



The hidden gender effect in online collaboration: An experimental study of team performance under anonymity



Hyang-gi Song^a, Michael Restivo^b, Arnout van de Rijt^{c,*}, Lori Scarlatos^d, David Tonjes^d, Alex Orlov^e

^a Department of Sociology, SUNY Stony Brook, United States

^b Department of Sociology, SUNY Geneseo, United States

^c Department of Sociology & Institute for Advanced Computational Science, SUNY Stony Brook, United States

^d Department of Technology & Society, SUNY Stony Brook, United States

^e Department of Material Science & Engineering, SUNY Stony Brook, United States

ARTICLE INFO

Article history:

Available online 21 April 2015

Keywords:

Experiment
Anonymity
Team gender composition
Team performance
Online education
Wiki

ABSTRACT

It has been argued that the generally positive effect that female participation exerts on team performance ceases to exist under conditions of anonymity. We evaluate this thesis in the context of an online learning environment in which the gender of fellow student team members was not disclosed to subjects. To circumvent selection effects in the composition of teams we employed an experimental design in which female and male students were randomly assigned to teams of varying gender composition. Against expectations, we find that under anonymity gender composition continues to impact team performance, with all-female teams being most productive. Counter-intuitively, this team effect occurred in our study without female students *individually* being more productive than their male counterparts. These findings indicate that the presence of females on anonymous teams can have a hidden effect on the productivity of other team members. Our results underscore that despite face-to-face interaction in higher education increasingly being substituted by Internet-enabled communication, a student's social environment continues to impact academic learning in important ways.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

The dramatic increase in the use of the Internet as a medium and meeting place for team work (Chu & Kennedy, 2011; Sulisworo, 2012) necessitates a better understanding of how conditions specific to online environments exacerbate or suppress features of group interaction traditionally observed in face-to-face settings. This question is relevant in a range of settings, from distributed organizations in which teams seek to effectively communicate despite being geographical dispersed to online educational settings where students' grades may be affected by the efforts of their classmates. A common feature of online groups is that members can participate anonymously or through non-identifying user names. Anonymous participation has been thought to reduce the salience of certain group processes that rely on member identification. Specifically, studies suggest that anonymity in online settings may reduce or even eliminate the otherwise positive effect of gender diversity on team performance

(Caspi, Chajut, & Saporta, 2008; Dubrovsky et al., 1991; Perkowski, 2012; Wade, Cameron, Morgan, & Williams, 2011; Yang, Cho, Mathew, & Worth, 2011).

The thesis that the composition of a team matters net of the individual qualities and attributes of its members has repeatedly been confirmed in non-anonymous settings (Chen, Ren, & Riedl, 2010; DiTomaso, Post, & Parks-Yancy, 2007; Hannagan & Larimer, 2010; Van Knippenberg & Schippers, 2007). Gender diversity in particular has been demonstrated to positively affect the performance of face-to-face teams in various studies involving students (Dufwenberg & Muren, 2006; Hoogendoorn, Oosterbeek, & Van Praag, 2013; Ivanova-Stenzel & Kübler, 2011), with mixed groups outperforming both male-dominated and female-dominated groups. However, whether gender composition continues to significantly impact team performance when the gender identities of members are cloaked by a web interface is an open question. As interactions are more and more taking place through online discussion boards, wikis, and social media, knowing how and to what degree team composition can continue to matter even when member identities are not salient is an increasingly important question.

The provisional answer from past scholarship is that in anonymous settings, team gender composition is not consequential for

* Corresponding author at: Department of Sociology, Stony Brook University, Nicolls Road 100, Stony Brook, NY 11794, United States. Tel.: +1 631 632 7704.

E-mail address: arnout.vanderijdt@stonybrook.edu (A. van de Rijt).

team performance. In particular, several studies have successfully avoided the key methodological obstacle of selection effects in group composition through experimental, non-voluntary assignment of student members to online teams (Flanagin, Tiyaamornwong, O'Connor, & Seibold, 2002; Herschel, 1994; Klein & Dologite, 2000). All found no statistically significant relationship between female participation and performance. However, measurement of group performance was limited in duration and scope. Two studies measured performance as the quantity and quality of ideas produced during brief anonymous brainstorming sessions (Herschel, 1994; Klein & Dologite, 2000). This excludes important dimensions of team work such as shirking and the long-term coordination of work effort. The third study measured productivity indirectly as subjects's perceptions of group performance (Flanagin, Tiyaamornwong, O'Connor, & Seibold, 2002). Furthermore, no separate individual-level measures of productivity were analyzed, making it impossible to tease out group and member effects.

In the present article we report on research employing a similar experimental design but allowing an evaluation of gender composition effects on both the quantity and quality of work performed during a semester-long course project, while measuring performance at both the individual and team level. Teammates were able to observe others' contributions and interact by sending comments and revising content through a collaborative interface, or "wiki". We find that female participation has a positive effect on both quality and quantity of work; however, we do not find beneficial effects of gender diversity per se. Rather, we find that productivity *monotonically* increases with the proportion of female students on a team. Interestingly, despite all-female teams generating the most and the best work, female students individually were no more productive than their male counterparts. These findings indicate that, even though gender identities were undisclosed, team members were influenced by the gender of other members.

2. Prior work

Theoretical arguments for why a team's gender composition may matter in an anonymous collaborative settings can be usefully grouped into three categories, as discussed in the next three subsections, respectively.

2.1. Skill diversity

Teams with members from more diverse backgrounds may combine a broader swath of non-overlapping competencies and experiences (Chen et al., 2010; Hong & Page, 2004; Jehn, Northcraft, & Neale, 1999) which could lead them to generally perform better than more homogeneous teams. Insofar as women possess different skills than men, a gender diverse team draws on a broader set of skills (Hamilton, Nickerson, & Owan, 2003; Hong & Page, 2004). As a result, gender diversity may enhance team performance.

Gender differentiation in skill sets is borne out of differences in socialization between men and women. This effect of gender diversity on team productivity transcends face-to-face interactions and continues to operate in anonymous online settings (Caspi et al., 2008; Friday, Friday-Stroud, Green, & Hill, 2006; Perkowski, 2012; Shollenberger, 2009; Wade et al., 2011; Yang et al., 2011). Gender-specific learning experiences simply equip a gender diverse team with more experience, whether or not gender is identified. Furthermore, individuals may continue to "do gender" (West & Zimmerman, 1987) as they are socialized to do, even when

they do not identify themselves. We may thus derive the following hypothesis:

Hypothesis 1. Team performance is higher in mixed-gender teams than in all-female or all-male teams.

2.2. Individual productivity

In some settings women tend to outperform men, which naturally aggregates to a situation in which all-female teams outperform other teams. This is particularly true in educational settings (DiPrete & Buchmann, 2013), where most experimental data come from. Over the past several decades, girls have come to receive higher grades than boys at most levels of education (Duckworth & Seligman, 2006). Starting from kindergarten all the way through college years, female students overall earn better grades (Perkins, Kleiner, Roey, & Brown, 2004). This owes not only to the fact that women display more advanced reading skills from an early age (Tach & Farkas, 2006) and recently have caught up with men in mathematics and science as well (Catsambis, 2005), but also to various noncognitive attributes where women seem to show an advantage over men. High school teachers consistently report that female students display greater interest in school and put more effort in their work than male students, who appear to be more disruptive during class. The difference in school grades between males and females may be partly attributable to these different characteristics (Downey & Vogt Yuan, 2005; Rosenbaum, 2001).

DiPrete and Buchmann (2013) argue that because boys receive intense peer pressure to take up a masculine identity, this often results in less emotional attachment to school. This kind of reinforced masculine identity may hinder them from developing close ties to school and coursework. Since these general differences between men and women have developed throughout years of education, one would expect that even when gender identity is not disclosed to group members, all these attributed differences should persist because roles have been internalized. Thus, we should observe that women in general outperform men in educational settings. We derive the following hypothesis:

Hypothesis 2. The performance of female team members exceeds that of their male counterparts.

It is important to note that it is difficult to observe a diversity effect on team performance in the presence of a strong individual gender effect. The latter may simply swamp the former, making all-female teams outperform other teams even when in mixed-teams female members do improve the performance of male members. It is thus critical that studies disentangle individual from group level effects.

2.3. Group-oriented behavior

A long tradition of functionalist and gender-role socialization theorists argue that through socialization, societal patterns of gender-role differentiation are produced and maintained (Bales, 1953; Parsons & Bales, 1955; Zelditch, 1955). Such socialization would generate, on average, a relatively uniform and stable set of differentiated gendered behaviors and behavioral expectations across individuals and social settings. For instance, women are thought to enact more expressive behaviors than men who perform more instrumental acts, mostly regardless of context (Walker, Iardi, McMahon, & Fennell, 1996). Thus, even under anonymity, women would continue to use communication in a more collaborative fashion with others and for strengthening of their personal

networks (Arbaugh, 2000; Barrett & Lally, 1999). Gender differences in socialization may also produce a tendency for women to be more group-oriented and cooperative in their behavior. Women in teams may help create more generous and egalitarian group dynamics (Dufwenberg & Muren, 2006). Women may add an element of cooperation and social sensitivity to otherwise competitive male environments (Ivanova-Stenzel & Kübler, 2011; Woolley, Chabris, Pentland, Hashmi, & Malone, 2010). As a result, the presence of women on a team may increase productivity and performance of other team members. We derive the following hypothesis:

Hypothesis 3. The performance of both female and male team members monotonically increases with the number of other team members who are female.

2.4. Previous experimental studies

We highlight three studies that have employed randomized experiments to test how the gender composition of teams affects team performance when gender identities are masked by anonymous online identities (Flanagin, Tiyaamornwong, O'Connor, & Seibold 2002; Herschel, 1994; Klein & Dologite, 2000). The key differences between these studies and the present study are that: (1) studies granted only limited space for participants to interact with other team members thereby limiting the role of team work; (2) performance was measured only at the level of student teams, not individual members, or it was not measured at all; and (3) in some cases not all relevant permutations of group composition were present (all-male teams were missing) due to a severe gender imbalance in the study population. These three limitations prevented these studies from effectively discriminating between the distinct arguments and hypotheses discussed in Sections 2.1–2.3.

Herschel (1994)'s study was the first to assess the gender composition effect in computer-mediated group projects through experimental methods. In a study of 61 undergraduate groups, consisting of 4–5 members with a gender composition ranging from 0% to 100% male, Herschel measured the gender composition effect on team brainstorming performance during 20-min brainstorming sessions. In this study, participants shared ideas on a particular issue presented by the researchers. Participants were seated behind their own computer workstations but could see all ideas so far proposed by other participants in the group. They were asked to type as many ideas as possible on the provided task decision items during the activity sessions, and a group was evaluated on its productivity by the total number of ideas generated by each group. Interaction was thus limited to the ability to view the groups' ongoing brainstorming results. Herschel found that the gender composition of a team did not affect brainstorming performance for groups working together anonymously through a computer-mediated setting.

Klein and Dologite (2000) conducted a study in which 114 undergraduate students were randomly assigned to teams of different gender composition. Participants were asked to come up with a one-paragraph proposal for either a new or improved operating system for a fast-food restaurant. As in Herschel (1994)'s study, participants were seated at individual computer station that were networked, and asked to type as many ideas as possible during 15 min. Then, the system electronically recorded all ideas and presented them at each computer station. Next, participants were asked to rate the ideas, discuss them, and select the best one as the group's proposed solution. Group performance was assessed by independent judges. Klein and Dologite also found no statistically significant difference in the group performance of

all-male, all-female, and mixed-sex groups in their anonymous online setting.

While these studies provide initial evidence with regards to how gender composition affects team performance under anonymity, both studies leave a number of issues unaddressed. As mentioned, in both studies, communication and interaction were limited. Moreover, neither study separately measured the performance of individual participants. It is therefore possible that the lack of a significant difference in performance across groups of different gender composition may be due to confounding. For example, if the performance of female participants on the individual level is stronger or weaker than male participants (Hypothesis 2), but at the same time a member's gender has the opposite effect on other team-members' performance (Hypothesis 3), then the magnitude of these two statistical effects could be canceling each other out in the combined analysis. Thus, we believe it is imperative to consider performance at both the individual and group level to avoid this type of statistical confounding. Given the statistical analysis employed in these studies, had a difference between groups shown up (i.e., if teams with more women had done better), it would have been unclear whether to attribute the effect to gender differences in performance between men and women (Hypothesis 2), or instead, to the differential performance of men in the presence of women and vice versa (Hypothesis 3). Further, both studies focused on a particular measure of outcome, brainstorming performance (Herschel, 1994; Klein & Dologite, 2000), which may not capture other important dimensions of group performance such as long-term commitment and willingness to coordinate and work together in a cooperative manner. Finally, both studies were conducted during a fairly short period of time (roughly in 20 min of group collaboration), leaving open the possibility that gender differences emerge during extended group collaboration over longer periods of time.

Flanagin, Tiyaamornwong, O'Connor, and Seibold (2002) conducted an experimental study spanning 10 weeks in which they tested gender composition effects at both individual and group level. With 69 undergraduate students (42 female and 27 male students), the researchers tested whether an all-female group would experience higher levels of process-oriented group behaviors when coordinating on the completion of course tasks involving joint authoring. This study design granted more room for communication and interaction among teammates. All participants were asked not to reveal their identity and work with non sex-specific pseudonyms anonymously. The researchers used a custom-designed, computer-based software application via Internet. With this software, participants were able to work on jointly authored group documents together by sending electronic messages to others and placing information in a common database. However, due to the skewed gender distribution in the study, the authors compared three all-female groups with nine mixed-sex groups, but no all-male group was studied, making it statistically harder to identify a gender composition effect. Moreover, the study focused on dimensions of "social or process-level group behavior" such as the level of group cohesion, group trust, task interdependence and group satisfaction, but neither group performance, nor output, nor productivity was directly measured as an outcome variable. Contrary to the authors' expectations, no group-level gender composition effect was found on any of the dependent measures.

In the present study, we follow the same experimental group assignment strategy as in the three aforementioned studies to circumvent selection effects in the composition of student teams. We seek to advance this line of research in four ways. First, we grant sufficient space for group members to communicate and interact with other teammates to produce a group project together without revealing one's true identity. Second, instead of focusing on one

form of productivity previously studied (brainstorming success), we measure individual contributions made to a shared group project as well as the quality of this jointly-created product. Third, we observe student team performance over the course of an entire semester, thus capturing long-term behavioral effects of gender composition that may emerge over such a time period. Fourth, in our statistical analysis we tease out the effects of gender composition on the quality and quantity of work performed at the individual and group level, allowing us to differentiate between [Hypotheses 1–3](#) and thus between distinct sets of causal mechanisms driving gender composition effects.

3. Method

3.1. Educational setting

This study's research protocol was approved by Stony Brook University's Institutional Review Board (project id #382719). To study how gender composition affects productivity in anonymous teams, we used a *wiki* as the online learning platform. The wiki platform had three main pages: home, students, and help. When students logged in to the platform, they first saw an announcement page where instructions for the weekly group task were posted. Each group received identical instructions. Students could then proceed to view their own team's page but not those of other groups. The wiki page was where the actual collaboration work was performed. Each team consisted of three students who were given exclusive access to one team-specific wiki page. They could not access other teams' wiki pages, thereby limiting interactions across team environments and preventing other forms of influence from occurring.

By using a wiki, students could immediately see how their input appeared as a contribution to their group's final product. Each student in a group used a different color text (assigned to them) in order to indicate which contributions were theirs. They could format the page freely and also add pictures. At the bottom of the page, group members could leave comments for other members of their group as well as respond to comments left by others. Finally, in addition to using the color of the text to determine which group member contributed what content, students could use the "history" function of the wiki to track contributions and compare changes across revisions.

Each student was designated a generic, anonymous username. Each group of three students was instructed to construct a wiki page detailing the "environmental impact profile of a beverage container" as part of their course requirements. Over the course of ten weeks, students were given a set of weekly instructions with a required task to complete, which would become part of their wiki page. At the end of the semester, the quality of their work was evaluated by course instructors on multiple dimensions. Students' individual contributions were automatically tracked by the wiki software, which facilitated the data collection process. Students received "participation credit" toward their final grade, but the quality of their team's page would not determine the individual students' grades. Students who opted not to participate in the experiment were grouped together in three-person teams and given an equivalent assignment to complete; however, only data from students who consented to participation was included in our analysis.

With three students assigned to each team, we varied the gender composition across the entire spectrum: groups could consist of all male students, or two male and one female student, or two female and one male student or all female students. The study aimed to assess both individual performance outcomes for students as well

as group-level performance of teams across these varying gender compositions.

3.2. Study subjects

Students were recruited into the experiment from five undergraduate courses at Stony Brook University. The courses included four sections of a freshman seminar in Information Technology as well as a medium size upper-division Sociology course. We selected these classes since several topics included in the curriculum related to the use of new media. Students were told that regardless of whether they consented to participate in the research project, all students would be required to complete an identical "semester-long group assignment" as part of their course participation credit. The instructors of these classes agreed to count this assignment as "participation credit" for the course. Students who consented to participate in the study were assigned to groups of three, while those who did not consent to be part of the research were grouped with others who also opted out. The total number of students enrolled in these courses was 176, of which 111 students (63%) consented to participate in the research project. We did not identify any factors that were associated with selecting into the research study.

Students were assigned a username and password to log in to their team's wiki page, which only members of their team and the course instructors could access. With three students per group, each student was assigned a username that combined a color (red, blue, or green) and a number representing their group (e.g. *red23* or *blue17*).

3.3. Experimental design and randomization

The gender distribution of the 111 students was mostly balanced with 60 female students and 51 male students. We randomly assigned these 111 students to 37 groups of three while maximizing balance: We created 8 groups of three male students, 8 groups of two male and one female student, 11 groups consisting of one male and two female students, and 10 groups of three female students. There were 58 students from Information Technology course sections and 53 from Sociology. We reduced the possibility for off-line, non-anonymous contact by composing teams of students from different classes. Given combinatorial constraints, all groups except two had a mix of students drawn from both Information Technology as well as Sociology courses. IT students on teams were chosen such that they were always drawn from different IT sections.

It was difficult for students to find out who else was in their group, because students did not know which other undergraduate courses besides the one they were enrolled in were included in the study, and because assigned user names were non-identifying and gender-blind. Based on informal debriefings with students at the end of the semester, the instructors concluded that there was no evidence that any of the students had discovered the identities of their team members. A post-hoc analysis of all comments made by students confirms the sustained nature of anonymity throughout the experiment. In none of the comments did a student share identifying information.

3.4. Group work procedures

Every Monday morning, all students received an identical weekly task in the form of an instructional paragraph posted on the homepage which they could immediately see when they logged in. For instance, students were asked to identify locations on campus where their beverage container was used or were asked to upload images showing their beverage containers.

Throughout the project, students worked with other teammates anonymously. Students could log in and work on their team page at any time or communicate with other team members through the comments function. Students could use the comments function rather freely for such things as encouraging participation or clarifying confusion regarding the assignments. Through careful post-hoc screening of the editing and commenting records, we verified that no identifiable information was exchanged to others in the group, such as gender, race, course, or year of academic standing.

3.5. Operationalization of variables

We measured dependent and independent variables at two different levels. Table 1 provides descriptive statistics for variables of both types. First, at the individual level, we record each student's gender, their respective class (IT freshman or SOC upper division), as well as individual work performance. To measure work performance we counted the number of logins, edits, words, photos, and comments each person contributed throughout the course of the assignment. Existing research on peer production has used similar measures of performance, including edits per day on Wikipedia (Kittur, Pendleton, & Kraut, 2009; Panciera, Halfaker, & Terveen, 2009; Suh, Chi, Pendleton, & Kittur, 2007; Wilkinson & Huberman, 2007; Restivo & van de Rijt, 2012; Zhu, Kraut, & Kittur, 2013; Restivo & van de Rijt, 2014; van de Rijt, Kang, Restivo, & Patil, 2014), number of code changes to an open source project (Apache) (Mockus, Fielding, & Herbsleb, 2000), and number of work items attempted on Mechanical Turk (Mason & Watts, 2010). These quantitative measures are an indicator of the level of involvement and contribution of a student to the group project. We assume that higher values on this quantitative measure indicate greater levels of student involvement and contributed to the group project.

At the team level, we measured the gender composition of the team (# female students), the class composition of the team (# SOC students), and overall team performance, both quantitatively and qualitatively. The quantitative measures simply aggregate all individual members' contributions, namely the numbers of logins, edits, words, photos and comments the team together contributed.

Table 1
Descriptive statistics of all variables.

	Mean	Standard deviation	Minimum	Maximum	Number of cases
<i>Student variables</i>					
Gender	.541	n.a.	0 (male)	1 (female)	111
Class	.477	n.a.	0 (ITS)	1 (SOC)	111
# Logins	5.18	4.54	0	34	111
# Edits	2.27	2.10	0	9	111
# Words	177	191	0	1057	111
# Photos	1.09	1.62	0	9	111
# Comments	.432	.870	0	4	111
<i>Team variables</i>					
Gender composition (# females)	1.62	1.11	0	3	37
Class composition (# SOC)	1.43	.603	0	2	37
# Team logins	15.5	7.70	6	44	37
# Team edits	6.81	3.91	0	18	37
# Team words	532	335	0	1315	37
# Team photos	3.27	2.84	0	11	37
# Team comments	1.30	1.90	0	7	37
Quality work, # criteria met	3.66	1.48	.5	6	37
Quality work, subjective	2.95	1.12	1	5	37

Additionally, the instructors and authors of this study evaluated the quality of each team's work on dimensions of adherence to the posted instructions, the aesthetics of their project, and coherence of their page's contents. We report the average of these quality dimensions as the first team-level measure reported in Table 1. Instructors also subjectively evaluated the overall quality of each project on a scale ranging from 1 (poor) through 5 (excellent). This is the second quality measure reported in Table 1.

4. Results

Fig. 1 displays the average performance observed for teams of different gender composition. Gender composition is represented as the number of women on a three-person team, being 0, 1, 2, or 3. Each panel presents this bivariate pattern of four group averages (circles) for a different performance indicator. Three confidence intervals are shown indicating the degree of significance with which performance differed between teams composed of only women (baseline) and the other three types of teams. These confidence intervals are based on negative binomial and ordinary least squares regression with dummy variables (without any controls). Panels A through E show no support for the thesis that diversity promotes team performance (rejecting Hypothesis 1). Although significance levels vary a bit from panel to panel, in each case the quantity of work achieved increases with the number of female students on a team, with female-only teams consistently showing the highest performance levels. Panel F, which measures the first factor from principal-factor analysis on the 5 indicators from Panels A through E, shows the same pattern. Panels G and H measure effects on qualitative dimensions of performance, again revealing a monotonic effect of female team membership.

One possibility is that the basic bivariate patterns in Fig. 1 represent a dominant individual-level effect of gender that masks a team-level diversity effect. If in the student population female students happen to be generally much better and harder working than male students, then diversity may still have had an additional small positive effect on each member's performance. Additionally, it may be that the participation credit did not sufficiently incentive men. Fig. 1 would then be confounding these two effects. An alternative possibility is that Fig. 1 represents a team-level effect of gender, with more female members improving other members' performance. Our quantitative measurement of performance at both the team and student level allows us to disentangle any possible team composition effect from a student effect.

Accordingly, Fig. 2 displays average performance broken down both by the gender of the focal student and by the gender of the two other team members, for each of the six quantitative performance measures from Panels 1A–F. Fig. 2 shows that the monotonic effects of female students' representation observed in Fig. 1 are not due to simple individual-level gender effects (rejecting Hypothesis 2). In collaborations with two male students, female students did roughly the same amount of work as their male counterparts, regardless of the measure of work. The same is true for collaborations with one male and one female student as well as for those with two female students. The only exception is commenting frequency (Fig. 2, Panel F), where we do see an individual-level gender difference. While male students seem to comment at a low rate regardless of team composition, female students' comment frequency appears to increase with the number of other female students on their team. For all other measures, we find that both male and female output generally increases with the number of other team members who are female (confirming Hypothesis 3).

Given that female students disproportionately came from the Sociology upper level class, we tested whether the general pattern

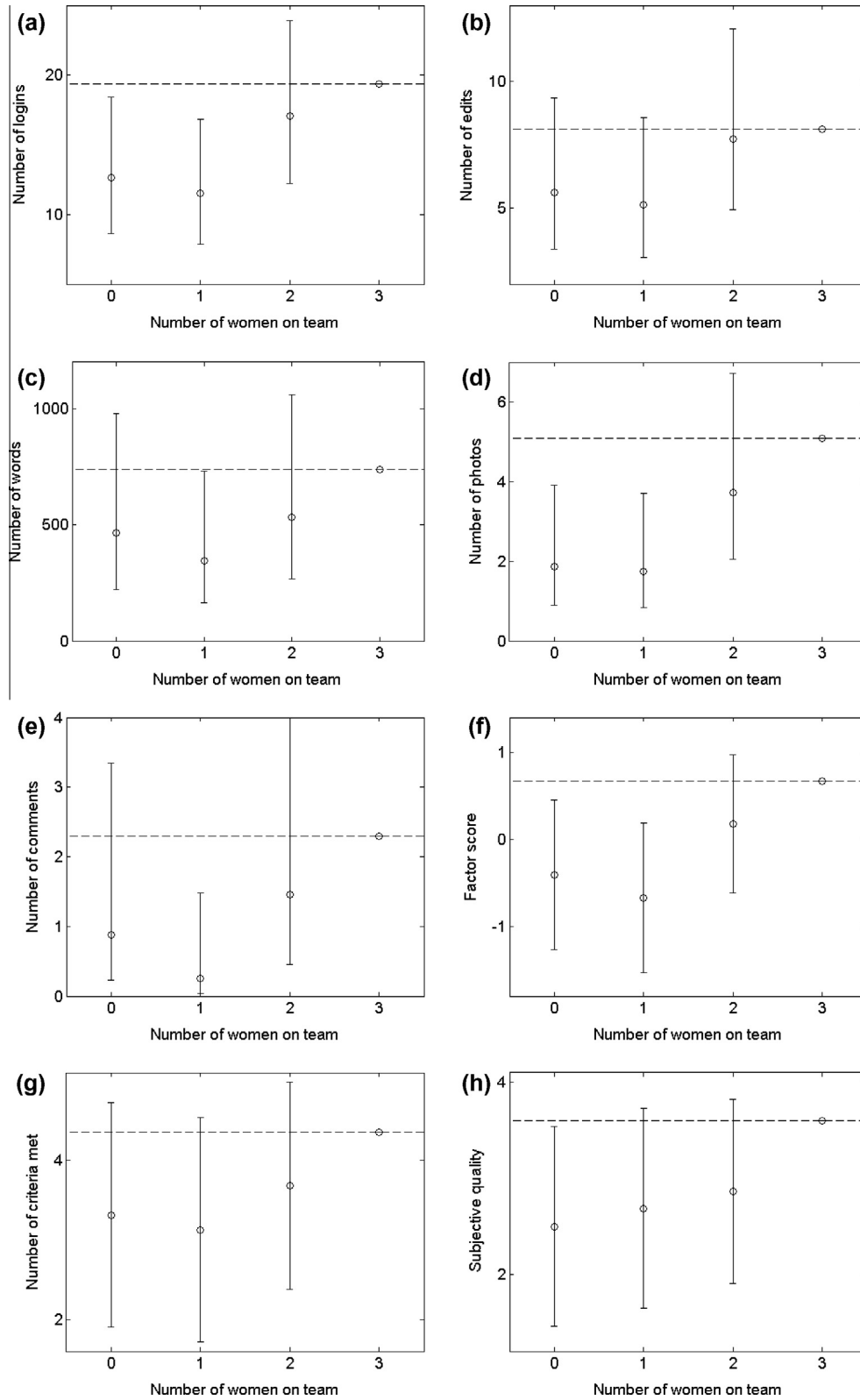


Fig. 1. Quantity and quality of team work by gender composition. Shown are group averages (black dots) for the four gender composition groups ($N = 37$). Confidence intervals (solid lines) are derived from negative binomial or ordinary least squares regression on group dummy variables, with “all female” as baseline (dotted line).

observed in Fig. 2 (of a positive effect of female team membership combined with a lack of an individual gender effect) held up when class membership was controlled. To this end we estimated random-effects negative binomial regression models predicting each

of the five measures of work from the students' gender and class as well as that of their team members. We also estimated a random-effects linear regression model with the first factor from principle factor analysis as dependent variable. The results in Table 2

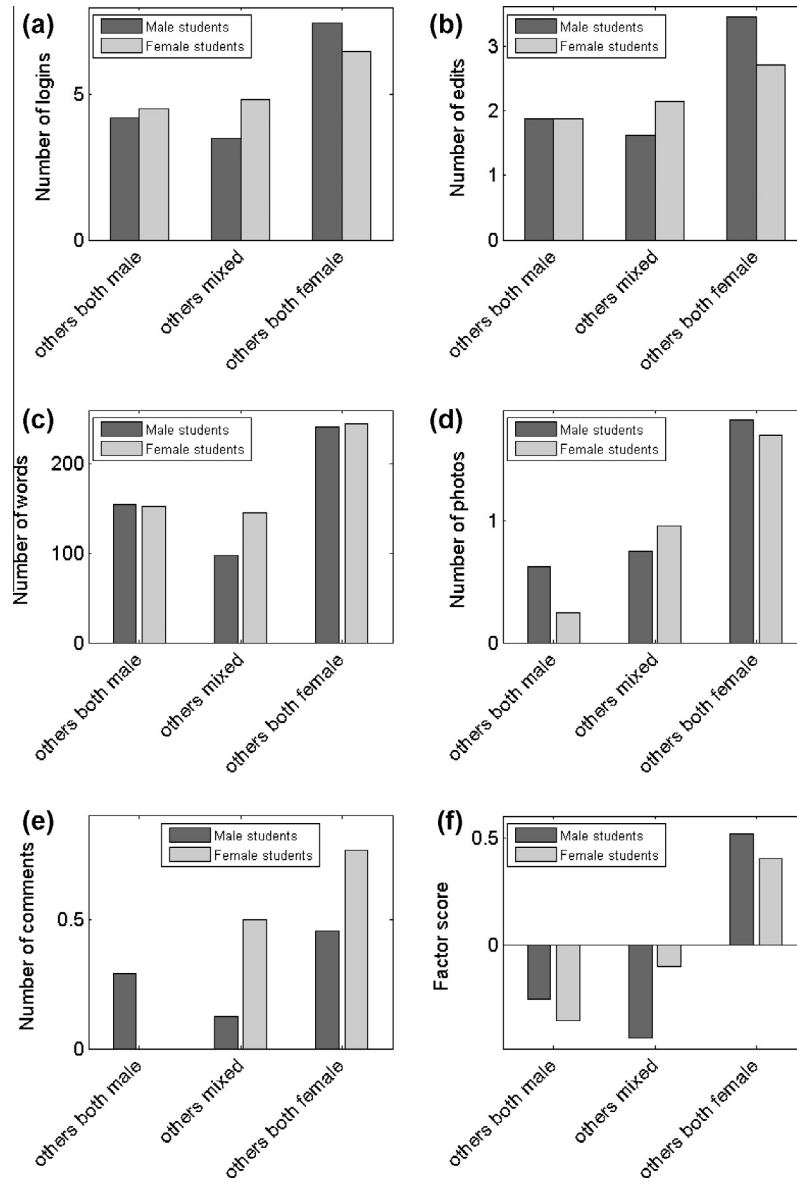


Fig. 2. Quantity of individual work by own gender and others' genders. Bars represent averages for each of the six combinations ($N = 111$).

Table 2
Random effects regression of quantity of individual work.

	# Logins	# Edits	# Words	# Photos	# Comments	Factor score
<i>Student variables</i>						
Female	.289 (.181)	.300 (.213)	.342 (.245)	.304 (.313)	.302 (.525)	.327 (.214)
Sociology student	-.563** (.192)	-.623** (.229)	-.915 (.268)	-.626 (.353)	.212 (.629)	-.586 (.227)
<i>Team variables</i>						
Gender of team members						
2 Males	-.442 (.236)	-.602* (.280)	-1.168*** (.328)	-1.222** (.458)	-1.577* (.783)	-.728** (.273)
Mixed	-.415* (.183)	-.471* (.210)	-.853*** (.243)	-.624* (.302)	-.854 (.528)	-.690** (.218)
2 Females (baseline)						
# Sociology students	-.262 (.165)	-.246 (.192)	-.464* (.223)	-.065 (.292)	-.337 (.608)	-.201 (.192)
Constant	1.398*** (.358)	1.578** (.466)	.345 (.402)	.535 (.595)	1.346 (1.491)	.741* (.351)
Number of teams	37	37	37	37	37	37
Number of individuals	111	111	111	111	111	111

Standard errors are shown in parentheses.

* $p < .05$.

** $p < .01$.

*** $p < .001$.

are consistent with Fig. 2. For none of the six quantitative measures of performance is there a significant individual effect of gender (rejecting Hypothesis 2), while for each of the six measures, there is a significant difference between teams where both other students are female compared to teams where one or none of the other students are female (rejecting Hypothesis 1 and confirming Hypothesis 3).

Finally, we explored the possibility that greater cooperativeness among female participants may have contributed to the positive effect of female membership on group performance in our study. To this end we coded all student comments based on their contents into three groups: misusing it as a text box (“text”), initiating collaborative atmosphere or asking questions (“initiative”), or responding to one’s initiation (“responsive”).

Six “initiative” and ten “responsive” comments were made by a total of ten individuals, three of which were male and the other seven female. All responsive comments were made by women. Among the three male students, two were in all-male groups, and their initiative was not echoed by other teammates. For example, one of them commented the following without receiving a response:

(1) “Guys, you may add more information about the last two questions which need some research. And maybe add some nice pictures either photographed or got from the Internet. Thanks a lot.” (2) “Guys, please make some efforts in the ninth assignment. Thanks a lot.”

The third male student who initiated a conversation did so in a team with two females:

“Are there any other members of the group who would like to make a contribution to the page??”

This comment received a response by one of the female teammates who wrote:

(bluet51) “I also found cups in tabler at dunkin donuts but [can not] seem to find them anywhere else on campus.”

Not only did females respond more readily, they also elicited responses more successfully. All three females who posted an initiative comment received comments all of which came from other women. For example, the comment

“I really like that picture of the wine bottles being used as the candle holder, very creative =)”

received the response

“Thank you!”

On another team the following exchange took place among women:

“Hello team, I am not sure if we are supposed to reveal our identities or not? But I think this is a good way of communicating between each other! Thanks! Also, do you think that Javits has canned beverages as an option for vending? Don’t remember off the top of my head!!”

“I don’t think so, but I don’t exactly remember. I’m pretty sure Javits just has water for vending.”

“ey blue.”

These results suggest that greater collaboration among women may partially explain the gender composition effect.

5. Discussion

Our study did not identify an effect of gender diversity on either the quantity or the quality of team effort in an anonymous online

educational setting. However, instead we identified a positive, monotonic effect of the number of female student team members on team performance. If there was a gender diversity effect, we would have observed productivity being highest in gender-mixed groups. Instead we found that a larger proportion of female students on a team was associated with both quantitative and qualitative improvements in a group’s final product. Our research strategy of measuring performance at both individual and team level allowed us to evaluate two competing accounts for this effect: (1) an individual-level phenomenon whereby female students simply did more and/or better work than their male counterparts, or (2) a team-level phenomenon whereby female students’ participation raised the performance levels of other members in the group. Our evidence clearly favors this second explanation over the first. This supports the notion that the presence of female students on a team exerted a positive influence on the efforts of other student members, even while gender was not disclosed. These results raise the question of what it is about female students that makes other students on the team more productive. Like all other experimental studies in which women and men are randomly assigned to groups of varying gender composition, also our study cannot definitively identify what unobserved features associated with gender are responsible for group-level gender effects. We speculate that while the female students in our study did not work harder or individually perform better than their male counterparts, their style of participation may yield the key to the differential outcomes. One possibility is that they may have acted more cooperatively. Our analysis of commenting behavior, in which we observed that female students engaged more in collaborative conversations than male students, provides some support for this mechanism. Students’ comments elicited more discussion in groups with more women, facilitating greater collaboration and providing opportunities to address and resolve possible group conflict. Future research should more rigorously investigate this possibility.

More generally our study suggests that the gender composition of groups may continue to impact group behavior even under anonymous online conditions. This includes learning outcomes in educational settings where the Internet is increasingly being used as medium for team cooperation. Our findings underscore the importance of the social environment in academic learning and we urge educators to carefully consider the significant impact that team composition may have on the achievement of group collaboration. Specifically, the disappearance of all gender differences in productivity in our study once team composition effects are controlled suggests that gender segregation in educational settings may be an underappreciated factor in the explanation of gender differences in student performance.

References

- Arbaugh, J. B. (2000). Virtual classroom characteristics and student satisfaction with internet-based MBA courses. *Journal of Management Education*, 24(1), 32–54.
- Bales, R. F. (1953). The equilibrium problems in small groups. In T. Parsons, R. F. Bales, & E. A. Shils (Eds.), *Working papers in the theory of action* (pp. 116–161). Glencoe, IL: Free Press.
- Barrett, E., & Lally, V. (1999). Gender differences in an on-line learning environment. *Journal of Computer Assisted Learning*, 15(1), 48–60.
- Caspi, A., Chajut, E., & Saporta, K. (2008). Participation in class and in online discussions: Gender differences. *Computers & Education*, 50(3), 718–724.
- Catsambis, S. (2005). The gender gap in mathematics: merely a step functions? In A. M. Gallagher & J. C. Kaufman (Eds.), *Gender differences in mathematics: An integrative psychological approach* (pp. 220–245). Cambridge, UK: Cambridge Univ. Press.
- Chen, J., Ren, Y., & Riedl, J. (2010). The effects of diversity on group productivity and member withdrawal in online volunteer groups. In *Proceedings of the 28th international conference on Human factors in computing systems – CHI ’10* (p. 821). New York, New York, USA: ACM Press. doi:<http://dx.doi.org/10.1145/1753326.1753447>.

- Chu, S. K.-W., & Kennedy, D. M. (2011). Using online collaborative tools for groups to co-construct knowledge. *Online Information Review*, 35(4), 581–597.
- DiPrete, T. H., & Buchmann, C. (2013). *The rise of women: The growing gender gap in education and what it means for American schools*. New York: The Russell Sage Foundation.
- DiTomaso, N., Post, C., & Parks-Yancy, R. (2007). Workforce diversity and inequality: Power, status, and numbers. *Annual Review of Sociology*, 33(1), 473–501. <http://dx.doi.org/10.1146/annurev.soc.33.040406.131805>.
- Downey, D. B., & Vogt Yuan, A. S. (2005). Sex difference in school performance during high school: Puzzling patterns and possible explanations. *Sociological Quarterly*, 46(2), 299–321.
- Dubrovsky, V. J., Kiesler, S., & B. N. Sethna. *The equalization phenomenon: Status effects in computer-mediated and face-to-face decision-making groups*.
- Duckworth, A. L., & Seligman, M. E. (2006). Self-discipline gives girls the edge: Gender in self-discipline, grades, and achievement test scores. *Journal of Educational Psychology*, 98(1), 198–208.
- Dufwenberg, M., & Muren, A. (2006). Gender composition in teams. *Journal of Economic Behavior & Organization*, 61(1), 50–54.
- Flanagin, A. J., Tiyaamornwong, V., O'Connor, J., & Seibold, D. R. (2002). Computer-mediated group work: The interaction of sex and anonymity. *Communication Research*, 29(1), 66–93.
- Friday, E., Friday-Stroud, S. S., Green, A. L., & Hill, A. Y. (2006). A multi-semester comparison of student performance between multiple traditional and online sections of two management courses. *Journal of Behavioral & Applied Management*, 8(1), 66–81.
- Hamilton, B., Nickerson, J., & Owan, H. (2003). Team incentives and worker heterogeneity: An empirical analysis of the impact of teams on productivity and participation. *Journal of Political Economy*, 111(3), 465–497.
- Hannagan, R. J., & Larimer, C. W. (2010). Does gender composition affect group decision outcomes? Evidence from a laboratory experiment. *Political Behavior*, 32(1), 51–67.
- Herschel, R. T. (1994). The impact of varying gender composition on group brainstorming performance in a GSS environment. *Computers in Human Behavior*, 10(2), 209–222.
- Hong, L., & Page, S. E. (2004). Groups of diverse problem solvers can outperform groups of high-ability problem solvers. *Proceedings of the National Academy of Sciences of the United States of America*, 101(46), 16385–16389. <http://dx.doi.org/10.1073/pnas.0403723101>.
- Hoogendoorn, S., Oosterbeek, H., & Van Praag, M. (2013). The impact of gender diversity on the performance of business teams: Evidence from a field experiment. *Management Science*, 59(7), 1514–1528.
- Ivanova-Stenzel, R., & Kübler, D. (2011). Gender differences in team work and team competition. *Journal of Economic Psychology*, 32(5), 797–808.
- Jehn, K. A., Northcraft, G. B., & Neale, M. A. (1999). Why differences make a difference: A field study of diversity, conflict and performance in workgroups. *Administrative Science Quarterly*, 44(4), 741–763.
- Kittur, A., Pendleton, B., & Kraut, R. E. (2009). Herding the cats. In: *Proceedings of the 5th international symposium on wikis and open collaboration – WikiSym '09* (pp. 1–9). New York, New York, USA: ACM Press.
- Klein, E. E., & Dologite, D. G. (2000). The role of computer support tools and gender composition in innovative information system idea generation by small groups. *Computers in Human Behavior*, 16(2), 111–139.
- Mason, W., & Watts, D. J. (2010). Financial incentives and the performance of crowds. *ACM SigKDD Explorations Newsletter*, 11(2), 100–108.
- Mockus, A., Fielding, R. T., & Herbsleb, J. (2000). A case study of open source software development: The apache server. In *Proceedings of the 2000 international conference on software engineering, 2000* (pp. 263–272). IEEE.
- Panciera, K., Halfaker, A. & Terveen, L. (2009). Wikipedians are born, not made. In *Proceedings of the ACM 2009 international conference on supporting group work – GROUP '09*. New York, New York, USA: ACM Press.
- Parsons, T., & Bales, R. F. (1955). *Family, Socialization and Interaction Process*. Glencoe, IL: Free Press.
- Perkins, R., Kleiner, B., Roey, S., & Brown, J. (2004). *The high school transcript study: A decade of change in curricula and achievement, 1990–2000*. Washington, DC: DIANME Publishing.
- Perkowski, J. (2012). The role of gender in distance learning: A meta-analytic review of gender differences in academic performance and self-efficacy in distance learning. *Journal of Educational Technology Systems*, 41(3), 267–278.
- Restivo, Michael, & van de Rijt, Arnout (2012). Experimental study of informal rewards in peer production. *PLoS ONE*, 7(3), e34358.
- Restivo, Michael, & van de Rijt, Arnout (2014). No praise without effort: Experimental evidence on how rewards affect wikipedia's contributor community. *Information, Communication and Society*, 17(4), 451–462.
- Rosenbaum, J. (2001). *Beyond college for all: Career paths for the forgotten half*. New York: Russell Sage Foundation.
- Shollenberger, S. L. (2009). *Perceptions of gender and leader emergence in virtual groups*. Doctoral Dissertation.
- Suh, B., Chi, E. H., Pendleton, B. A., & Kittur, A. (2007). Us vs. them: Understanding social dynamics in Wikipedia with revert graph visualizations. In *2007 IEEE symposium on visual analytics science and technology* (pp. 163–170).
- Sulisworo, D. (2012). Designing the online collaborative learning using the Wikispaces. *International Journal of Emerging Technology and Learning (IJET)*, 7(1), 58–61.
- Tach, L. M., & Farkas, G. (2006). Learning-related behaviors, cognitive skills, and ability grouping when schooling begins. *Social Science Research*, 35(4), 1048–1079.
- van de Rijt, Arnout, Kang, Soong Moon, Restivo, Michael, & Patil, Akshay (2014). Field experiments of success-breeds-success dynamics. *Proceedings of the National Academy of Sciences*, 111(19), 6934–6939.
- Van Knippenberg, D., & Schippers, M. C. (2007). Work group diversity. *Annual review of psychology*, 58, 515–541. <http://dx.doi.org/10.1146/annurev.psych.58.110405.085546>.
- Wade, C. E., Cameron, B. A., Morgan, K., & Williams, K. C. (2011). Are interpersonal relationships necessary for developing trust in online group projects? *Distance Education*, 32(3), 383–396.
- Walker, H. A., Ilardi, B. C., McMahon, A. M., & Fennell, M. L. (1996). Gender, interaction, and leadership. *Social Psychology Quarterly*, 255–272.
- West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender & Society*, 1(2), 125–151. <http://dx.doi.org/10.1177/0891243287001002002>.
- Wilkinson, D. M. & Huberman, B. A. (2007). Assessing the value of cooperation in Wikipedia. In *Proceedings of the 2007 international symposium on Wikis, vol. 12* (pp. 157–164). ACM.
- Woolley, A., Chabris, C., Pentland, A., Hashmi, N., & Malone, T. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004), 686–688.
- Yang, Y., Cho, Y., Mathew, S., & Worth, S. (2011). College student effort expenditure in online versus face-to-face courses the role of gender, team learning orientation, and sense of classroom community. *Journal of Advanced Academics*, 22(4), 619–638.
- Zelditch, M. Jr., (1955). Role differentiation in the nuclear family: A comparative study. In T. Parsons & R. F. Bales (Eds.), *Family, socialization and interaction process* (pp. 307–352). Glencoe, IL: Free Press.
- Zhu, H., Kraut, R. E., & Kittur, A. (2013). Effects of peer feedback on contribution: A field experiment in Wikipedia. In *CHI'13: Proceedings of the 2013 annual conference on human factors in computing systems*. New York: ACM Press.