Giovanna Devetag, Francesca Ceccacci and Paola De Salvo

Do Reputation Concerns Make Behavioral Biases Disappear? The Conjunction Fallacy on Facebook and Mechanical Turk

CEEL Working Paper 3-13

Do Reputation Concerns Make Behavioral Biases Disappear? The Conjunction Fallacy on Facebook and Mechanical Turk[*]

Giovanna Devetag[†]

Francesca Ceccacci[‡]

Paola De Salvo[§]

May 10, 2013

Abstract

This paper reports the results of experiments designed to test whether individuals interacting on Facebook are more likely to succumb to the conjunction fallacy when they post their answers publicly and are exposed to the answers of others. Using the experimental design in Kahneman and Tversky (1983), we find that the proportion of individuals violating the conjunction rule on Facebook is substantially lower than that reported by previous experiments conducted in the lab, regardless of whether responses are public or private. When responses are posted in a public form, however, the participation rate is substantially higher. The violation rate on Facebook is also significantly lower than the rate of violation from the same experiment run on Mechanical Turk, a popular online labor market, with monetary incentives. Adding a bonus for the correct answer reduces the violation rate on Mechanical Turk when answers are private,

but not when they are public, suggesting that peer effects may indeed counteract the effect of monetary incentives. Our experiment casts doubts about the robustness of behavioral biases for the understanding of real life decisions in environments in which interaction is not anonymous and people are reputation conscious, and suggests the power of social networks to mitigate their effects.

1   Introduction

Fads, herd behavior, information cascades, "tipping point" phenomena and observational learning are all instances of peer effects, i.e. the influence exerted on our own judgment, choice and behavior by the observation of the choices, judgments and behavior of others. Peer effects are universally recognized as having potentially important and at times dramatic aggregate consequences, leading to bank runs and asset bubbles, to financial crises, recessions, and even to revolutions. Accordingly, in recent years social scientists have devoted increasing attention to the understanding of this class of social phenomena, both through the analysis of field data and through extensive laboratory experimentation.

To the extent that peer effects all have to do with being influenced in one's judgment and behavior by the exposure of what other people think and do, the recent large popularity gained by applications related to the world of Internet 2.0, and of social networks like Facebook and Twitter in particular, has amplified their chances of occurrence enormously as well as the speed and magnitude of their potential aggregate consequences. In fact, social networks allow every participant to instantly share news, opinions, images, and videos worldwide, and hence to contribute to the onset of collective movements based on imitation and closely related mechanisms; several political analysts, for example, have pointed out the key role played by social networks in the recent "Arab spring" uprisings in the Middle East, in the Iranian green movement that took place worldwide after the 2009 presidential elections in Iran, and in the more recent "Occupy" movements originated in the US and diffused across Europe in response to the recent financial crisis. As we are writing the introduction to this article, Italian political experts are commenting on how Facebook and Twitter are influencing the ongoing presidential elections by creating social pressure towards members of the Parliament.

In this paper we use the leading worldwide social network, Facebook, to set up a natural

field experiment on one of the most classic and widely studied biases in individual judgment: the *conjunction fallacy* (Kahneman and Tversky 1983). The conjunction rule, stating that the probability of the intersection of two independent events cannot be greater than the probability of either of the two events taken separately, is one of the most basic rules of modern probability theory, and therefore a basic principle of any theory of rational choice under uncertainty. Despite its simplicity, several experimental studies have shown that many people violate the rule when expressing probability judgments, thus exhibiting the conjunction fallacy. The first and most famous experiment on the conjunction fallacy, also known as the "Linda" problem, was conducted by Kahneman and Tversky (1983), using the text reported below (Kahneman and Tversky, 1983, p. 297:)

> Laura is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which of these two alternatives do you think is more probable?
>
> 1. Laura is a bank teller
> 2. Laura is a bank teller and is active in the feminist movement

Kahneman and Tversky (1983) report that 85% of their experimental subjects judged option 2 as being more probable than option 1, thus violating the conjunction rule. The authors report similar violation rates when variants of the reported problem are introduced, and explain their findings by suggesting that individuals use intuitive heuristics such as "representativeness" to formulate their judgments, thus committing errors as in the case of "Linda" (Kahneman and Tversky 1972, Kahneman 1972).

The conjunction fallacy has been widely studied and replicated, and findings are mixed. They will be reviewed in the next section. In the present study we investigate the extent to which the fallacy may be replicated in a more "natural" environment where peer effects of the type mentioned above are very likely to occur; more specifically, we want to test whether a specific environment - Facebook - where people spend much of their time posting information and diffusing information posted by others, and where, therefore, the likelihood of peer effects of various sorts is particularly high, is more or less conducive to massive violations of the conjunction rule with respect to more traditional lab environments where judgments are typically made in isolation.

Today Facebook is the #6 ranked Web property worldwide, reaching 275 million visitors in February 2011, a 175-percent increase on the year before. One of the global regions that

3

best illustrates Facebook's growth is Europe, where the social network has experienced a 314-percent increase to nearly 100 million visitors. The Facebook audience is highly engaged, with the average user spending three hours per month on the site. The site audience is largest in the U.K. with 22.7 million visitors, followed by France with 13.7 million visitors and Turkey with 12.4 million visitors. It has also soared in Italy, growing more than 2,700 percent in one year only, reaching 10.8 million visitors in 2009. Data show that while Facebook already reaches over half of the total internet audience worldwide, its audience is still increasing, and it is spreading from the US and Europe to emerging economies. The first four countries for absolute numbers of subscribers to Facebook are the United States, Brazil, India and Indonesia. Italy ranks eleventh, while the fastest growing Facebook countries are Indonesia, Bolivia, Bosnia Hercegovina and Algeria[1]. The leading role of Facebook in the world of internet users and its absolute predominance among social networks makes it a natural candidate to study phenomena related to peer effects in judgment and decision making.

Facebook is designed in such a way that everything a person posts on their personal profile (in forms of, e.g., status updates, links to articles or videos from external Internet sources, links to documents in Word or .pdf format, uploads of pictures, short notes, etc.) is made visible to that person's network of "friends" (Facebook contacts) and can usually, by default, be shared by any one of them (unless the original publisher poses some restrictions on diffusion) in order to be seen and possibly shared by their networks of friends, and so on and so forth. A person can also choose to make a content "public", i.e., visible to all Facebook members beyond their own network of friends. Besides one's personal profile page, whose updates are promptly signaled in the *News Feed* page for others to see, other very popular Facebook features are discussion forums or groups, joined by persons interested in exchanging information and opinions on a particular topic, and events, which are temporary groups created for specific occasions. Anybody can share content coming from an individual profile and from outside Facebook both by publishing it on their own profile page and by sharing it within groups, forums, and events, and anybody can express their appreciation for a specific content by pressing the *Like* button or by writing a comment. Usually all comments posted on a profile page are visible to all of the profile owner's friends, and all comments posted in a group by members are usually visible to all Facebook users if the group settings allow this (and the default option is for a group to be visible to all). Hence, a specific content can potentially reach thousands or even millions of people in a matter of days, hours, or even minutes. The average Facebook user is continuously exposed to a

---

[1]Source: Business Insider, technology.inquirer.net

considerable number and variety of both contents and opinions on those contents from other users.

In such a closely connected virtual environment, it is natural to ask whether the appearance of behavioral biases may be favored by mechanisms of imitation, conformity, herd behavior or observational learning. If this is the case, then, one could conclude that social networks like Facebook, with their enormous power to connect people and diffuse information and opinions may risk "amplifying" phenomena of individual irrationality.

In order to test this conjecture, we designed a natural field experiment in the following way: we identified four Facebook accounts with no connection to one another, and we instructed each account owner (henceforth: *administrator*) to create two parallel "events" simultaneously: a "private" and a "public" event, corresponding to our two experimental treatments. Half of each administrator's Facebook friends were to be invited to the public event, and the other half to the private event, according to a random mechanism. The "event" consisted in an invitation to participants to answer a simple question without consulting others. The text of the "Linda" problem was displayed and participants were invited to state whether option 1 or option 2 were more probable. The only difference between the two events was in the response mode: in the "private" event, participants were asked to send their answer to the account administrator in a private form (so that nobody could see the other participants' responses), while in the "public" event, participants were asked to write their answer directly on the event's "wall" (a public space where every person invited to an event can post a message), so that everybody could see the sequence of previous responses before giving their own. The public modality is the typical Facebook format to visualize comments on a specific content, which are displayed sequentially in chronological order. So in this respect, the *public* treatment is the one where we expected to observe peer effects. The *private* treatment is designed as a control treatment. However, the peculiar feature of our experiment on Facebook (which, if on the one hand seemed unavoidable, on the other hand renders our experiment more "natural" as it makes it similar to all other types of interaction occurring on the social network) is voluntary participation. That is, participants were invited to answer the question, and they could decide whether or not to answer the question *after* reading it. This feature affects both treatments equally, therefore it is not supposed to matter for our results. However, to introduce a further control for voluntary participation and self-selection, we conducted the same experiment by using participants recruited through Mechanical Turk (Paolacci et al. 2010), an online labor market where requesters hire workers to perform simple tasks under for payment of small sums of money. Mechanical Turk is increasing in popularity as a reliable source of experimental subjects for

surveys and experiments on individual judgment and decision making. We conducted the experiment on Mechanical Turk in order to add financial incentives to our design, and in order to control for self-selection. We conducted a *flat fee* treatment, where participants responded to the "Linda" problem for a fixed payment, and a *bonus* treatment, where we added a bonus payment for the correct answer. In both treatments, we replicated the private and public response modes, with every participant being randomly assigned to either modality.

We formulated the hypothesis that peer effects may amplify the magnitude of the conjunction fallacy, leading to higher violation rates in the *public* treatment than in the *private* treatment, for both Facebook and Mechanical Turk. For the same reason (peer effects) we expected violation rates in Facebook and Mechanical Turk- public treatments to be higher than those typically found in traditional lab experiments. Additionally, we assumed that adding monetary incentives would reduce the magnitude of the violation, in line with previous studies. We motivate our assumptions in detail in the next section. Our results somewhat contradict our initial hypotheses; in the Facebook experiment, we fail to replicate previous results on the conjunction fallacy: in fact, in both the *private* and the *public* treatment, the large majority of participants identify the correct answer, with no significant differences between the two distributions. What differs between the *private* and the *public* treatment is the participation rate, which is significantly higher in the latter. In the Mechanical Turk experiment, on the contrary, the majority of our experimental subjects commit the fallacy, thus replicating previous findings. The addition of a bonus for the correct answer significantly reduces the violation rate in the *private* treatment, but not in the *public* treatment. These findings combined suggest that the anomalous results of our Facebook experiments can be explained by a combination of self-selection and willingness to signal one's status to others (where "status" in this context refers to the ability to identify the correct answer to a problem whose correct answer is not obvious).

The rest of the paper is organized as follows: section 2 reviews previous literature and specifies our motivation; section 3 describes the experimental design and implementation and section 4 presents the results of our experiments: in section 5 we discuss our findings, and finally, section 6 offers some concluding remarks.

2   Related Literature

The study of social influences on individual judgment and decision making has a long and important tradition in social psychology (see Levine 2013 for a review). In the famous experiment by Asch (1956), subjects were asked to judge the length of a bar in a task where

the correct answer (among the three available) was made self-evident. When experimental subjects were exposed to previous wrong answers (purposely provided by fake subjects who were in fact confederates of the experimenter), roughly one-third on average were induced to provide the wrong answer, whereas the error rate in the control group was less than 1%; the conformity effect was proportional to the size of the dominant majority, although it decreased substantially in the absence of unanimity (Asch 1951, 1952, 1955, 1956). Later studies introduced the important distinction between normative social influence and informational social influence, the latter occurring for judgments or choices in the case of intellective tasks, such as problems that have a single correct solution. While normative social influence is defined as the willingness to conform in order to gain social reward or avoid punishment without necessarily changing one's own private beliefs to align them with those of the majority, Informational social influence derives from the acceptance of persuasive arguments by others or simply from the belief that others possess more knowledge or information, and/or superior cognitive ability which allow them to identify the correct answer to a problem (Hogg and Turner 1987). Research on social influence in psychology suggests that the most likely outcome of social influence is the convergence of judgments and decisions to some group "norm". Extreme pressure to conform may create judgment distortions in the form of groupthink (Janis 1982, Turner and Pratkanis 1998) and group polarization (Myers and Lamm 1975, Isenberg 1986).

In the last twenty years economics has paid increasing attention to social learning environments too: herding behavior and informational cascades are well known aggregate effects of our tendency to be influenced in our judgments and decisions by exposure to the judgments and decisions of others (Banerjee 1992, Bikhchandani, Hirshleifer and Welch 1992 for a review). The influence can be rational insofar as it is justified by the belief that others are more informed. It becomes "irrational" when the behavior of others is given disproportionate weight in one's judgment compared to the weight assigned to private signals (Devenow and Welch, 1996, Bohren 2010). Under conditions of scarce information, even rational actors can herd on the incorrect action with positive probability; moreover, if players are naive and incorrectly think that the actions of others are motivated solely by private signals, herding on the incorrect action can occur even in information-rich settings (Eyster and Rabin 2010). Goeree and Yariv (2007) conduct an experiment to disentangle conformity, intended as a mere preference to follow the majority, from rational, information driven herding. Their results show that herd behavior is explained by conformity rather than by best responding to lack of private signals. A recent field experiment conducted on Facebook shows evidence of herding behavior, in that participants were on average more likely to *like* a status update

if other participants had done so (Egebark and Ekstrm 2011). The probability of expressing a positive opinion increased both with the number of predecessors who had expressed a positive opinion and with social proximity. The authors isolate conformity as the main driver of herd behavior in their experiment.

Facebook is an ideal environment to study pressure to conform, given the high exposure of any action taken, and given that much of Facebook activity consists in publicly expressing opinions and beliefs and commenting on them. Our working hypothesis is very simple and derives from the classic studies on conformity and peer pressure. Given that the conjunction fallacy and other cognitive biases are fairly well known violations of the principles of individual rationality, which emerge when the majority of a population provides the incorrect answer to a choice or judgment problem, if peer pressure and conformity produce herding on the incorrect response as in Asch's classic experiment, then we would expect the magnitude of the violation to be greater when responses are posted in public form. The reason is intuitive: even if not everybody feels the pressure to conform, individuals who do not feel the pressure will be likely to commit the fallacy in any case, given previous findings. Among the individuals who are more subject to the pressure to conform, the ones who would have committed the fallacy even if performing the task in isolation will contribute to increasing the violation rate, and even those who otherwise would have been uncertain or would have identified the correct answer with some effort may be inclined to "follow the crowd". Hence, it is reasonable to expect a higher violation rate in the *public* treatment than in the *private* one. In addition, the conjunction fallacy is the ideal task to test the emergence of social informational influence beyond normative influence. The correct answer is not at all obvious, which makes it natural to observe what answer other people provide (Turner et al. 1987). Our experiment does not allow us to disentangle normative from informational influence or learning by imitation: it is not possible, in our set up, to know whether, provided that herding does occur, it is generated by the mere desire to conform to the majority or by rational observational learning in the presence of uncertainty. Our goal is simply to test whether the magnitude of the conjunction fallacy is amplified by interaction on a social network. Next we discuss our experimental design in detail.

3   The Experiments

3.1   The experiment on Facebook

The experiment on Facebook was organized as follows: first, we identified four administrators who would agree to implement the experiment through their Facebook account. None of

these persons belonged to the academic milieu, to minimize the probability that some of their contacts could be familiar with the problem. The age of the administrators ranged from 18 to 59, and the number of Facebook "friends" was equal to at least 100 for each administrator. There was no overlapping among the administrators' list of contacts.

Each administrator was instructed to create two parallel events on Facebook, a *private* event and a *public* one. Both events asked participants to answer a simple question without consulting other people. The two events were to be created simultaneously, and *only* differed in the modality with which invited contacts were asked to answer the question "contained" in the event. Each administrator's list of Facebook friends had been previously divided into two groups according to a random criterion: those with name initials in the (A-K) range were to be invited to one of the two events on the basis of the toss of a coin, while those with name initials in the (L-Z) range were invited to the other event. The coin was tossed separately for each administrator, for a total of four times. Contacts invited to either event were unaware of the existence of the other event (each event was visible only to invited contacts).

The text was identical in the two events *except* for the instructions on how to answer the question. The Linda scenario was displayed and the two possible answers were numbered 1 and 2. The *public* version of the event asked participants to provide their answers by writing the answer number, "1" or "2" directly on the event's wall, while the *private* version of the event asked participants to give their answers in a private form by sending a message to the administrator. In this way, in the *public* treatment, answers were publicly available and the history of previous responses was visible to all participants.

We used the "transparent" version of the Linda problem also used in Charness et al. (2009) to maximize comparability with previous studies (the text is available in the Appendix). Only the name of the character was changed from Linda to Laura to further reduce the probability that some of the invited participants might recall having heard about the problem before. Furthermore, to minimize possible order effects, the correct answer was placed as the first answer in one version of the text, and as second in another version.

We decided to collect all the answers received by the tenth day after the event publication. Over 90% of the responses, however, were collected within the first hour following publication of the event. Table 1 summarizes the experimental design.

There are several peculiarities of our experimental design that need attention. First of all, in our experiment responses are not anonymous. This is an important difference with respect to traditional laboratory experiments which, however, is unavoidable and makes our experimental test useful for external validity purposes. Given that our main goal is to

9

investigate the role of social networks on herd behavior, it is essential that our experiment mirrors the most important features of interaction through social networks. Secondly, our subjects could easily "google" the Linda text and find the right answer in a matter of few seconds. Third, our subjects in the Facebook experiment were not financially incentivized. Since all these features occur in both the public and private treatment, none of them is supposed to matter for our results. However, in order to introduce a further control, we implemented the same experiment on an online platform called Mechanical Turk introducing monetary incentives. The experiment is described in the next section.

3.2 The experiment on Mechanical Turk

Mechanical Turk (MT, henceforth) is a web service created to coordinate worldwide supply and demand of tasks requiring human intelligence (called HITs - Human Intelligence Tasks in MT terminology). The platform functions as an online labor market where employees (workers) are recruited by employers (requesters) for the execution of simple tasks in exchange for a wage. Both workers and requesters are anonymous, although each is identified through an automatic ID provided by Amazon at registration. Requesters can ask that workers satisfy predefined criteria (e.g., country of residence, age, education, or reputation). Tasks are typically simple enough (e.g., survey completion, manuscript proofreading, image tagging, audio transcriptions, etc.) to be completed in a few minutes, and payment usually occurs immediately after task completion. Typical hourly wages are about $1.40. Recently, the MT platform has been used increasingly by scholars as a source of subjects for experimental research. Recent studies (Paolacci et al. 2010) have compared the results from a set of famous experiments (including the "Linda" problem) performed on MT, on other online platforms and by using subjects from traditional sources: results show that subjects recruited through MT exhibit the canonical cognitive biases and pay at least as much attention to instructions as more traditional subject pools, confirming the validity of Mechanical Turk as a source of reliable data for judgment and decision making experiments. Results on the conjunction fallacy experiment in particular show no significant difference in the percentage of correct responses with respect to other online platforms and to traditional lab experiments performed with student subjects (Paolacci et al. 2010).

Our experimental design included two conditions: in the *flat fee* condition, participants (workers) were paid only a fixed amount for participating in the survey; in the *bonus* condition, they were paid a bonus, in addition to the fixed reward, for choosing the correct answer.

Implementation was as follows: workers interested in participating were shown a message

10

stating that the task would consist in answering a single question for a fixed payment equal to $.25 (plus a bonus of $.25 for responding correctly in the *bonus* treatment); then the screen in the *private* treatment reported the scenario asking subjects to write their answer ("1" or "2") in the appropriate window and click "next". Then a short demographic survey followed and the task was over. In the *public* treatment, after the initial message was displayed, a public pad appeared showing the problem text and asking subjects to provide their answer by writing their worker ID in the appropriate column followed by their answer ("1" or "2"). In this way, any respondent after the first one could read previous answers before giving his own. Furthermore, responses were not given anonymously, since workers had to write down their IDs in order to get paid[2]. A further similarity with the Facebook experiment is the possibility to "google" the answer.

Half of the scenarios in both conditions and treatments reported "Laura is a bank teller" as response n.1, the other half reported it as response n.2. The assignment of a subject to either scenario was random. The number of subjects was specified in advance, and the experiment was terminated when the predetermined number of responses had been collected. In both conditions, the data collection phase lasted approximately 48 hours.

Table 1 reports a summary of the experimental design.

In accordance with our hypothesis on peer effects, we expect to observe a higher violation rate in the *public* treatment than in the *private* treatment, in both the *flat fee* and the *bonus* condition. We also expect to observe a lower violation rate in the *bonus* condition as opposed to the *flat fee* condition, in line with previous results on the effect of financial incentives in mitigating the incidence of behavioral biases (see, e.g., Camerer 1995). Concerning the comparison between Facebook and Mechanical Turk, on the one hand we would expect to observe a lower violation rate in Mechanical Turk as opposed to the corresponding treatment in Facebook, due to the presence of financial incentives in the former. On the other hand, the possibility to self-select into the task, present in Facebook but not in Mechanical Turk, might in principle offset the effect of financial incentives, if correct responders are overrepresented in the portion of invited subjects who decide to participate. Hence, we remain impartial as to the effects of the two environments on our results.

The Appendix reports experimental instructions for both Facebook and MT experiments; figure 1 reports a screenshot from one of the events in the *public* treatment on Facebook,

---

[2]Although providing a worker ID is not as strong as providing one's full name as happens on Facebook, a worker's ID in Mechanical Turk is linked to that worker's reputation, which influences the probability of being hired in the future. Moreover, one cannot be sure that identities on Facebook are truthful, given the high estimated frequency of "fake" accounts. Hence, it is fair to consider that both our experimental environments (MT and Facebook) include non-anonymity

	Facebook	Mechanical Turk flat fee	Mechanical Turk with bonus
Private	N= 74	N=136	N=103
Public	N= 207	N=87	N=99

Table 1: A summary of the experimental design

and fig. 2 reports a screenshot of the Public pad that subjects in the MT experiment used to post their answers in the *public* treatment.

4 Results

Tables 2 and 3 show the results from the Facebook experiment in the *private* and *public* treatments respectively. Several interesting findings emerge: first, contrary to our hypothesis, there is no significant difference in the percentage of correct responses between the *private* and the *public* version (the two distributions are not significantly different according to a Fisher exact test; $p-value = .50$, one-tailed); second and most important, in all events, and in both the *private* and the *public* treatment of each event the majority of respondents identify the correct answer, thus failing to exhibit the conjunction fallacy.

Hence, not only were we unable to replicate the 85% violation rate reported in Kahneman and Tversky (1983), but the majority of our subjects responded correctly. The two distributions (ours and that of K&T) look markedly different, with a Chi-square test yielding a significance value of $p < .0001$ for both the *private* and the *public* treatments. Our reported violation rate is also significantly lower than that obtained by Charness et al. (2009) for single individuals without incentives ($p < .0001$), and it is comparable with that obtained by the same authors using single individuals with monetary incentives (33%), and trios without incentives (25.6%).

	Correct	Incorrect	Total	Response rate
account 1	7 (77%)	2 (23%)	9	21.4%
account 2	9 (81%)	2 (19%)	11	12%
account 3	12 (52%)	11(48%)	23	12%
account 4	22 (71%)	9 (29%)	31	8%
total	50 (67%)	24 (35%)	74	

Table 2: The experiment on Facebook: results of the *private* treatment

Moreover, it is also noteworthy that, while violation rates do not differ between the *private*

	Correct	Incorrect	Total	Response rate
account 1	12 (54.5%)	10 (45.5%)	22	68%
account 2	18 (54.5%)	15 (45.5%)	33	47%
account 3	62 (74%)	22 (26%)	84	27.6%
account 4	46 (68%)	22 (32%)	68	18.8%
total	138 (67%)	69 (33%)	207	

Table 3: The experiment on Facebook: results of the *public* treatment

and the *public* treatment, the response rate in the two treatments does differ significantly, being substantially higher in the public version (27%) than in the private version (10%, $p < .0001$, Chi-square test). Although part of the difference may be explained by taking into account that sending a private answer to the administrator was slightly more time-consuming than responding directly on the event's wall, we do not consider it very likely that this slight difference in response effort alone was responsible for the highly significant difference in participation rates. A more "social" explanation, therefore, must be envisaged.

Table 4 shows the results of the Mechanical Turk experiment. Let us first analyze the results without the bonus for the correct answer. The table shows that percentages in the *private* and the *public* version are almost identical (Fisher exact test; $p - value = .89$, two-tailed). In addition, roughly 70% of respondents fall prey to the conjunction fallacy, a datum which closely replicates previous conjunction fallacy experiments conducted on Mechanical Turk and on Internet boards (Paolacci et al. 2010), although the violation rate is, again, significantly different from that reported in Kaheman and Tversky (1983) ($p < .0001$).

	MT flat fee	MT bonus
Private	41/136 (30%)	49/103 (47.5%)
Public	36/123 (29%)	36/99 (36%)
Total	77/259 (29.7%)	85/202 (42%)

Table 4: The experiment on Mechanical Turk: percentages of correct responses

Introducing a bonus for the correct answer improves performance by more than 50% in the *private* treatment: in fact, with the bonus the percentage of correct responses increases from 30% to 47%, and the difference is highly significant ($p - value = .004$, Fisher exact test, one-tailed). On the contrary, no significant differences emerge as a function of bonus payment in the *public* treatment ($p - value = .16$, one-tailed). Comparing the *private* and the *public* treatments in the bonus condition does not yield significant differences ($p = .12$, two-tailed). Hence, these findings seem to corroborate our hypothesis about the influence of

conformity, in that the *public* treatment seems to mitigate the positive effect induced by the monetary incentive.

Overall, data from the Mechanical Turk experiment reveal a highly significant difference with respect to the Facebook experiment. Pooling public and private treatments from both environments, differences in correct responses in the two experiments are significant at the 1% level (Chi-square, $p = .0001$). Results in the Mechanical Turk *private* treatment with bonus, in which the rate of violation is the lowest, are still significantly higher than those in the Facebook experiment, both from the *private* treatment ($p - value = .009$) and from the *public* treatment ($p - value = .001$, Fisher's exact test).

Hence, while the data on Mechanical Turk substantially confirm both the presence of the conjunction fallacy and the power of monetary incentives to mitigate its magnitude, thus replicating previous results even in the presence of unusual conditions such as non-anonymity and public responses, the findings from Facebook appear anomalous. In order to shed more light on our data, we performed a runs test on the sequences of responses from the *public* treatments of both the Facebook and Mechanical Turk experiment. The test is aimed at verifying whether the probability of giving a certain answer depends on the specific position in the sequence of answers. If this is the case, participants were influenced by previous responses. Results of the runs test performed on the strings of responses indicate that one of the two sequences from the Mechanical Turk with bonus treatment is significantly different from a random sequence ($p = .05$, two-tailed)[3].

Figure 3 reports the cumulative frequency of responses conforming to the *majority* response for account 3 and 4 in the Facebook experiment and for the *public* treatments in Mechanical Turk (all data are averages). The graph reports how the cumulative frequency of *majority* answers changes as a function of the specific position of the participant in the sequence. Only the first 48 answers in the sequence were considered, in order to compare strings of equal length. The figure reveals no particular influence of the position in the sequence, although the conformity to the majority answer grows at a slower pace in the Mechanical Turk with bonus treatment with respect to the remaining two treatments.

Fig. 3 about here

---

[3]The two sequences in the treatment correspond to the two scenarios in which "Laura is a bank teller" was stated as first answer and as second answer respectively. Once a worker agreed to perform the task, he/she was automatically directed to one of the two different scenarios with equal probability.

14

5 Discussion

Several considerations emerge from our data: first, in our Facebook environment the majority of respondents did not commit the fallacy, whereas in Mechanical Turk the majority did indeed commit the fallacy, and the difference in the rate of violation when no bonus for the correct answer is involved is huge; moreover, in both Facebook and Mechanical Turk, we found no significant difference between the rate of violation when respondents answered in private form and when they answered publicly. The lack of any difference between private and public response modes suggests that participants in both environments were not paying excessive attention to previous responses in the *public* treatment, or, if they were, this attention did not translate automatically into herding behavior. It may well be that participants looked at a sample of previous responses, and, being unable to resolve their uncertainty (given lack of unanimity) opted to ignore the answers of others altogether and rely on their independent judgment. In addition, the lack of unanimity or of a striking majority (say, more than 90% of respondents choosing the same answer) may have reassured participants about the fact that neither answer, once provided, would be detrimental to their reputation. In fact, the famous studies on conformism by Asch and colleagues have shown that the tendency to conform to the majority judgment is closely linked to the size of the majority, and it decreases significantly when the majority is not unanimous with respect to when unanimity is present. As is well known, the most famous experiment by Asch involved answering a question whose correct answer was self-evident. In the Linda problem, the correct answer is far from obvious: therefore, in this case, answers were probably diversified enough to deactivate any form of herding or imitation. Even observational learning may have acted weakly in the absence of unanimity or of an overwhelming majority. Further experiments may test this conjecture by conducting a treatment with a certain number of predetermined "incorrect " answers posted by confederates to test whether unanimity triggers herd behavior, in line with the original experiment by Asch (1955). This further treatment should be run on both Facebook and Mechanical Turk, given that, at least when no monetary incentives were involved, in both environments participants did not seem to be influenced by previous answers[4].

Hence, given that the response mode does not lead to detectable differences in error rates, there is something else involved in the Facebook experiment that must account for the difference in results. In both environments, given the experimenter's lack of control, a

---

[4]Whether or not this treatment would involve deception is an open question, given that it would not imply any untruthful statement on the part of the experimenter, but only omission of information regarding some of the participants.

subject was virtually allowed to "google" the problem text and find the correct answer on the Internet, therefore this possibility must be ruled out as an explanation; moreover, subjects in the MT *bonus* condition had additional incentives to do so, but this did not prevent most answers from being incorrect. Hence, cheating on the part of subjects is not likely to be of any relevance in explaining our data.

Having ruled out alternative explicatory factors, the most important difference between our Facebook experiment and the Mechanical Turk experiment is the possibility for a subject to decide to participate or not *after* having read the question[5]. It is therefore plausible that self-selection into the task partly explains the difference between our Facebook and Mechanical Turk results, as well as those between Facebook and previous experiments on the conjunction fallacy. Subjects who decided to respond were probably more reflective, and hence less susceptible to committing the fallacy. Or, some subjects may have decided to write their answer only after some careful thinking and after having experienced the "Eureka" feeling associated with discovering the correct answer in these types of problems. However, self-selection alone does not explain why these more "rational" subjects were overrepresented in the sample of respondents, given the absence of monetary incentives; furthermore, it does not explain the higher participation rate observed in the *public* treatment compared to the *private* treatment. Hence, self-selection must be combined with the desire on the part of subjects to present themselves in a positive light by signaling their status, where status in this context refers essentially to being smarter than the average; participants in the *private* treatment could signal their status only to the account administrator, while participants in the *public* treatment could publicly signal it to all other participants as well. These two psychological phenomena combined might explain both the significant difference in the participation rate between the two treatments, and the very low error rate observed. Participants who had identified the correct solution were eager to publicly reveal it in order to acquire a good reputation among their Facebook peers. Participants who were more uncertain about the correct solution and were likewise concerned about their reputation simply refrained from participating. If reputation-induced self-selection was indeed the cause of our anomalous results, and if the mechanism is generalizable beyond the specific instance of

---

[5]A further difference between our two experiments concerns language and subject pool: the experiment on Facebook was conducted in Italian and the subject pool was predominantly Italian, while the experiment on MT was conducted in English and the subject pool was more international. Although in principle this difference may contribute to explain differences in results, we doubt it would constitute the main explanatory factor, given that the nature of the task is not sensitive to cultural differences, and given that the "Linda" problem has been extensively replicated in Italian, confirming results from previous experiments (see, e.g., Girotto 1994).

our study - further experiments should be devised to test the robustness of these results - then both the conjunction fallacy and other similar judgment and decision making biases might turn out to be less widespread in real world "social" contexts where participation is voluntary and where status and reputation among one's network of peers are highly valued. In addition, self-selection and concerns for reputation (or willingness to signal one's status) may similarly mitigate the magnitude of other collective phenomena likely to be fueled by social networks, making the results of traditional lab experiments on herd behavior, information cascades, imitation, observational learning, and the like less immediately transferable to real world environments where interaction is not anonymous.

It is noteworthy that in our experiment subjects invited to the same "event" were not necessarily Facebook friends, although they were all friends of the event administrator. It would be interesting to test the extent to which the effect which emerged in our experiment would vary in magnitude as a function of participants' social proximity. An easy proxy for social proximity in Facebook may be provided by a combination of two factors: whether or not two individuals are directly connected (they are Facebook "friends") and the number of contacts ("friends") they have in common. If the magnitude of judgment and decision making biases changed as a function of social proximity, the reputation-based explanation would gain further support. Other experiments should be conducted by varying features such as the type of bias under investigation, number of participants, presentation format, as well as participants' socio-demographic characteristics.

6 Conclusions

Our political views, our opinions on a variety of issues, our personal tastes, choices and personal experiences are being shared increasingly with others on Facebook and other popular social networks. Our field experiment on a simple and well-know judgment fallacy shows that such intense interaction may act counterintuitively by mitigating the effects of individual "irrationality", and suggests that concerns for status and reputation may in some instances mitigate the effects of herding behavior and prove more effective than financial incentives. Our data also fail to replicate previous results on the conjunction fallacy, which poses some doubts on the external validity of traditional lab experiments on individual judgment and decision making in which subjects' responses are typically provided in isolation. Given the pervasiveness of our exposure to the opinions of others subsequent to the advent of Internet 2.0 applications, the systematic investigation of peer effects through the tool offered by field experiments on social networks may offer useful and novel insights into the determinants of

many contemporary social phenomena that are relevant for economics.

References

[1] Asch, S.E. (1951). Effects of group pressure on the modification and distortion of judgments. In H. Guetzkow (Ed.), *Groups, leadership and men*, pp. 177190. Pittsburgh, PA: Carnegie Press.

[2] Asch, S.E. (1952). *Social psychology*. Englewood Cliffs,NJ: Prentice Hall.

[3] Asch, S.E. (1955). Opinions and social pressure. *Scientific American*, 193, 3535.

[4] Asch, S.E. (1956). Studies of independence and conformity. A minority of one against a unanimous majority. *Psychological Monographs*, 70(9), 170.

[5] Banerjee, A. V. (1992). A Simple Model of Herd Behavior. *Quarterly Journal of Economics*, 107(3): 797-817.

[6] Bikhchandani, S., Hirshleifer, D., Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change in Informational Cascades. *Journal of Political Economy*, 100(5): 992-1026.

[7] Bohren, J. A. (2010). Information-Processing Bias in Social Learning. Mimeo.

[8] Camerer, C. F. (1995). Individual decision making. In J. Kagel and A. Roth (eds), *Handbook of Experimental Economics*, Princeton: Princeton University Press.

[9] Charness, G., Karni, E., Levin, D. (2009). On The Conjunction Fallacy in Probability Judgment: New Experimental Evidence Regarding Linda, *Games and Economic Behavior* 68, March, 2010, 551-556.

[10] Devenow, A., Welch, I. (1996). Rational herding in financial economics in *European Economic Review* 40, Issues 3-5, 603-615.

[11] Egebark, J., Ekstrm, M. (2011). Like What You Like or Like What Others Like? Conformity and Peer Effects on Facebook. IFN Working Paper No. 886. Available at SSRN: http://ssrn.com/abstract=1948802 or http://dx.doi.org/10.2139/ssrn.1948802

[12] Eyster, E. and Rabin, M. (2010). Nave Herding in Rich-Information Settings. *American Economic Journal: Microeconomics*, 2(4): 221-43.

[13] Girotto, V. (1994). *Il ragionamento*. Bologna: Il Mulino.

[14] Goeree, J., Yariv, Y. (2010). Conformity in the Lab. Mimeo

[15] Hogg, M. A.; Turner, J. C. (1987). Social identity and conformity: A theory of referent informational influence. In Doise, W., Moscivici, S., *Current issues in European social psychology*. Cambridge: Cambridge University Press, 2: 139182.

[16] Isenberg, D.J. (1986). Group Polarization: A Critical Review and Meta-Analysis. *Journal of Personality and Social Psychology* 50 (6): 1141-1151.

[17] Janis, I. L. (1982). *Groupthink: Psychological Studies of Policy Decisions and Fiascoes*. Boston: Houghton Mifflin.

[18] Kahneman, D (1972). *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

[19] Kahneman, Tversky, Daniel, Amos (1972). Subjective probability: A judgment of representativeness. In Kahneman, D., Slovic, P., Tversky, A. *Judgment under uncertainty: Heuristics and biases*. Cambridge: Cambridge University Press.

[20] Levine, J. M. (Ed.). (2013). *Group processes*. New York: Psychology Press.

[21] Myers, D.G.; Lamm, H. (1975). The polarizing effect of group discussion. *American Scientist* 63 (3): 297303.

[22] Paolacci, G., Chandler, J., Ipeirotis, P. G. (2010). Running Experiments Using Amazon Mechanical Turk. *Judgment and Decision Making* 5 (5), 411-419.

[23] Turner, M. E.; Pratkanis, A. R. (1998). Twenty-five years of groupthink theory and research: lessons from the evaluation of a theory. *Organizational Behavior and Human Decision Processes* 73: 105115.

[24] Tversky, A. and Kahneman, D. (1983). Extensional versus Intuitive Reasoning: The Conjunction Fallacy in Probability Judgment. *Psychological Review* 90, 293-315.

Appendix: Instructions

This is the text of one the two events that were created on Facebook (the text used in the *private* version is between parentheses):

> Dear friends, I have created this 'event' because I am kindly asking your collaboration in a piece of a research that a friend of mine is conducting. What I am asking you is very simple and will take you no more than a couple of minutes. You should read the question reported below and provide your answer, that is the answer that you judge as more appropriate among the two answers available. Please, read, think and provide your answer without asking for other people's advice, as I am interested in your individual judgment.
>
> You are asked to provide your answer directly on the event's wall (after clicking on "Join") by writing "1" if you think that answer 1 is more appropriate, or "2" if you think that answer 2 is more appropriate. I ask you to answer ONLY in this way - by writing the number corresponding to the correct answer on the event's wall - and NOT to add any other comment. Any comment you write along with your answer will be deleted immediately (You are asked to send your answer in the form of a private message to me, which must contain only "1" if you think that answer 1 is more appropriate, or "2" if you think that answer 2 is more appropriate. I will not publish any of your answers, I will only consider your aggregate responses. Please do NOT write any comment on the event's wall. Any comment you may write will be deleted immediately.)
>
> Although a mandatory deadline for this event has not been set, I hope to be able to gather the majority of responses in the next few days. If you do not intend to participate, just click on "refuse" or simply do not do anything.
>
> Many thanks in advance for your collaboration! Best regards, (name of the administrator)

Then the following scenario was presented, with the alternatives listed either in this order or the reverse order.

> Laura is 31 years old, single, outspoken and very bright. She majored in philosophy. As a student, she was deeply concerned with issues of discrimination and social justice, and also participated in anti-nuclear demonstrations. Which of these two alternatives do you think is more probable?

1. Laura is a bank teller

2. Laura is a bank teller and is active in the feminist movement

Figure 1: Screenshot of one of the events in the Facebook *public* treatment; names have been obscured for privacy reasons.
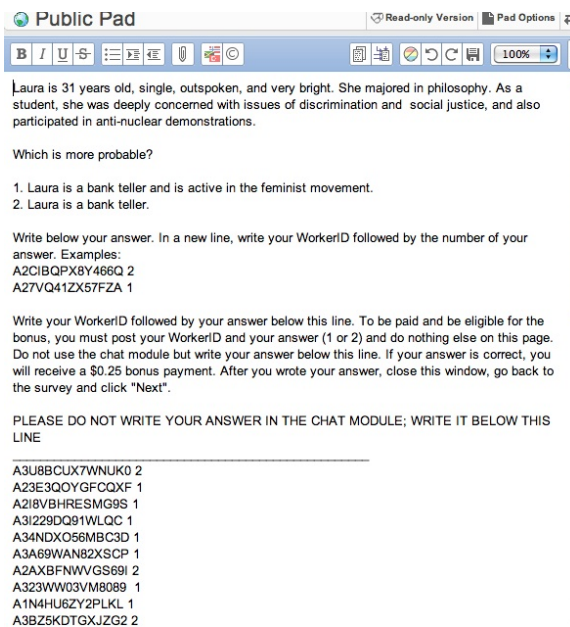

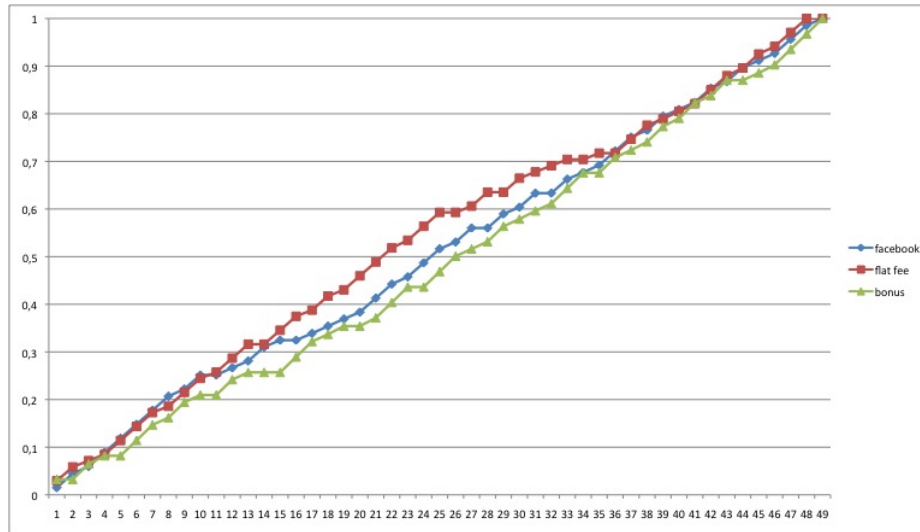Figure 2: Screenshot of instructions in the Mechanical Turk *public* treatment with bonus

Figure 3: Cumulative frequency of majority responses as a function of position in the sequence (averages)