RESEARCH INSTITUTE OF INDUSTRIAL ECONOMICS

Does Gender Diversity Promote Nonconformity?

Makan Amini, Mathias Ekström, Tore Ellingsen, Magnus Johansson and Fredrik Strömsten

This is an author-produced version of a paper accepted for publication in the Management Science*. The paper has been peer-reviewed but does not include the final proof corrections or pagination.* License information.

DOI/Link: https://doi.org/10.1287/mnsc.2015.2382

Reference: Amini, Makan, Mathias Ekström, Tore Ellingsen, Magnus Johansson and Fredrik Strömsten (2017). "Does Gender Diversity Promote Nonconformity?". *Management Science* 63(4), 1085–1096.

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

Does Gender Diversity Promote Non-Conformity?*

Makan Amini
Advent International

Mathias Ekström
NHH Norwegian School of Economics and Research Institute of Industrial Economics (IFN)

Tore Ellingsen
Stockholm School of Economics and NHH Norwegian School of Economics

Magnus Johannesson
Stockholm School of Economics

Fredrik Strömsten
McKinsey & Company

Final version: 8 October 2015

Abstract: Failure to express minority views may distort the behavior of company boards, committees, juries, and other decision-making bodies. Devising a new experimental procedure to measure such conformity in a judgment task, we compare the degree of conformity in groups with varying gender composition. Overall, our experiments offer little evidence that gender composition affects expression of minority views. A robust finding is that a subject's lack of ability predicts both a true propensity to accept others' judgment (informational social influence) and a propensity to agree despite private doubt (normative social influence). Thus, as an antidote to conformity in our experiments, high individual ability seems more effective than group diversity.

Key Words: Conformity, Gender Differences, Group Composition, Skill

JEL-codes: C90; D02; D71; D83; J16

* *Acknowledgments:* We thank David Cesarini, Per-Henrik Hedberg, Torun Lindholm, Carmit Segal, Roberto Weber and seminar participants at SSE and University of Zürich for helpful discussions. We thank five anonymous referees and the editors Stephan Meier and John List for insightful comments on previous drafts. Thanks also to The Choice Lab at NHH for administering the second experiment, with the assistance of Ranveig Falch, Reidun K. Malvik, Andreas Olden, Frøya Stavem, Kaja K. Toset. Finally, thanks to Abel Schumann for econometrics coding and to the Jan Wallander and Tom Hedelius Foundation and the Swedish Council for Working Life and Social Research for financial support.

1. Introduction

The Western world is experiencing a long wave of increased female participation in political and economic decision-making.[2] Besides gains in fairness and changes in priorities, it is widely believed that greater gender diversity can improve the quality of group decisions by increasing the amount of available information and decreasing the degree of conformity among group members.[3] As an illustration of the folk wisdom that diversity reduces conformity, consider the following statement from a recent U.K. Treasury Report to the House of Commons on the desirability of increasing female board representation in the U.K. financial industry:

> "Concern about the under representation of women on boards can be about business performance as much as fairness. There is a consensus that an effective challenge function within a board is required in financial institutions, and that diversity on boards can promote such challenge." House of Commons (2010, paragraph 13)

[2] For example, the fraction of females among the world's parliamentarians increased from about 3% after World War II to about 12% in 1985 and about 22% in 2015; in the Nordic countries, female MP's now constitute between 38% (Denmark) and 44% (Sweden) of all national parliamentarians (International Parliamentary Union, 2015). Similarly, female presence on company boards is growing in most countries. Although the fraction remains small worldwide, with females holding 17.3% of seats in the world's 200 largest companies in 2014 (CWDI, 2014), this is up from only 10.4% in 2004. The change is heterogeneous, and greatest in Europe. Already in 2008, women held more than 25% of the board seats of large companies in Finland and Sweden – and more than 44% in Norway (Egon Zehnder, 2008). Likewise, female membership of juries has grown dramatically; as late as World War II twenty U.S. states prohibited women from serving on juries (Fowler, 2005), whereas current juries have roughly equal gender representation, as most citizens are eligible for jury service.

[3] Of course, diversity could influence group processes and performance in a whole host of other ways too; see Williams and O'Reilly (1998) for a broad literature survey.

Indeed, both the report itself and the Government's formal reply to it explicitly argue that more diverse boards will be less prone to conformity.[4]

In order to cleanly identify the link between a group's gender composition and the conformity of its members, we conduct laboratory experiments. But before describing these experiments, let us briefly explain why we think laboratory evidence is a useful complement to more immediately relevant field evidence. Two important recent field studies, one observational and the other experimental, suggest that gender homogeneity could threaten performance and indicate that excessive conformity could be a central mechanism:

(i) Based on a large sample of US firms, Adams and Ferreira (2009) argue that the presence of women on company boards has a causal impact on board behavior, especially in the area of monitoring, where a poorly performing CEO is more likely to be replaced if there are more women on the board.[5] One possible reason for this relationship is that a diverse board creates an atmosphere in which participants feel free to express uncomfortable opinions. But another possibility is that most CEOs are males, and individual board members simply feel less loyalty to a CEO of the opposite sex. Yet another possibility is that female board members on average are more skillful or diligent than their male counterparts. Additionally, as Adams and Ferreira acknowledge, even with apparently plausible instrumental variables one cannot be sure of identifying causal impacts when groups' gender composition is endogenous.

---

[4] Electronic copies of both the report and the reply are available free, at http://www.publications.parliament.uk/pa/cm200910/cmselect/cmtreasy/482/482.pdf and http://www.official-documents.gov.uk/document/cm79/7900/7900.pdf respectively.

[5] This is not to say that increased female representation is profitable. Performance is positively related to the rate of female presence only in firms with weak shareholder protection. If anything, the average effect is negative. The legislation to improve the gender balance of company boards in Norway suggests that the sudden infusion of females had a variety of impacts on behavior and performance, but because many of the females might not have been board members absent the legislation the data do not allow a clean estimate of a gender-mix effect; see Ahern and Dittmar (2012).

(ii) In an extensive yearlong field experiment on business start-ups, Hoogedoorn, Oosterbeek, and van Praag (2013) demonstrate that teams with an equal gender mix perform better than all-male teams in terms of both sales and profits. The direction of causality is here unambiguous, since the experimenters control the start-ups' gender composition. However, the authors cannot pin down the exact reasons for the performance difference; the gender mix is uncorrelated with measures such as individual learning, complementarities in skill composition, as well as with measures of personal relations between group members.[6]

In this paper, we report evidence from two experiments based on a new design that we have developed to diagnose unwillingness to express a minority view. The first experiment suggests that subjects are significantly more willing to state their true belief about the correct solution to a problem in a mixed gender group than in a same-gender group. The evidence thus suggests that conformity to the behavior of similar others might be responsible for some of the links between gender diversity and behavior that have previously been observed in field settings (e.g., Apesteguia, Azmat, and Iriberri, 2012; Hoogedoorn, Oosterbeek, and van Praag, 2013, and the references therein).

Persuaded by arguments that our initial finding might be a fluke, and that journals should combat publication bias by requesting replications, we conducted a second experiment designed to fulfill desirable replication criteria (as outlined by, e.g., Maniadis, Tufano, and List, 2014). In order to corroborate our interpretations of the first study's findings, we also added a treatment with unbalanced gender composition, hypothesizing that an intermediate degree of group homogeneity would generate an intermediate level of conformity.

However, the second experiment offers no support for any of our hypotheses concerning the impact of gender diversity. There is no significant difference in conformity

[6] Needless to say, identification of precise mechanisms is typically even more difficult in non-experimental studies that consider the impact of gender composition on team performance, such as, e.g., Pelled, Eisenhardt, and Xin (1999), Lee and Fahr (2004), and Herring (2009).

between homogeneous and perfectly balanced groups. And if anything unbalanced groups display *more* conformity than any of the two others.

Let us now describe our design and findings a little more precisely. The experiments investigate individuals' behavior in groups of six. The experimental task requires cognitive ability, but occasionally one subject – without knowing it – faces a different task than the other subjects face. By studying the behavior of such minority-opinion subjects, we seek to determine both (i) *normative social influence,* i.e., the extent to which minority-opinion subjects falsely pretend to possess cognitive ability by publicly agreeing with others' judgment, and (ii) *informational social influence,* i.e., the extent to which minority-opinion subjects honestly, but in this case erroneously, change their mind when facing a large majority with a different opinion.[7] In particular, we are concerned with comparing the behavior of minority-opinion subjects in all-male and all-female groups to their behavior in groups with an even gender distribution

In Experiment I, the main finding is that minority subjects are significantly more susceptible to normative social influence in single-gender groups than in mixed-gender groups. That is, diversity promotes non-conformity. On the other hand, group composition plays no significant role for informational social influence. Thus, the study lends some credence to the hypothesis that gender-diversity affects group behavior through the encouragement of minority opinions. In Experiment II, the overall level of conformity is closely similar to the levels observed in Experiment I. However, the replication treatments indicate that there is no difference between homogeneous and perfectly diverse groups; if anything, conformity is greater in the perfectly diverse groups than in the homogeneous groups.

One finding that is robust across both studies is that conformity is greater among subjects whose judgment is less accurate. Lack of ability displayed in solitary problem-solving

[7] These definitions are due to Deutsch and Gerard (1955). The classic studies of normative social influence are due to Asch (1951,1956,1957), whereas Sherif (1937) is the classic study of informational social influence. For a comprehensive survey of early studies of conformity see Cialdini and Trost (1998).

significantly predicts both informational social influence and normative social influence. Altogether, our findings thus suggest that high quality of individual group members' judgment is a better antidote to conformity than is gender diversity.

Previous experimental research concerning the relationship between groups' gender composition and individuals' conformity consists of two main studies. One study finds that males' conformity is greater in single-gender groups, but that females' conformity is greater in mixed-gender groups (Tuddenham, MacBride, and Zahn, 1958). The other study finds, for both males and females, that conformity is greater in mixed-gender groups than in single-gender groups (Reitan and Shaw, 1964). However, the present experiment differs from these previous studies along several dimensions. We consider conformity in judgment tasks whereas the previous studies considered visual perception tasks; our subjects interact directly, whereas previous subjects were sitting in separate cubicles and communicated through light-switches; we explicitly separate normative from informational influence;[8] finally, our subjects are born fifty years later, in another country, and pursue an education intended to produce high-level professional decision makers.[9] For all these reasons, we thought that the present study could come out quite differently from the previous studies.

Our study contributes a new experimental design for the study of conformity. Compared to previous designs, it has one main advantage. It involves neither confederates, as in Asch (1951), nor costly equipment as in the alternative paradigm of Crutchfield (1955). Thus, the design is both cheaper to implement and less prone to the subject suspicion that has plagued earlier designs (Stricker, Messick, and Jackson, 1967).

[8] To be clear: many previous studies have disentangled informational and normative influence, for example by comparing responses given under different magnitude of social pressure while keeping information constant. However, these studies have not considered the role of the group's gender composition.

[9] Even for visual perception tasks, conformity is known to vary across cultures, across individuals, across situations, and over time (Bond and Smith, 1996; McIlveen and Gross, 1999; Wren, 1999).

Before presenting our experiment, let us briefly mention some relevant theoretical work. There is a large literature that discusses why people refrain from voicing minority views, or more generally avoid standing out from the crowd. For a textbook treatment, see Aronson (2004, Chapter 2); see also the literature review of Griskevicius et al (2006). Informational social influence is typically explained by the direct benefit of imitating others who may be better informed about or adapted to the relevant circumstances. Pioneering formal models of informational social influence, or social learning, are due to Banerjee (1992) and Bikhchandani, Hirshleifer, and Welch (1992). Normative social influence on the other hand is usually ascribed to the desire for affiliation or appreciation.[10] A formal model of normative influence in the context of compliance with social norms is due to Bernheim (1994).[11] Arguably, models of reputation for expertise (Scharfstein and Stein, 1990) better capture normative influence in the context of judgment – where agents, in communicating their views, desire to display appropriate skill to an outside audience. But even these models do not fully address the problem that we study here, since our subjects if anything are concerned with the impression that they make on other participants. It is an open problem how best to model normative social influence in this setting.[12]

[10] It is also possible that some people simply have an inherent desire or instinct to imitate. See Goeree and Yariv (2014) for supportive experimental evidence as well as for references to other recent experiments on social learning.

[11] It is also understood that the desire for affiliation can sometimes entail non-conformity. For example, when many males compete for the attention of a single female (or vice versa), standing out from the pack could be the best strategy.

[12] We conjecture that the best model of why people fail to report their honest beliefs will involve a departure from common knowledge of rationality. For example, there might be a positive probability that some people (are thought to) report naïvely, failing to take into account the information conveyed in previous reports. The fear of being thought of as being such a naïve and possibly ill-informed person could keep sophisticated people from speaking their mind. Incidentally, overweighting of private information relative to public signals has previously been found to

The literature suggests at least one plausible channel through which gender homogeneity could foster normative influence, namely through group cohesion. Several studies find that group cohesion is strengthened by demographic homogeneity in general (George, 1971) and by gender homogeneity in particular (Rosen, Bliese, Wright, and Gifford, 1999), and there is typically a positive effect of group cohesion on group norm compliance (Berkowitz, 1954; Lott and Lott, 1961; Wyer, 1966). As Aronson (2004, p 21) puts it: "A group is more efficient at inducing conformity if (1) it consists of experts, (2) the members (individually or collectively) are important to the individual, or (3) the members (individually or collectively) are comparable to the individual in some way." In view of group cohesion theory, the natural interpretation of our findings is that the other group members are seen as sufficiently important to motivate significant conformity, but that the similarities that matter to these Scandinavian business school students are unrelated to gender.

2. Design of Experiment

Subjects interacted in groups of six. The six subjects were seated in a circle, facing each other, but with screens preventing them from seeing others' tables. The instructor informed the subjects that the study would consider how their problem solving ability was affected by different external factors and that further explanation would be given after the completion of all experiments. In order to decrease the level of anonymity between group members, subjects were then asked to introduce themselves to the rest of the group with their name, age and grade. To reduce the risk of experimenter influence, subjects were also informed that their names and other identifiers would not be collected.

---

prevent informational cascades in experiments; see Goeree, Palfrey, and Rogers (2007). In those situations, there can be social gains from private shortcomings.

The experiment consisted of three consecutive problem-solving sessions or parts (hereafter denoted P1, P2 and P3). P1 and P3 contained six problems, while P2 contained thirty. The six problems in P1 and P3 were identical, though presented in a different order. These six problems also appeared among the thirty problems in P2.

All problems were in the form of Raven's progressive matrices: The subject is confronted with a three times three matrix with one of the nine squares missing and is asked to use the logical patterns in the matrix to pick one of five alternatives (A-E) to fit the missing square.

P1 and P3, with a time limit of three minutes each, were solved individually and silently with subjects using pens to mark their answers. In P2 subjects were asked to state their answers out loud. The instructor told the subjects when to move on to the next problem. Subjects were given twenty seconds to choose an answer and were then asked to give their answer, one after each other. The answers were given in a clockwise direction with the sequence shifting for each problem so that all subjects started (and ended) the answering rounds five times each. Subjects were informed that they should make an individual decision during the silent twenty seconds. They were also instructed to only state their answer loudly and clearly without any further explanation or justification.

All six subjects were given the same problems in each round of sessions P1 and P3, but in twelve of the thirty rounds in session P2 the last subject to answer was, unknowingly, presented with a different problem than the rest of the group. Thus, twice in P2 each subject was facing a problem with a correct answer that differed from that of the five other subjects. The two problems that differed from the rest of the group's problems were the same for all subjects. Henceforth, we refer to these two problems as *test problems*. Both test problems were included in P1 and P3 as well. Figure 1 graphically illustrates the design.

[Figure 1 about here]

In the twelve rounds comprising a test problem, the five other subjects' problems were chosen to be non-trivial yet easy enough to make it likely that all five would give the same answer, thus leaving the sixth subject with the decision of either conforming to the group's incorrect answer or deviate from the group by stating the correct answer.

By dividing the experiment into three different sessions, we sought to separate conformity due to normative influence from conformity due to informational influence. Consider a subject in session P2 answering last in a round following what appears to the subject to be a sequence of five identical (and wrong) answers. If the subject answers with the group in P2 but not in P1 and P3, the behavior is classified as being due to normative social influence. If the subject answers with the group in P2 and P3, but not in P1, the behavior is classified as being due to informational social influence. If the subject does not answer with the group in P2, the behavior is classified as non-conformist.

However, let us already now state one caveat regarding this classification scheme. We cannot exclude the possibility that some subjects fail to remember the group's (and therefore their own) answer in round 2 when given the same question in round 3. For such a subject, the behavior that we classify as normative influence could instead be due to informational influence. In other words, our measure of normative influence may be upward biased and the measure of informational influence correspondingly downward biased. We discuss this issue in more detail below.

Observe that we did not provide any material incentives, apart from a show-up fee, neither at the level of the group nor at the level of the individual. We discuss this design choice in the final section.

Two additional features are worth mentioning. First, in order to attenuate any suspicion that we were studying conformity, we minimized the gap between the correct and the incorrect

answers as follows. Prior to the experiment, we ran pilot rounds on another set of subjects. These pilots informed us which answers were second most likely to be chosen (after the correct one) for each problem. The correct answer to each test problem could thus be matched to the other subjects' second most likely answer, and their correct answers were likewise matched to the second most likely answer to the test problems. Second, we passively pretended that the experiment studied the impact of caffeine on behavior. Before subjects answered P1 they were asked to fill out a questionnaire regarding their intake of caffeinated beverages. Between P1 and P2, group members were given one glass of a well-known caffeinated energy drink. However, note that we did not tell the subjects that this was the purpose of our study.

Although we passively pretended that the study was about the impact of caffeine and that all subjects were solving the same problems, we argue that the design complies with established codes of conduct among experimental economists, for the following three reasons. First, we never lied to the subjects. Second, the honesty norm in behavioral economics does not imply telling subjects everything. Quite the opposite: a major concern in most experimental research is to minimize "experimenter demand effects." In order to prevent subjects from doing what they believe the experimenter desires, subjects are usually not informed about the purpose of the study, even though (or precisely because) such information would affect their behavior. Third, our design minimizes the threat to subject integrity, since any potentially shameful behavior on the part of subjects – such as yielding to normative influence – is hardly considered by the perpetrator to be detectable by the other subjects. Indeed, compared to most previous studies of normative social influence, we would argue that our methodology is preferable both with respect to deception and subject integrity. For example, as noted above, the methods used in the classical studies of Asch (1951) and Crutchfield (1955) involve outright deception, and many subjects also report suspicion that they are being deceived, and hence that their conformity is being noted by at least some of the other subjects (Stricker, Messick, and Jackson, 1967).

3. Results from Experiment I

The first experiment was conducted at the Stockholm School of Economics (hereafter SSE) over seven days at the end of April 2010. A total of 113 students, 56 male and 57 female, participated in the experiment. Subjects' age ranged from 18 to 29 years with a mean of 21.6. The subjects were randomly assigned to nineteen groups. Six groups contained only women, six groups contained only men, and seven groups contained an equal proportion of women and men. All groups consisted of six subjects except from one group, which consisted of five subjects due to a late dropout.

Although the subjects were similar with respect to age and education, most of them barely knew the other subjects in their group.[13]

As 113 subjects answered two test problems each, we received 226 answers to test problems. In our main specification, we only consider test answers following unanimous answers by the previous five group members. Of the total 226 test answers, 68 are discarded due to lack of unanimity of prior responders. An additional 16 answers are discarded because the tested subject answered incorrectly, and in accordance with the group's erroneous answer, in P1. In total, 84 individual responses were thus excluded, leaving 142 valid responses. Of the 113 subjects, 100 subjects gave at least one valid response. While the high rate of discarded test problems reduces the statistical power of our tests, it does not introduce a bias in favor of finding conformity. To the contrary, the regular presence of heterogeneous answers ought to make it easier for the subjects to speak their mind.

[13] In an interpersonal relations questionnaire distributed after the experiments, subjects were asked to rate, on a five-point scale, how well they knew each of the other group members. Out of the 560 ratings of dyadic inter-group relations, only about 12% received a score of 4 or 5, which corresponded to "I know him/her somewhat well", and "I know him/her very well". By far the most common score was 1, corresponding to "I do not know him/her at all". The average score was 1.8.

Of the 100 subjects with a valid response, 43 conformed at least once. In total, subjects conformed 52 times out of 142 opportunities, or 37 per cent of the time. Since subjects in session P1 gave correct answers for the two test problems at roughly the same rate, around 77%, we have no reason to expect differences in conformity across problems. We therefore combined the responses for the two test problems into one measure of conformity for each individual. This variable was assigned the value 0 for subjects that did not conform in any of the valid test questions, the value 0.5 for subjects conforming in one out of two valid test problems, and 1 for subjects conforming in all valid test problems.[14]

Our conformity measure is in turn decomposed into a measure of informational influence and a measure of normative influence depending on whether, in P3, a conforming subject stays with the majority's answer or changes answer.

To formally test whether group composition affects our conformity measures, we run OLS regressions of conformity on diversity, a dummy variable taking the value 1 in a mixed group and 0 in a homogeneous group.[15] We run these regression models both with and without controlling for the gender of the subject. To take account of any within-group correlation, we compute robust standard errors clustered at the group level (also called sandwich standard errors, Liang & Zeger, 1986). Because this test is only valid asymptotically, and tends to reject the null hypothesis too often when there are fewer than thirty clusters, we have also computed standard errors using the Wild bootstrap, as recommended by Cameron, Gelbach & Miller (2008). However, as this yields very similar standard errors and does not affect our conclusions about significance, we only report the conventionally computed sandwich standard errors.

[Figure 2 about here]

[14] In the Appendix, we also report a non-parametric specification (so that this effect is not assumed to be linear).

[15] In the Appendix, Table A1 and Figure A1, we divide the homogenous (same gender) groups into "all male" and "all female" groups.

[Table 1 about here]

Figure 2 depicts our main findings, and Table 1 summarizes the regression analysis. Normative social influence is more than three times larger in same-gender groups than in mixed-gender groups ($p<0.01$). The average normative influence rate in same-gender groups is also large in absolute terms. When in lone minority, a subject concords with the majority view only to return to the original answer in the subsequent private round, in more than a third of all cases. On the other hand, there is no significant effect of group composition on informational influence, ($p=0.75$), and persistent changes in behavior are also rare in absolute terms, occurring only in about one case out of ten.[16] Since normative influence is so much more frequent than informational influence, total conformity too is significantly smaller in mixed-gender groups, ($p<0.05$). Thus, even if one takes a maximally skeptical view of our distinction between normative and informational influence, the fact remains that conformity broadly construed is affected by the group's gender composition.

Of course, the extent to which a person is willing to comply with the majority might depend on the strength of the subject's confidence. To evaluate this possibility, we investigate whether conformity is related to performance in P1; it seems natural that subjects who make more mistakes will be less certain about their judgment. These regressions are also reported in Table 1. On average, for all groups, each additional error is associated with an 11 percentage point greater normative social influence, ($p<0.01$), and an 8 percentage point greater

[16] On a few additional occasions, subjects fail to conform in P2 but conform in P3 instead. This could be due to a process of delayed conviction that the group's answer might be correct. These responses (6 responses from 5 different subjects) were not coded as conformity in the analysis. Coding these responses as conformity increases the level of informational influence somewhat to 15.6% in same-gender groups and 15.4% in mixed-gender groups, but it does not change the reported conclusions in the paper. In Experiment II there is only one response of this type (where the subject fails to conform in Session 2 but conforms in Session 3 instead).

informational social influence, (p=0.05).[17] While the increase in informational influence is perhaps easiest to explain, the increase in normative influence is consistent with previous literature relating self-doubt to normative social influence (Tesser, Campbell, and Mickler, 1983).

The effect of skill is largely orthogonal to the impact of gender composition. Specifically, when we include a control variable for P1 behavior in an OLS regression, normative influence is still significantly larger in same-gender groups, p=0.01, as one would expect.[18] Thus, we conclude that normative influence is robustly smaller in mixed groups.

4. Results from Experiment II

The second experiment was conducted at the Norwegian School of Economics (NSE) in March 2014. NSE is the Norwegian counterpart to SSE, and the student bodies have very similar characteristics. We chose this location because we could think of no other place that would be a better match for a replication study.[19]

The experimental procedure was the same as in Experiment 1, with the only difference that two questions were added towards the end of P2.[20]

---

[17] In principle, this estimate could be biased by differences in ability diversity across groups, but apparently it is not; see the robustness analysis in the online Appendix. There, we also report interactions between ability and group to see whether the impact of ability differs depending on circumstances.

[18] There is no reason to think that subjects would differ in the extent to which they make mistakes in P1 on the basis of the group's gender composition.

[19] NSE was established as a blueprint of SSE. The programs are almost identical. Like SSE, NSE attracts the country's best students and prepares them for a business career. Since we wanted to avoid the possibility that new subjects had participated in or heard about the first experiment, NSE was a preferable location to SSE.

[20] Each subject then received the two questions that the group received when they received their test questions, and the correct answers to these questions were labeled in order to yield the same correct response for all subjects for

In total, 252 subjects participated in Experiment II, 127 male and 125 female. Subjects' age ranged from 19 to 37 years, with an average of 23.0. There were 42 groups. As in Experiment I, we composed groups with only men, groups with only women and groups with an equal proportion of men and women. In addition we added minority groups with either one man and five women or one woman and five men. We designed the experiment to have an equal number of same gender groups (with an equal number of male and female groups), mixed groups, and minority groups (with an equal number of male and female minority groups). The sample size was chosen to have at least 80% statistical power to replicate the results of Experiment I at the 5% significance level. Using the observed means and standard deviations of Experiment I, 64 subjects in each treatment are needed to have 80% statistical power to replicate a difference in conformity between the mixed and same gender groups of the same magnitude as in Experiment I. To merely replicate the difference in normative social influence from Experiment I, 37 subjects are needed in each group to have 80% statistical power.

The experiment was scheduled at 7 different times in a single day (8, 9, 10, 11, 12, 13, and 14). At each time we would simultaneously run six groups; one male group, one female group, two mixed groups, one male minority group, and one female minority group. Invited subjects could sign up for a specific time, and were then randomly allocated to the 6 groups at that time (and the 6 experimenters were randomly allocated to groups at each time). Due to logistical issues, we deviated slightly from the planned design: We only ran five groups at 8 as too few subjects showed up at that time. The remaining group (a male minority group) was run at 15 based on reserves not utilized at the other times. As subjects were not randomly allocated to this group (conditional on the time of the experiment), we exclude the group from the main analysis,

---

the two last questions in P2. In the (unlikely) case that a subject asked whether everyone faced the same set problems in P2, we could thus respond that everyone would solve the same 32 problems. (One subject raised this question during the experiment).

but whether this group is included or not is unimportant for our findings. In one mixed group at 10 there was, due to naming inconsistencies (names incorrectly classified into gender), 4 men and 2 women. This group was also excluded from the analysis. In one supposed male group (at 8) there was, for the same reason, one woman, and in one female group (at 11) there was one man. These two groups are included in the analysis as female and male minority groups (rather than the intended male and female groups). To control for the resulting slight imbalance in the fraction of subjects across treatments at each time, we include six dummy variables for the time of the data collection. In the regressions we also cluster standard errors on the 40 groups included in the analyses.

The answers were coded in the same way as for Experiment I. A subject answering with the group in P2, but not in P1 was coded as conformity. If the subject gave the same answer also in P3 it was coded as informational social influence, and if the subject gave a different answer in P3 it was coded as normative social influence.

For the 240 subjects that participated (after excluding the two groups above), we received 480 answers to the two test problems. Of these test answers, 162 were discarded due to a lack of unanimity in the answers of prior responders in the round. An additional 34 answers were discarded because the tested subject answered incorrectly, and in accordance with the group's erroneous answer, in P1. Finally, 4 answers were discarded because the group made a unanimous incorrect answer that was the same as the answer of the subject in P1. In total 200 responses were thus excluded, leaving 280 valid responses. Of the 240 subjects, 191 gave at least one valid response and were included in the analyses (64 subjects in the mixed treatments; 58 subjects in the same gender treatments, and 69 subjects in the minority treatments).

Of the 191 subjects with a valid response, 84 conformed at least once. In total, subjects conformed 100 times out of 280 opportunities, or 36 percent of the time (which is almost exactly the same rate as for Experiment I). As for Experiment I, we combined the responses to the two

test problems into one measure of conformity for each individual. This variable was assigned the value 0 for subjects that did not conform in any of the valid test questions, the value 0.5 for subjects conforming in one out of two valid test problems, and 1 for subjects conforming in all valid test problems.

To formally test whether group composition affects our conformity measures, we run OLS regressions of conformity on diversity, with two dummy variables for mixed groups and minority groups (with same gender groups as the baseline category). To take account of any within-group correlation, we compute robust standard errors clustered at the group level. We control for the time of the experimental session in all regressions as discussed above (with 6 dummy variables, with 8 am as the baseline category).[21] As in study 1, we add the gender of the subject in a second regression specification to test whether gender affects the level of conformity. In a third regression we also, as in Study 1, add the number of correct answers in P1 to test how skill affects the level of conformity.[22]

[Figure 3 about here]

[Table 2 about here]

Figure 3 depicts the main results, and Table 2 summarizes the regression analysis. While the average levels of conformity are similar to Study 1, it is clear from Figure 3 that the impact of gender composition is very different. Now, conformity is smaller in the same-gender groups than in the mixed groups. Although the difference is not statistically significant, it goes in the "wrong"

[21] Conformity is significantly higher at the sessions conducted at 11 am, reflecting either the selection of subjects that signed up for this time or the fact that this time slot comes immediately before lunch, when energy levels may be depleted.

[22] In the Appendix, Table A2 and Figure A2, we divide the same gender groups into "all male" and "all female" groups and the minority groups into male and female minority groups.

direction, and we can only admit a complete failure to replicate Study 1's finding that more gender diversity is associated with less conformity.

A finding that does replicate is the relationship between skill and conformity. As in Study 1, subjects doing worse in the first six solo-rounds of privately solved problems are more susceptible to both normative and informational influence. Even the effect sizes are quite similar across the two studies. Evaluated at the average skill level, a one standard-deviation decrease in skill entails about 50% greater conformity.

The new treatment, with 5 subjects of one gender and 1 of the other, offered a surprising finding. Subjects in these unbalanced groups conform more than any of the other two groups, with the difference being statistically significant at the 10% level in comparison with the homogeneous groups. The greatest difference is between the all male groups, which have the lowest level of overall conformity, at about 20%, and the mostly male groups, which have an overall conformity level of about 50%. At first sight, this is highly counterintuitive. We had expected the unbalanced groups to display an intermediate level of conformity – more than the perfectly diverse groups, less than the homogeneous groups. An alternative theory, promoted by Griskevicius et al (2006) is that male strategies for mate attraction involve non-conformity whereas female strategies involve conformity. According to this theory, males ought to conform *less* in the presence of a female. However, in our design there is a confounding factor that does not allow a clean test of that theory: When a male in a mostly male group decides whether or not to conform, he also decides whether or not to agree with the sole female group member. Presumably, the best strategy from a mating perspective is then to agree. Indeed, a possible interpretation of the high degree of conformity in the unbalanced groups is that subjects avoid disagreeing with scarce members of the opposite sex. To investigate this possibility, we compare the behavior of minority members to the behavior of majority members of the unbalanced groups. We find that on average the minority members are about a third less likely to conform

than are the majority members, and the difference is even larger for normative influence.[23] But since there are only 15 minority groups, the estimate is too imprecise to be statistically significant. Altogether, we think that these findings from the unbalanced treatments are intriguing, but until they have been replicated they should all be viewed as highly speculative, whether significant or not.

5. Discussion

Both our experiments document significant conformity among gifted young Scandinavian business students. In particular, many of our subjects seem reluctant to voice a judgment that differs from that of a large majority. That is, normative social influence, rather than informational social influence, accounts for most of the observed conformity. Both types of influence are more prevalent among less skilled subjects, suggesting that high skill not only increases the probability that individuals privately identify correct solutions, but also makes them more resistant to normative social influence. But the main purpose of our study was neither to assess the magnitude of social influence nor to document how conformity varies with personal traits; there are already large literatures on these issues. Instead, our purpose was to investigate how conformity depends on the gender composition of the group. Here, the findings are more ambiguous.

Our first study suggests that gender diversity promotes non-conformity. However, the second study offers no such evidence; if anything, diversity promotes conformity. While we cannot rule out the possibility that the different findings are due to a difference in subject pools,

[23] Specifically, the minority member is 18.5 percentage points less likely to conform (p=0.265) and has 16.4 percentage points lower compliance (p=0.187). Recall that average conformity is about 0.44 and average normative influence is about 0.32 in the mixed groups.

that explanation strikes us as implausible. In view of the pooled effects, which are close to zero, a more likely explanation is that the first finding is a false positive. And even if there is in fact a difference between the student populations at these two business schools, it is difficult to escape the conclusion that the initial finding concerning the role of gender composition lacks robustness. The outcome illustrates yet again the value of replication.

The second study additionally suggests that conformity could be extra large in groups with unbalanced gender composition. However, as this finding was unexpected and is not replicated, this finding is at most a call for further research.

As always, one should be cautious in making far-reaching inferences from laboratory studies. We think it is hard to tell whether conformity is likely to be greater or smaller in relevant field settings. Some factors speak in favor of more conformity in the field: Our subjects' low connectedness limits group cohesion, which tends to raise conformity (see references in Introduction). Conformity also tends to be greater when group membership is more important (Guimond, 1990; Wolf, 1985). Since membership was hardly at stake in our temporary group of loosely connected subjects, field subjects may be more prone to conform. Other factors speak in favor of less conformity in the field: In particular, normative influence tends to decrease as the group's task becomes more important, at least if the task is also relatively easy (Baron, Vandello, and Brunsmann, 1996; Levine, Higgins, and Choi, 2000).[24]

There is one feature of our design that behavioral economists might object to, namely, the lack of material incentives. Material stakes are often large in the field settings that the experiments seek to shed light on. Therefore behavioral economists usually advocate the use of monetary incentives as a means to enhancing the external validity of experimental findings

[24] However, Baron, Vandello, and Brunsmann (1996) find that normative social influence *increases* with task importance if the task is sufficiently difficult. One reason might be that people "choke" under pressure; see, e.g., Ariely et al. (2009).

(Camerer, 2003, 38-40). Especially when it comes to judgment and decision-making tasks, there is evidence that subjects make fewer mistakes when stakes are larger (Camerer and Hogarth, 1999; Hertwig and Ortmann, 2001). On the other hand, there is also evidence that a small monetary incentive entails behavior that is further from the behavior under large incentives than is the behavior under no monetary incentive – perhaps because a small incentive legitimizes careless behavior. Hence an alternative view is that experimenters should either pay "enough" or not at all (Gneezy and Rustichini, 2000; Gneezy, Meier, and Rey-Biel, 2011). Ultimately, we saw two reasons for avoiding strong incentives. First, as we wanted to study the impact of gender composition on normative social influence, we needed the overall level of normative influence not to be too low. In order to identify a difference in mean proportions, the average proportion should ideally be around ½. We thought it unlikely that the average proportion would exceed this level. If sizeable monetary incentives were to significantly reduce the normative social influence, as suggested by the findings of Baron, Vandello, and Brunsmann (1996) and Levine, Higgins, and Choi (2000), the impact of gender composition would be correspondingly harder to detect. The fact that we fail to find a gender composition effect, *despite* a relatively high overall rate of normative social influence of about thirty percent, suggests that we would also have been unlikely to find an effect in the presence of monetary incentives.[25] Second, it is delicate to implement a design that gives individual incentives without making the situation unduly competitive.[26]

[25] If we would have found strong gender composition effects, there might have been a concern that these would vanish in the presence of strong monetary incentives. However, we see less reason to expect an interaction between gender composition effects and the strength of monetary incentives than between conformity and monetary incentive strength.

[26] One possibility would be to randomly draw an answer and reward the whole group on the basis of that answer.

A natural next step, which would at least partly address the question of external validity, is to run field experiments using features of our design.[27] As the design relies neither on confederates nor on costly equipment, it is portable to field settings. Furthermore, because each group member receives an own set of instructions, the experimental task does not have to be of the form that we used here, but can be adapted to the context at hand.

Let us end by returning to the big question that motivated our research: Why does gender composition affect the performance of groups – as it seems to do in many cases? We still do not know. But compared to our priors we are now less inclined to ascribe beneficial effects of gender diversity to a decrease in normative social influence.

[27] To date, there are few field experiments on conformity, but see Egebark and Ekström (2013) for a recent field experiment to investigate conformity on Facebook, a social network on the internet.

References

Adams, R., Ferreira, D. (2009). Women in the boardroom and their impact on governance and performance. *Journal of Financial Economics* 94(2), 291-309.

Ahern, K., Dittmar, A. (2012). The changing of the boards: The impact on firm valuation of mandated female board representation. *Quarterly Journal of Economics* 127(1), 137-197.

Apesteguia, J., Azmat, G., Iriberri, N. (2012). The impact of gender composition on team performance and decision-making: Evidence from the field. *Management Science* 58(1), 78-93.

Ariely, D., Gneezy, U., Loewenstein, G., Mazar, N. (2009). Large stakes and big mistakes. *Review of Economic Studies* 76, 451-469.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgment. In H. Guetzkow (Ed.), *Groups, Leadership, and Men* (pp. 76-92). Pittsburgh, PA: Carnegie Press.

Asch, S. E. (1956). Studies of independence and conformity: A minority of one against a unanimous majority. *Psychological Monographs* 70(9), (Whole No. 416).

Asch, S. E. (1957). An experimental investigation of group influence. In Walter Reed Army Institute of Research, *Symposium on preventive and social psychiatry* (pp. 15-17). Washington, DC: U.S. Government Printing Office.

Banerjee, A.V. (1992). A simple model of herd behavior. *Quarterly Journal of Economics* 197, 797-817.

Baron, R.S., Vandello, J.A., Brunsmann, B. (1996). The forgotten variable in conformity research: Impact of task importance on social influence. *Journal of Personality and Social Psychology* 71, 915-927.

Berkowitz, L. (1954). Group standards, cohesiveness, and productivity. *Human Relations* 7, 509-519.

Bernheim, B.D. (1994). A theory of conformity. *Journal of Political Economy* 102, 841-877.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascade. *Journal of Political Economy* 100, 992-1026.

Bond, R., Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin* 119, 111-137.

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

Camerer, C.F. (2003). *Behavioral Game Theory.* Princeton: Princeton University Press.

Camerer, C.F., Hogarth, R.M. (1999). The effects of financial incentives in economic experiments: A review and capital-labor-production framework. *Journal of Risk and Uncertainty* 18, 7-42.

Cialdini, R. B., Trost, M. (1998) Social influence: Social norms, conformity, and compliance. In Gilbert, D. T., Fiske, S. T., Lindzey, G. (Eds.), *The Handbook of Social Psychology* (Vol. 2, 4th ed., pp. 151-192). New York: McGraw-Hill.

Crutchfield, R. S. (1955). Conformity and character. *American Psychologist* 10*,* 191-198.

CWDI (2014). *Corporate Women Directors International 2014 Report: Women Board Directors of Fortune Global 200*. https://enterprisingwomen.com/the-connector/2014-cwdi-report-women-board-directors-of-fortune-global-200-2004-2014.html [downloaded 4 March 2015].

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *Journal of Abnormal and Social Psychology* 51*,* 629-636.

Egebark, J., Ekström, M. (2013). Is liking contagious? Manuscript, Stockholm University.

Egon Zehnder (2008). *Women on European Boards*. http://www.egonzehnder.com/global/thoughtleadership/hottopic/id/78402633/article/id/11900485 [downloaded 2 September 2010].

Fowler, L. (2005). Gender and jury deliberations: The contributions of social science. *William & Mary Journal of Women and the Law* 12(1), Article 2.

George, A. L. (1971). Primary groups, organizations, and military performance. In Little, R. W., (Ed.), *Handbook of Military Institutions* (pp. 293-318). Beverly Hills, CA: Sage.

Gneezy, U., Meier, S., Rey-Biel, P. (2011). When and why incentives (don't) work to modify behavior. *Journal of Economic Perspectives* 25(4), 191-210.

Gneezy, U., Rustichini, A. (2000). Pay enough or don't pay at all. *Quarterly Journal of Economics* 115(3), 791-810.

Goeree, J.K., Palfrey, T.R., Rogers, B.W., McKelvey, R.D. (2007). Self-correcting informational cascades. *Review of Economic Studies* 74, 733-762.

Goeree, J.K., Yariv, L. (2014). Conformity in the lab. Manuscript, CalTech.

Griskevicius, V., Goldstein, N.J., Mortensen, C.R., Cialdini, R.B., Kenrick, D.T. (2006) Going along versus going alone: When fundamental motives facilitate strategic (non)conformity. *Journal of Personality and Social Psychology* 91, 281-294.

Guimond, S. (1990). Attitude change during college: Normative or informational social influence? *Social Psychology of Education* 2, 237-261.

Herring, C. (2009). Does diversity pay? Race gender, and the business case for diversity. *American Sociological Review* 74(2), 208-224.

Hertwig, R., Ortmann, A. (2001). Experimental practices in economics: A methodological challenge for psychologists? *Behavioral and Brain Sciences* 24, 383-403.

Hoogedorn, S., Oosterbeek, H., van Praag, M. (2013). The impact of gender diversity on team performance. *Management Science* 59(7), 1514-1528.

House of Commons (2010): *Treasury Tenth Report – Women in the City*, London: British Parliament.

International Parliamentary Union (2015). *Women in National Parliaments: Situation as of 1st January 2015*. http://www.ipu.org/wmn-e/classif.htm [downloaded 4 March, 2015].

Levine, J. M., Higgins, E. T., Choi, H.-S. (2000). Development of strategic norms in groups. *Organizational Behavior and Human Decision Processes* 82, 88-101.

Lee, C., Fahr, J. (2004). Joint effects of group efficacy and gender diversity on group cohesion and performance. *Applied Psychology* 53(1), 136-154.

Liang, K.-Y., Zeger, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13-22.

Lott, A. J., Lott, B. E. (1961). Group cohesiveness, communication level, and conformity. *Journal of Abnormal and Social Psychology* 62, 408-412.

Maniadis, Z., Tufano, F., List, J.A. (2014). One swallow does not make a summer: New evidence on anchoring effects. *American Economic Review* 104(1), 277-290.

Murray, D.R., Trudeau, R., Schaller, M. (2011). On the origins of cultural differences in conformity: Four tests of the pathogen prevalence hypothesis. *Personality and Social Psychology Bulletin* 37, 318-329.

McIlveen, R., Gross, R. (1999), *Social Influence*. London: Hodder & Stoughton.

Pelled, L.H., Eisenhardt, K.M., Xin, K.R. (1999). Exploring the black box: An analysis of work group diversity, conflict, and performance. *Administrative Science Quarterly* 44, 1-28.

Reitan, H. T., Shaw, M. E. (1964). Group membership, sex composition of the group, and conformity behavior. *Journal of Social Psychology* 64, 45-51.

Rosen, L. N., Bliese, P. D., Wright, K. A., Gifford, R. K. (1999) Gender composition and group cohesion in U.S. Army units: A comparison across five studies. *Armed forces & Society* 25, 365-386.

Sherif, M. (1936). *The Psychology of Social Norms*. New York: Harper.

Stricker, L. J., Messick, S., Jackson, D. N. (1967) Suspicion of deception: Implications for conformity research. *Journal of Personality and Social Psychology 5,* 379-389.

Tuddenham, R. D., MacBride, P., Zahn, V. (1958). The influence of the sex composition of the group upon yielding to a distorted norm. *Journal of Psychology* 46, 243-251.

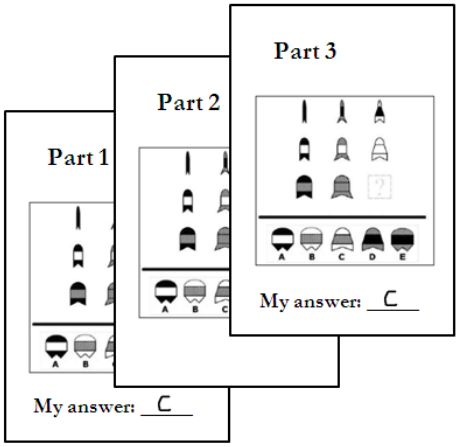
Williams, K.Y., O'Reilly, III, C.A. (1998). Diversity in organizations: A review of 40 years of research. *Research in Organizational Behavior* 20, 77-140.

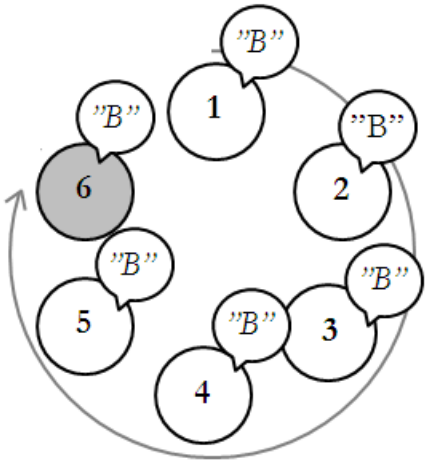
Wolf, S. (1985). Manifest and latent influence of majorities and minorities. *Journal of Personality and Social Psychology* 48, 899-908.

Wren, K. (1999) *Social Influences*. Florence, KY : Taylor and Francis/Routledge.

Wyer, R. S., Jr. (1966). Effects on incentive to perform well, group attraction and group acceptance on conformity in a judgmental task. *Journal of Personality and Social Psychology* 6, 255-263.

Fig. 1. Description of test problems and procedures





Two "test problems" measure subjects' conformity. The test problem shown above appears in all three sessions (P1, P2 and P3) and has the correct answer "C". In P1 and P3 the subjects solve six problems privately using a pen.
In this example, the subject correctly answers C in both P1 and P3

In P2, subjects solve thirty problems, including the six problems that appear in P1 and P3. Subjects report their solutions publicly, in a rotating sequence. The starting role also rotates. When facing a test problem, the subject is the last to report. The tested subject does not know that the first five subjects face a different problem with the correct answer B. In the example, the tested subject conforms and answers "B" as well. This answer would be recorded as normative social influence since the subject reverted to "C" in P3.

Fig. 2. (Experiment I): Means of the two conformity measures in same-gender groups and mixed-gender groups respectively. Total conformity is the sum of normative and informational social influence. Error bars represent standard errors of the mean.

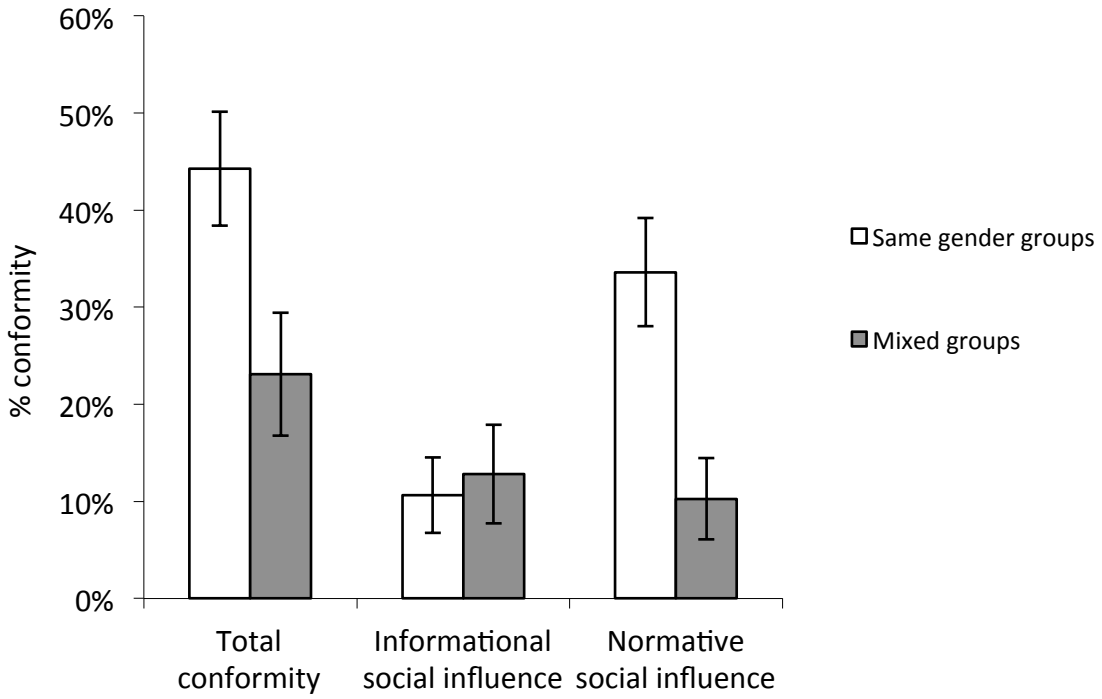


Fig. 3 (Experiment II): Means of the two conformity measures in same gender groups and mixed gender groups. Total conformity is the sum of normative and informational social influence. Error bars represent standard errors of the mean.

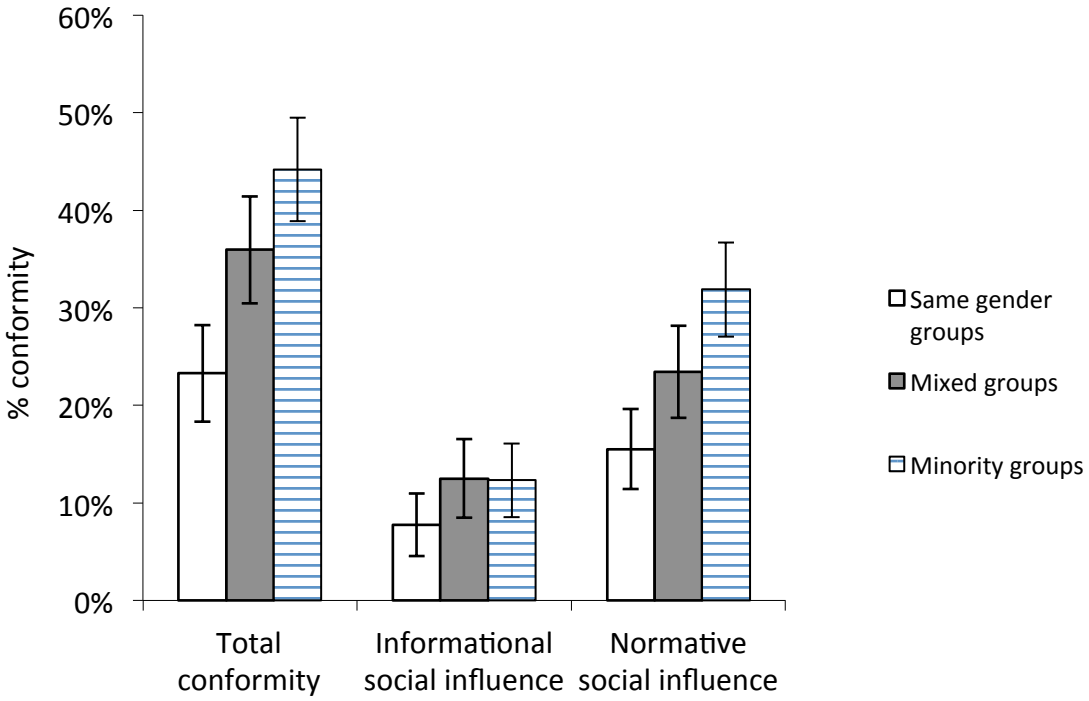


Table 1. OLS regressions on conformity in Experiment I (conducted at Stockholm School of Economics). Clustered standard errors (on groups) in parentheses and p-values in squared brackets.

	Total Conformity			Informational Social Influence			Normative Social Influence		
Constant	0.443 (0.047) [<0.001]	0.460 (0.074) [<0.001]	1.318 (0.173) [<0.001]	0.107 (0.030) [0.003]	0.084 (0.041) [0.056]	0.440 (0.196) [0.038]	0.336 (0.057) [<0.001]	0.376 (0.078) [<0.001]	0.878 (0.172) [<0.001]
Mixed group[1]	-0.212 (0.084) [0.021]	-0.212 (0.084) [0.022]	-0.147 (0.075) [0.066]	0.022 (0.067) [0.752]	0.021 (0.067) [0.754]	0.048 (0.070) [0.496]	-0.234 (0.070) [0.004]	-0.233 (0.071) [0.004]	-0.195 (0.068) [0.010]
Male		-0.033 (0.067) [0.626]	0.052 (0.074) [0.492]		0.045 (0.041) [0.279]	0.081 (0.053) [0.146]		-0.079 (0.076) [0.316]	-0.029 (0.067) [0.673]
Correct answers in part 1			-0.186 (0.037) [<0.001]			-0.077 (0.037) [0.053]			-0.109 (0.034) [0.005]
n	100	100	100	100	100	100	100	100	100
R-squared	0.055	0.056	0.241	0.001	0.007	0.073	0.085	0.095	0.176
F-value [p-value]	6.38 [0.021]	3.42 [0.055]	15.72 [<0.001]	0.10 [0.752]	0.63 [0.542]	1.71 [0.200]	10.99 [0.004]	5.75 [0.012]	12.73 [<0.001]

[1] The baseline category is same gender group.

Table 2. OLS regressions on conformity in Experiment II (conducted at the Norwegian School of Economics and Business Administration). Clustered standard errors (on groups) in parentheses and p-values in squared brackets.[1]

	Total Conformity			Informational Social Influence			Normative Social Influence		
Constant	0.194 (0.059) [0.002]	0.181 (0.068) [0.011]	0.868 (0.154) [<0.001]	0.022 (0.045) [0.630]	0.038 (0.050) [0.452]	0.312 (0.134) [0.025]	0.172 (0.052) [0.002]	0.143 (0.053) [0.011]	0.556 (0.148) [0.001]
Mixed group[2] (β1)	0.112 (0.069) [0.114]	0.113 (0.070) [0.115]	0.088 (0.067) [0.195]	0.045 (0.058) [0.441]	0.044 (0.057) [0.444]	0.035 (0.057) [0.550]	0.067 (0.056) [0.238]	0.068 (0.058) [0.244]	0.054 (0.057) [0.350]
Minority group[3] (β2)	0.160 (0.085) [0.068]	0.163 (0.085) [0.064]	0.143 (0.078) [0.075]	0.022 (0.060) [0.708]	0.019 (0.062) [0.762]	0.011 (0.060) [0.858]	0.137 (0.069) [0.052]	0.144 (0.070) [0.047]	0.132 (0.068) [0.060]
Male		0.022 (0.079) [0.782]	0.030 (0.071) [0.676]		-0.030 (0.060) [0.620]	-0.027 (0.057) [0.645]		0.052 (0.058) [0.375]	0.057 (0.055) [0.310]
Correct answers in part 1			-0.139 (0.025) [<0.001]			-0.055 (0.025) [0.035]			-0.083 (0.028) [0.004]
p-value β1=β2	0.536	0.522	0.421	0.711	0.696	0.705	0.234	0.209	0.164
n	191	191	191	191	191	191	191	191	191
R-squared	0.114	0.114	0.231	0.079	0.082	0.121	0.053	0.058	0.114
F-value [p-value]	2.45 [0.030]	2.25 [0.039]	10.45 [<0.001]	3.77 [0.002]	3.13 [0.006]	2.64 [0.015]	2.25 [0.044]	1.75 [0.110]	3.64 [0.002]

[1] All regressions also include six dummy variables for the time of the sessions (subjects signed up for one of 7 different time slots during a day and were randomized to treatments at each time slot), whose coefficients are not shown in the table.

[2] The baseline category is same gender group.

[3] Groups with one man or one woman.

For online publication

Appendix 1: Additional figures, tables, and robustness checks

Fig. A1 (Experiment I). Means of the two conformity measures; same gender groups separated into male and female groups. Total conformity is the sum of normative and informational social influence. Error bars represent standard errors of the mean.

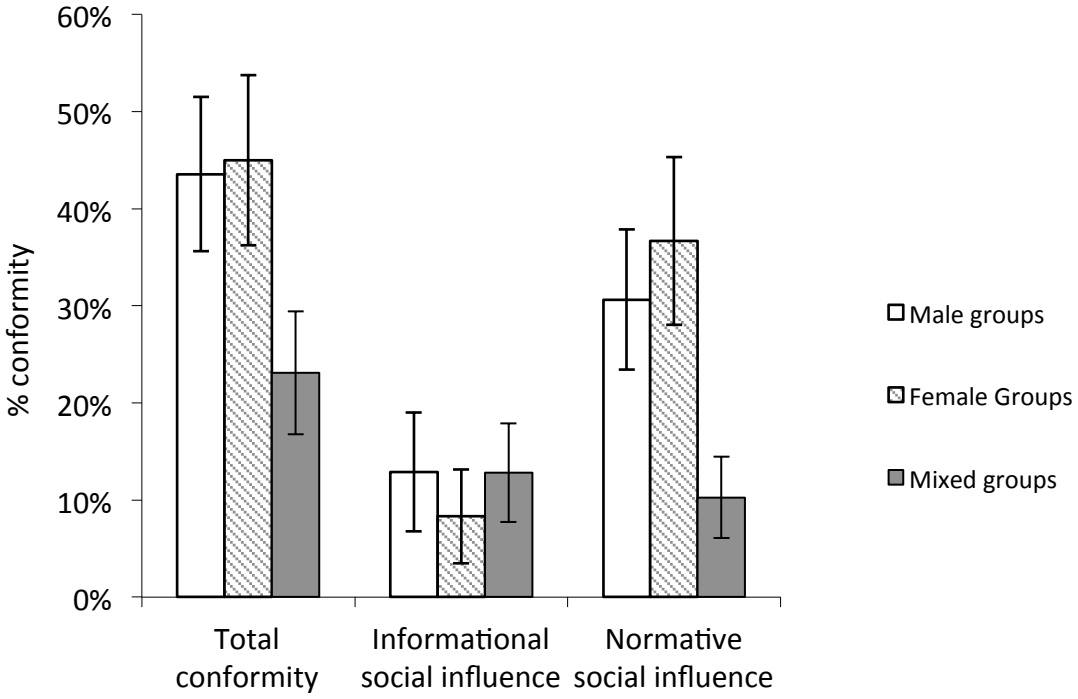


Fig. A2 (Experiment II). Means of the two conformity measures; same gender groups separated into male and female groups and the minority groups separated into male and female minority groups. Total conformity is the sum of normative and informational social influence. Error bars represent standard errors of the mean.

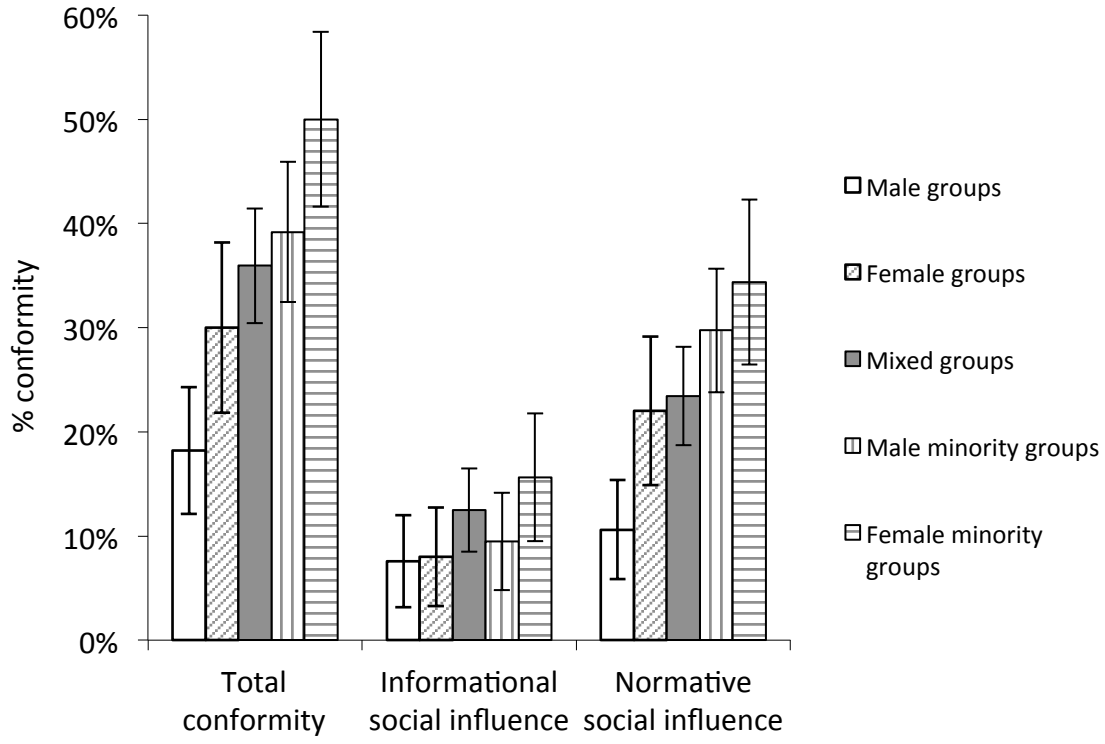


Table A1. OLS regressions on conformity in Experiment I (conducted at the Stockholm School of Economics). Same gender groups separated into male and female groups. Clustered standard errors (on groups) in parentheses and p-values in squared brackets.[1]

	Total Conformity		Informational Social Influence		Normative Social Influence	
Constant	0.435 (0.033) [<0.001]	1.357 (0.178) [<0.001]	0.129 (0.035) [0.002]	0.493 (0.203) [0.025]	0.306 (0.063) [<0.001]	0.864 (0.163) [<0.001]
Mixed group[1] (β1)	-0.205 (0.077) [0.016]	-0.173 (0.064) [0.014]	-0.001 (0.070) [0.991]	0.012 (0.071) [0.871]	-0.204 (0.075) [0.015]	-0.185 (0.070) [0.017]
Female group (β2)	0.015 (0.096) [0.881]	-0.052 (0.106) [0.627]	-0.046 (0.060) [0.458]	-0.072 (0.077) [0.363]	0.060 (0.114) [0.603]	0.020 (0.102) [0.849]
Correct answers in part 1		-0.183 (0.036) [<0.001]		-0.072 (0.036) [0.059]		-0.111 (0.034) [0.005]
p-value β1=β2	0.070	0.301	0.571	0.351	0.020	0.051
N	100	100	100	100	100	100
R-squared	0.055	0.240	0.005	0.065	0.089	0.176
F-value [p-value]	3.68 [0.046]	15.51 [<0.001]	0.31 [0.736]	1.47 [0.257]	5.69 [0.012]	15.03 [<0.001]

[1] The baseline category is male group.

Table A2. OLS regressions on conformity in Experiment II (conducted at the Norwegian School of Economics and Business Administration). Same gender groups separated into male and female groups and the minority groups separated into male and female minority groups. Clustered standard errors (on groups) in parentheses and p-values in squared brackets.[1]

	Total Conformity		Informational Social Influence		Normative Social Influence	
Constant	0.081 (0.094) [0.394]	0.786 (0.190) [<0.001]	-0.009 (0.075) [0.906]	0.276 (0.151) [0.075]	0.090 (0.080) [0.267]	0.510 (0.187) [0.010]
Mixed group[2] (β1)	0.177 (0.087) [0.048]	0.125 (0.090) [0.173]	0.056 (0.074) [0.456]	0.035 (0.075) [0.648]	0.121 (0.073) [0.104]	0.090 (0.077) [0.250]
Female group (β2)	0.152 (0.093) [0.112]	0.085 (0.102) [0.413]	0.023 (0.062) [0.708]	-0.004 (0.065) [0.954]	0.128 (0.075) [0.094]	0.088 (0.082) [0.287]
Male minority group[3] (β3)	0.143 (0.114) [0.218]	0.118 (0.109) [0.288]	-0.014 (0.074) [0.847]	-0.024 (0.074) [0.744]	0.157 (0.092) [0.095]	0.142 (0.091) [0.128]
Female minority group (β4)	0.311 (0.120) [0.013]	0.238 (0.117) [0.049]	0.083 (0.095) [0.384]	0.054 (0.093) [0.567]	0.228 (0.100) [0.028]	0.184 (0.104) [0.085]
Correct answers in part 1		-0.132 (0.026) [<0.001]		-0.054 (0.024) [0.032]		-0.079 (0.029) [0.009]
p-value β1=β2	0.731	0.591	0.539	0.465	0.900	0.970
p-value β1=β3	0.728	0.936	0.253	0.307	0.634	0.446
p-value β1=β4	0.184	0.194	0.738	0.808	0.179	0.220
p-value β2=β3	0.935	0.754	0.470	0.668	0.730	0.505
p-value β2=β4	0.130	0.125	0.412	0.420	0.227	0.238
p-value β3=β4	0.211	0.304	0.238	0.290	0.470	0.651
N	191	191	191	191	191	191
R-squared	0.135	0.239	0.089	0.125	0.064	0.113
F-value [p-value]	2.02 [0.057]	10.21 [<0.001]	2.77 [0.011]	2.75 [0.010]	3.33 [0.003]	5.15 [<0.001]

[1] All regressions also include a set of dummy variables for the time of the sessions, whose coefficients are not shown in the table.

[2] The baseline category is male group.

[3] Groups with one man or one woman.

Robustness checks

Pooled data

Pooling the data for both studies and estimating the effect of mixed-gender groups and same-gender groups in a regression with a dummy variable for experiment (and the six dummy variables for the time of the data collection in Experiment II), yields the following result: As can be expected, with effects in different directions in the two experiments, the pooled effect is not significant for total conformity (mixed group coefficient=-0.040; p-value=0.514), informational social influence (mixed group coefficient=0.032; p-value=0.456) or normative social influence (mixed group coefficient=-0.072; p-value=0.160).

Non-linearity

It is possible that group characteristics have a non-linear effect on conformity. Thus, instead of our numerical conformity variable, we conduct an ordered probit analysis (with clustering on group) based on not conforming on any of the valid test questions, conforming on one of the valid test questions, and conforming on all valid test questions. This results in very similar significance levels of the variables as in the OLS analysis. In Experiment I, the p-value of the mixed group variable was 0.020 (0.021 for OLS) for total conformity and 0.003 (0.004 for OLS) for normative social influence. In Experiment II, the p-value of the minority group variable is 0.059 (0.068 in OLS) for total conformity and 0.049 (0.052 in OLS) for normative social influence.

More permissive specification of social influence

In our main specification, we confine attention to those cases in which all previous group members agree on the correct answer. As a robustness check we include all observations in which a majority of them do so, including the fraction of prior responders who gave the correct answer as an explanatory variable. Significance levels are broadly the same as in the regressions reported in Table 1 and 2, with somewhat smaller effect sizes (as can be expected as conformity is likely to decrease when we include observations where there is not unanimity in the group). The mixed group coefficient in Experiment I decreases from -0.212 (p=0.021) to -0.194 (p=0.009) for total conformity and from -0.234 (p=0.004) to -0.193 (p=0.005) for normative social influence. The minority group coefficient in Experiment II decreases from 0.160 (p=0.068) to 0.132 (p=0.075) for total conformity and from 0.137 (p=0.052) to 0.111 (p=0.058) for normative social influence. As expected, the coefficient on the new explanatory variable has a positive sign. In Experiment I, the significance levels are p=0.076 for total conformity and p=0.127 for normative social influence; with controls also for gender and the number of correct answers in part 1 the variable is significant at the 5% level for both total conformity (p=0.019) and normative social influence (p=0.031)). In Experiment II, the variable is significant for both total conformity (p=0.002) and normative social influence (p=0.005).

Ability diversity

It is conceivable that some estimates are affected by differences in heterogeneity across groups. Thus, we investigate whether ability diversity affects our results. We measure ability diversity as the difference between the highest and the lowest performance in part 1 within each group, and add this variable to those regressions in Tables 1 and 2 that control for the individual performance in part 1 (as this may be correlated with the ability diversity). The new variable is not significant at the 5% level in any of these regressions (and has little effect on the other coefficients). In the data from Experiment II, it is significant at the 10% level for informational social influence (p=0.062) and for normative social influence (p=0.068), but with effects in opposite directions (a negative effect for informational social influence and a positive effect for normative social influence).

Minority behavior

In Experiment II, we also test whether the behavior differs within minority groups between subjects in majority and subjects in minority (with and without controls for gender and the number of correct answers in part 1). The coefficient on a minority dummy variable was not significant at the 10% level in any of these regressions for total conformity, informational social influence or normative social influence.

Interaction between ability and group

It is possible that ability might impact behavior differently depending on group composition. We thus added an interaction term between performance in P1 (ability) and the group(s). In Experiment 1 this interaction variable was not significant for normative social influence (coefficient=-0.008; p=0.912). For informational social influence the interaction variable was significant at the 10% level (coefficient=-0.132; p=0.074), implying that each additional error is associated with a 3 percentage point greater informational social influence in the same gender group and a 17 percentage point greater informational influence in the mixed gender groups. For total conformity the interaction variable was significant at the 5% level (coefficient=-0.140; p=0.019), implying that each additional error is associated with a 14 percentage point greater total conformity in the same gender group and a 28 percentage point greater total conformity in the mixed gender groups.

In Experiment 2 the interaction variables between ability and the mixed group and the minority group were not significant at the 10% level for total conformity (coefficient=-0.023 (p=0.648) for the interaction with the mixed group, and coefficient=0.005 (p=0.932) for the interaction with the minority group), informational social influence (coefficient=0.027 (p=0.703) for the interaction with the mixed group, and coefficient=0.012 (p=0.866) for the interaction with the minority group), or normative social influence (coefficient=-0.051 (p=0.420) for the interaction with the mixed group, and coefficient=-0.007 (p=0.919) for the interaction with the minority group).

Script

1. Place the following material at each desk before subjects enter the room:
    - 1 caffeine survey
    - 1 part 1
    - 1 part 2 (IMPORTANT: make sure that the small number on the front page matches the number on the desk)
    - 1 part 3
    - 1 energy drink
    - 1 pen

2. Let subjects in and tell them to sit down at a desk of their own choice, point out that they are not supposed to browse the material before you say so!
3. When everyone is seating: Say hello, thank everyone for coming, and inform them that it will take 30-45 min and that they will earn 200 NOK afterwards
4. Make sure that everyone has turned off their cell phone
5. Inform about the procedure (the four parts): "You have 4 booklets in front of you. We do one booklet at a time, and let them be placed next to you until I say you can start. You will start with a caffeine survey. After you have filled out the survey you will continue with part 1, which consists of 6 pattern recognition problems (SHOW EXAMPLE QUESTION!). After that, you will be asked to drink an energy drink, and once you are finished we will do part 2 and 3. Part 2 consists of 32 questions and they are answered orally, part 3 contains 6 questions and they are answered on paper. We ask you to answer all questions; please make your best guess if you are uncertain. Any questions? I will not be able to answer questions during the sessions."
6. Do caffeine survey
    - "This survey will take 3 min, please answer as well as you can."
    - "You can now take the survey from your desk, circle the number that corresponds to your desk number, after that you can begin answering the questions."
    - Collect the caffeine survey and place them in your folder (IMPORTANT: Make sure that subjects have circled the correct seat number!)
7. Tell people to introduce themselves with their name and their current year of study
    - Make sure that this feels natural: "while I am collecting the survey you can introduce yourself for each other with your name and year of study"
8. Do part 1
    - "This part will take 3 min, you are supposed to answer all 6 questions, and only one answer per question."
    - "You can now take Part 1 from your desk, circle the number that corresponds to your desk number, after that you can begin answering the questions."
    - Collect Part 1 and place them in your folder (IMPORTANT: Make sure that subjects have circled the correct seat number!)
9. Tell people to drink the energy drink
10. Do part 2

- Before they start you need to explain the procedure and make sure that everything is clear: "In front of you there is a booklet containing 32 problems of the same type as in part 1. You will have about 15 seconds to find the correct answer to a problem, after that I will ask you to provide your answer one at a time. For each new question we rotate so next time you will answer first (point at the person sitting at desk number 2) and so on. I will write down your answer so there is no need to write them down in your booklet. Any questions? I remind you that I cannot answer questions during the session."
    - Possible question 1: Does everyone have the same questions? YES
    - Possible question 2: Does everyone have the same question at the same time? NO. Continue as normal but make a note since we want to exclude this group from the analysis!
- Circle each subject's gender (m or f) on your answering sheet
- As subjects calls out their answers you write them down on the answering sheet (gray color on the sheet indicates who is supposed to start answering that particular question!)
- Collect Part 2 and place them in the folder
- Place the answering sheet in the folder

11. Do part 3
    - "This part will take 3 min, you are supposed to answer all 6 questions, and only one answer per question."
    - "You can now take Part 3 from your desk, circle the number that corresponds to your desk number, after that you can begin answering the questions."
    - Collect Part 3 and place them in your folder (IMPORTANT: Make sure that subjects have circled the correct seat number!)

12. Thank everyone and point out that they should not talk to friends about the experiment, at least not today.
13. Give subjects the show-up fee!
14. Return to the meeting point and hand in the folder for this group. Get a new folder with material and repeat the procedure.