross-National Perspective” in Moscow, “Social Capital” in Konya, at the Global Economic Symposium “Restructuring Economies, Transforming Societies” in Kuala Lumpur, at research seminars at Erasmus University of Rotterdam, Seinan Gakuin University of Fukuoka, Technical University of Munich, and University of Zaragoza. The research of Angelo Antoci and Fabio Sabatini has been funded by [Regione Autonoma della Sardegna](#) (Grant no. [CRP-60619](#)) (L. R. n. 7, 2007; research project “Capitale sociale e divari economici regionali”). The research of Tommaso Reggiani has been funded by the [Czech Science Foundation](#) (grant: [GA18-19492S](#)). The research of Laura Bonelli and Fabio Paglieri has been funded by the Italian Ministry of Education, University and Research, [MIUR](#) (PON research project CLARA, Cloud plAtform and smart underground imaging for natural Risk Assessment). Grant number: [SCN 00451](#).

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.jebo.2019.02.026](https://doi.org/10.1016/j.jebo.2019.02.026).

References

- Accetturo, A., De Blasio, G., Ricci, L., 2014. A tale of an unwanted outcome: transfers and local endowments of trust and cooperation. *J. Econ. Behav. Organ.* 102, 74–89.
- Aitken, R.C., 1969. Measurement of feelings using visual analogue scales. *Proc. R. Soc. Med.* 62 (10), 989–993.
- Alesina, A., La Ferrara, E., 2002. Who trusts others? *J. Public Econ.* 85 (2), 207–234.
- Altavilla, C., Luini, L., Sbriglia, P., 2006. Social learning in market games. *J. Econ. Behav. Organ.* 61, 632–652.
- Anderson, A.A., Brossard, D., A., S.D., 2014. The nasty effect: online incivility and risk perceptions of emerging technologies. *J. Comput.-Med. Commun.* 19 (3), 373–387.
- Anderson, M., 2018. A Majority of Teens Have Experienced Some Form of Cyberbullying. Technical Report. Pew Research Center, Washington, DC.
- Anderson, M., Perrin, A., 2017. Tech adoption climbs among older adults. Technical Report. Pew Research Center.
- Andrighetto, G., Zhang, N., Ottone, S., Ponzano, F., D'Atoma, J., Steinmo, S., 2016. Are some countries more honest than others? evidence from a tax compliance experiment in sweden and italy. *Front. Psychol.* 7, (article 472).
- Antoci, A., Delfino, A., Paglieri, F., Panebianco, F., Sabatini, F., 2016. Civility vs. incivility in online social interactions: an evolutionary approach. *PLoS ONE* 11 (11), e0164286.
- Antoci, A., Sabatini, F., 2018. Online networks, social interaction and segregation: an evolutionary approach. *J. Evolut. Econ.* 28 (4), 859–883.
- Antoci, A., Sabatini, F., Sodini, M., 2011. The solaris syndrome: social capital in a growing hyper-technological economy. *J. Econ. Behav. Organ.* 81 (3), 802–814.
- Antoci, A., Sabatini, F., Sodini, M., 2015. Online and offline social participation and social poverty traps. *J. Math. Sociol.* 39 (4), 229–256.
- Arrow, K.J., 1972. Gifts and exchanges. *Philos. Public Affa.* 1 (4), 343–362.
- Banfield, E.C., 1958. *The Moral Basis of a Backward Society*. Free Press, New York.
- Barberà, P., Jost, J., Nagler, J., Tucker, J., Bonneau, R., 2015. Tweeting from left to right: is online political communication more than an echo chamber? *Psychol. Sci.* 26 (10), 1531–1542.
- Barberà, P., Rivero, G., 2015. Understanding the political representativeness of twitter users. *Soc. Sci. Comput. Rev.* 33 (6), 712–729.
- Bauernschuster, S., Falck, O., Woessmann, L., 2014. Surfing alone? the internet and social capital: evidence from an unforeseeable technological mistake. *J. Public Econ.* 117, 73–89.
- Ben-Ber, A., Putterman, L., 2009. Trust, communication and contracts: an experiment. *J. Econ. Behav. Organ.* 70 (1–2), 106–121.
- Berg, J., Dickhaut, J., McCabe, K., 1995. Trust, reciprocity, and social history. *Games Econ. Behav.* 10, 121–142.
- Berggren, N., Bjørnskov, C., 2011. Is the importance of religion in daily life related to social trust? cross-country and cross-state comparisons. *J. Econ. Behav. Organ.* 80 (3), 459–480.
- Bigoni, M., Bortolotti, S., Casari, M., Gambetta, D., 2018. At the root of the north-south cooperation gap in italy. preferences or beliefs? *Econ. J.* doi:[10.1111/eoj.12608](https://doi.org/10.1111/eoj.12608).
- Bigoni, M., Bortolotti, S., Casari, M., Gambetta, D., Pancotto, F., 2016. Amoral familism, social capital, or trust? the behavioral foundations of the italian north-south divide. *Econ. J.* 126 (594), 1318–1341.
- Bigoni, M., Fort, M., Nardotto, M., Reggiani, T., 2015. Cooperation or competition? a field experiment on non-monetary learning incentives. *B.E. J. Econ. Anal. Policy* 15 (4), 1753–1792.
- Bigoni, M., Suetens, S., 2012. Feedback and dynamics in public good experiments. *J. Econ. Behav. Organ.* 82, 86–95.
- Boxell, L., Gentzkow, M., Shapiro, J., 2017. Greater internet use is not associated with faster growth in political polarization among us demographic groups. *Procee. Nat. Acad. Sci. U.S.A.* 114 (40), 10612–10617.
- Burdin, G., Halliday, S., Landini, F., 2018. The hidden benefits of abstaining from control. *J. Econ. Behav. Organ.* 147, 1–12.
- Campante, F., Durante, R., Sobbrío, F., 2017. Politics 2.0.: the multifaceted effect of broadband internet on political participation. *J. Eur. Econ. Assoc.* forthcoming.
- Carpenter, J., Danieri, A.G., Takahashi, L.M., 2004. Cooperation, trust, and social capital in southeast asian urban slums. *J. Econ. Behav. Organ.* 55 (4), 533–551.
- Casari, M., Ichino, A., Michaeli, M., De Paola, M., Marandola, G., Scoppa, V., 2018. Civicness drain. IZA Discussion Paper No. 11955.
- Castellacci, F., Schwabe, H., 2017. Internet use and well-being in different domains of life. Paper presented at the conference The effects of ICTs on Well-being, Oslo, 4–5 September 2017.
- Castellacci, F., Tveito, V., 2018. Internet use and well-being: a survey and a theoretical framework. *Res. Policy* 47 (1), 308–325.
- Chiapello, M., 2018. BALANCETABLE: Stata module to build a balance table. Technical Report. Boston College - Department of Economics.
- Conrads, J., Reggiani, T., 2017. The effect of communication channels on promise-making and promise-keeping: experimental evidence. *J. Econ. Interac. Coord.* 12 (3), 595–611.
- Del Vicario, M., Vivaldo, G., Bessi, A., Zollo, F., Scala, A., Caldarelli, G., Quattrociocchi, W., 2016. Echo chambers: emotional contagion and group polarization on facebook. *Sci. Rep.* 6, 37825.
- Diener, E., 1979. Deindividuation, self-awareness, and disinhibition. *J. Pers. Soc. Psychol.* 37 (7), 1160–1171.
- Duggan, M., 2014. Online Harassment 2014. Technical Report. Pew Research Center, Washington, DC.
- Duggan, M., 2017. Online Harassment 2017. Technical Report. Pew Research Center.
- Duggan, M., Smith, A., 2016. The political environment on social media. Washington, DC: Pew Research Center.
- Eckel, C.C., Füllbrunn, S.C., 2015. Thar she blows? gender, competition, and bubbles in experimental asset markets. *Am. Econ. Rev.* 105 (2), 906–920.
- Ellison, N.B., Steinfield, C., Lampe, C., 2007. The benefits of facebook “friends”: social capital and college students’ use of online social network sites. *J. Comput.-Med. Commun.* 12, 1143–1168.
- Falck, O., Gold, R., Heblich, S., 2014. E-Lectons: voting behavior and the internet. *Am. Econ. Rev.* 104 (7), 2238–2265.
- Fehr, E., Fischbacher, U., Kosfeld, M., 2005. Neuroeconomic foundations of trust and social preferences: initial evidence. *Am. Econ. Rev.* 95 (2), 356–351.
- Fischbacher, U., 2007. Z-Tree: zurich toolbox for ready-made economic experiments. *Exper. Econ.* 10 (2), 171–178.

- Fukuyama, F., 1995. *Trust: the Social Virtues and the Creation of Prosperity*. Free Press, New York.
- Geraci, A., Nardotto, M., Reggiani, T., Sabatini, F., 2018. Broadband internet and social capital. IZA Discussion Paper 11855.
- Glaeser, E., Laibson, D., Sacerdote, B., 2002. An economic approach to social capital. *Econ. J.* 112 (483), F437–F458.
- Glaeser, E., Laibson, D., Scheinkman, J.A., Soutter, C.L., 2000. Measuring trust. *Q. J. Econ.* 115 (3), 811–846.
- Greenwood, S., Perrin, A., Duggan, M., 2016. Social Media Update 2016. Technical Report. Pew Research Center.
- Greiner, B., 2015. Subject pool recruitment procedures: organizing experiments with orsee. *J. Econ. Sci. Assoc.* 1 (1), 114–125.
- Guerra, A., Harrington, B., 2018. Attitude–behavior consistency in tax compliance: a cross-national comparison. *J. Econ. Behav. Organ.* doi:10.1016/j.jebo.2018.10.013.
- Guerra, G., Zizzo, D.J., 2004. Trust responsiveness and beliefs. *J. Econ. Behav. Organ.* 55 (1), 25–30.
- Guido, G., Peluso, A., Capestro, M., Miglietta, M., 2015. An italian version of the 10-item big five inventory: an application to hedonic and utilitarian shopping values. *Pers. Individ. Dif.* 76, 135–140.
- Gupta, G., Mahmud, M., Maitra, P., Mitra, S., Neelam, A., 2018. Religion, minority status, and trust: evidence from a field experiment. *J. Econ. Behav. Organ.* 146, 180–205.
- Hampton, K., Wellman, B., 2003. Neighboring in netville: how the internet supports community and social capital in a wired suburb. *City Commun.* 2 (4), 277–311.
- Harrison, G., List, J., 2004. Field experiments. *J. Econ. Lit.* 42 (4), 1009–1055.
- Hollander, M., Wolfe, D., 1999. *Nonparametric statistical methods*. Wiley, Hoboken, NJ.
- John, O.P., Srivastava, S., 1999. The Big-five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. In: Pervin, L.A., John, O.P. (Eds.), *Handbook of personality: Theory and research*, 2. Guilford Press, New York, pp. 102–138.
- Johnson, N.D., Mislin, A.A., 2011. Trust games: a meta-analysis. *J. Econ. Psychol.* 32, 865–889.
- Kiesler, S., Siegel, J., McGuire, T.W., 1984. Social psychological aspects of computer-mediated communication. *Am. Psychol.* 39 (10), 1123–1134.
- Kirschner, P., 2015. Facebook as learning platform: argumentation superhighway or dead-end street? *Comput. Human Behav.* 53, 621–625.
- Knack, S., 2001. Trust, Associational Life and Economic Performance. In: Heliwell, J.F., Bonikowska, A. (Eds.), *The Contribution of Human and Social Capital to Sustained Economic Growth and Well-Being*. Human Resources Development Canada and OECD, Ottawa and Paris, pp. 172–202.
- Knack, S., Keefer, P., 1997. Does social capital have an economic payoff? *Q. J. Econ.* 112 (4), 1251–1288.
- Lea, M., O'Shea, T., Fung, P., Spears, R., 1992. Flaming in computer-mediated-communication, Contexts of computer-mediated communication. *Harvester Wheatsheaf, Hertfordshire, England*, pp. 89–112.
- Ljunge, M., 2014. Trust issues: evidence on the intergenerational trust transmission among children of immigrants. *J. Econ. Behav. Organ.* 106, 175–196.
- Madden, M., Lenhart, A., Cortesi, S., Gasser, U., Duggan, M., Smith, A., Beaton, M., 2013. Teens, social media, and privacy. Technical Report. Pew Research Center, Washington, DC.
- Mutz, D.C., Reeves, B., 2005. The new videomalaise: effects of televised incivility on political trust. *Am. Polit. Sci. Rev.* 99 (1), 1–15.
- Nunn, N., Wantchekon, L., 2011. The slave trade and the origins of mistrust in africa. *Am. Econ. Rev.* 101 (7), 3221–3252.
- Paglieri, F., 2017. A plea for ecological argument technologies. *Philos. Technol.* 30 (2), 209–238.
- Pénard, T., Poussing, N., Suire, R., 2013. Does the internet make people happier? *J. Socio-Econ.* 46, 105–116.
- Putnam, R.D., 2000. *Bowling Alone: the Collapse and Revival of American Community*. Simon & Schuster, New York.
- Putnam, R.D., Leonardi, R., Nanetti, R.Y., 1993. *Making Democracy Work: civic Traditions in Modern Italy*. Princeton University Press, Princeton, NJ.
- Quattrociocchi, W., Scala, A., Sunstein, C., 2016. Echo chambers on facebook. Harvard Law School, Discussion Paper N. 877.
- Rainie, L., 2017. The reckoning for social media. Keynote speech at the New Media and Society conference in Toronto on July 29, 2017.
- Rainie, L., 2018. Americans' complicated feelings about social media in an era of privacy concerns. Technical Report. Pew Research Center.
- Rainie, L., Anderson, J., Albright, J., 2017. The Future of Free Speech, Trolls, Anonymity and Fake News Online. Technical Report. Pew Research Center, Washington, DC.
- Rainie, L., Kiesler, S., Kang, R., Madden, M., 2013. Anonymity, Privacy, and Security Online. Technical Report. Pew Research Center, Washington, DC.
- Rainie, L., Lenhart, A., Smith, A., 2012. The tone of life on social networking sites. Technical Report. Pew Research Center, Washington, DC.
- Rammstedt, B., John, O.P., 2007. Measuring personality in one minute or less: a 10-item short version of the big five inventory in english and german. *J. Res. Pers.* 41, 203–212.
- Rosenberg, M., 1956. Misanthropy and political ideology. *Am. Sociol. Rev.* 21, 690–695.
- Rost, K., Stahel, L., Frey, B., 2016. Digital social norm enforcement: online firestorms in social media. *PLoS ONE* 11 (6), e0155923.
- Sabatini, F., Sarracino, F., 2017. Online networks and subjective well-being. *Kyklos* 70 (3), 456–480.
- Sent, E.-M., van Staveren, I., 2018. A feminist review of behavioral economic research on gender differences. *Fem. Econ.* 1–35.
- Siegel, J., Dubrowsky, W., Kiesler, S., McGuire, T.W., 1986. Group processes in computer-mediated communication. *Organ. Behav. Hum. Decis. Process.* 37 (2), 157–187.
- Sproull, L., Kiesler, S., 1986. Reducing social context cues: electronic mail in organizational communication. *Manage. Sci.* 32 (11), 1492–1513.
- Steinfeld, C., Ellison, N.B., Lampe, C., 2008. Social capital, self-esteem, and use of online social network sites: a longitudinal analysis. *J. Appl. Dev. Psychol.* 29, 434–445.
- Valenzuela, S., Park, N., Kee, K.F., 2009. Is there social capital in a social network site?: facebook use and college students' life satisfaction, trust, and participation. *J. Comput.-Med. Commun.* 14 (4), 875–901.
- Yamagishi, T., 1986. The provisioning of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* 51, 110–116.
- Yamagishi, T., Yamagishi, M., 1994. Trust and commitment in the united states and Japan. *Motiv. Emot.* 18 (2), 129–166.
- Yamamura, E., 2009. Dynamics of social trust and human capital in the learning process: the case of the Japan garment cluster in the period 1968–2005. *J. Econ. Behav. Organ.* 72 (1), 377–389.
- Yasseri, T., Sumi, R., Rung, A., Kornai, A., Kertész, J., 2012. Dynamics of conflicts in wikipedia. *PLoS ONE* 7 (6), e38869.
- Zizzo, D.J., 2010. Experimenting demand effects in economic experiments. *Exper. Econ.* 13 (1), 75–98.