Journal of Economic Behavior and Organization 203 (2022) 543–562

Contents lists available at ScienceDirect

Journal of Economic Behavior and Organization

journal homepage: www.elsevier.com/locate/jebo

Information exchange and multiple peer groups: A natural experiment in an online community[☆]

Lingqing Jiang[a,*], Zhen Zhu[b]

[a] Department of Economics, University of Essex, Wivenhoe Park, CO4 3SQ, United Kingdom
[b] University of Kent, Giles Ln, Canterbury CT2 7NZ, United Kingdom

ARTICLE INFO

Article history:
Received 4 January 2022
Revised 15 September 2022
Accepted 22 September 2022

JEL classification:
J13
J18
C36

Keywords:
Peer groups
Information exchange
Natural experiment
Instrumental variables
Online communities

ABSTRACT

We utilise a quasi-experimental setup to identify causal effects of having additional peer groups on information exchange in a large online maternity community. The information exchange is a key performance indicator for the community as well as a public good among users. Pregnant users join default peer groups based on estimated due date (EDD). Natural uncertainties of EDD can lead to multiple peer groups. Using EDD as an instrumental variable, we find that additional peer group(s) reduces information exchange in both default peer group and total peer groups. Having more advanced groups mitigates the reduction, likely due to information spillovers.

1. Introduction

It is increasingly common for individuals to be members of multiple groups simultaneously, both in organisations and in social life. Firms adopt multiple-team structure to leverage resources more effectively and promote knowledge transfer (Milgrom and Roberts, 1992); researchers collaborate with different sets of coauthors to exploit synergies in expertise and knowledge spillovers (Borjas and Doran, 2015); individuals expand their social groups to connect social ties and relations (Granovetter, 2018). While there are clear incentives and benefits to access additional groups and resources both in organisations and in social life, there are also challenges as it creates competing pressures on attention and time devoted among groups.[1]

Identifying the causal effects of having additional peer groups on relevant outcomes is challenging as the choices of joining which and how many groups are often endogenous (Heckman, 1990; Evans et al., 1992). For example, individuals who have higher abilities may be invited to join additional working units; individuals who are endowed with more social

☆ The authors would like to thank Bo Honore, Sonia Bhalotra, Jordi Brandt, Gabriella Conti, Thomas Cornelissen, Monica Costa Dias, Subhasish Chowdhury, Francesco Drago, Marco Francesconi, Lorenz Goette, Angus Holford, Marco Manacorda, Friederike Mengel, Marie Claire Villeval, Christian Zehnder, and participants at the conferences of Royal Economic Society 2019, Reading Behavioural Workshop 2019, ISER internal seminar, General Online Research 2019, NetMob 2019, and IC2S2 2019 for their insightful comments. We obtained ethical approval for this study from the University of Essex.

* Corresponding author.
E-mail addresses: lingqing.jiang@essex.ac.uk (L. Jiang), z.zhu@kent.ac.uk (Z. Zhu).

[1] See Margolis (2020) for a meta analysis of multiple team membership with 44 studies.

https://doi.org/10.1016/j.jebo.2022.09.019
0167-2681/

resources may have access to additional social groups. Although the causal effects of having peers or peer groups have been identified using randomised group assignment or instrumental variable approach in various individual behaviours and contexts, to our knowledge, the effects of having *additional* peer groups on individual outcomes are rarely investigated beyond correlational studies. Ignoring the potential endogeneity may overestimate the positive effects or overlook potential negative effects, which could mislead individual choices or policy implications.

Our study exploits a real-world setup in a large online maternity community. First of all, this online maternity community adopts a group-based structure and the group assignment resembles a quasi-experimental approach which allows clean identification. Moreover, such online maternity communities are emerging in countries all over the world including Asia, Europe, and the US.[2] Therefore, they are of general interest per se. The community automatically assigns pregnant users into different peer groups based on the month of their estimated due date (EDD) at registration. For example, users who are expecting to give birth on any day in June 2021 are assigned to the peer group (aka *birth club*) named *June 2021*. Each month, a new peer group is created and new users can join. Such assignment rules are common in real life. The closest example would be that schools assign pupils to different grades according to the year of their birth.

Under the default group assignment rule, we observe a non-trivial fraction of users who manually join additional peer group(s), which is partially driven by the natural uncertainty of EDD.[3] The standard deviation of EDD is about ± 15 days (Hoffman et al., 2008). Users seem to be aware of this fact and internalise it as we observe a clear pattern that those whose EDD falling in the beginning (end) of the month tend to additionally join the peer group of the previous (next) month. This pattern allows us to use the day of the EDD as the instrument for having multiple peer groups as i) it directly affects the propensity of joining one of the adjacent peer groups; and ii) it is a natural lottery (Angrist, 1990).

The outcome of interest is users' information exchange in the peer groups on this online community. The creation of peer groups provides a common space for users to exchange information and knowledge among themselves throughout the entire pregnancy. The users typically exchange information about their pregnancy status and symptoms, which often involves sharing check-up results, updating progress, asking for confirmation, consultation and help from other peers in the group.[4] The exchange information in the peer groups takes the form of initiating a post or responding to a post. A post typically initiates information exchange by either sharing own experience or asking a question; a response typically continues the information exchange by either answering a posted question or asking follow-up questions. For the online community, active information exchange is of crucial importance to its business success. The volume of the user-generated traffic is a key performance indicator and needs to be closely monitored in order to sustain its position in the market and to fulfill the expectations of the stakeholders. For the users, the information exchange among themselves serves as a public good and constitutes the essence of social support (Jiang and Zhu, forthcoming), defined as *interpersonal exchange of potentially useful information or things* (Cohen and Syme, 1985).

We focus on two important mechanisms that could make a difference between having one single peer group and having multiple groups in the context of information exchange. The first is the mechanical substitution and the second is the asymmetric information flow driven by time-sensitive information. On the one hand, under the mechanical substitution, the two-peer-groups users simply spread their activities into two peer groups without differentiating between peer groups. Consequently, it would cause the reduction in the information exchange in either peer group and the reduction should be the same for joining either the earlier group or the latter group. On the other hand, the peer groups are created along the timeline and the information sets expands with the pregnancy status. Given the time difference, user who join the earlier group is exposed to peers who are more advanced in the pregnancy status and has access to time-sensitive information in advance comparing to user who join the later group. This could have two potential consequences: one is accessing something new as an activator which could boost the information exchange in the default group, and the other is causing a potential strategic shift to the earlier group which could further reduce the information exchange in the default group.

We perform the main empirical analysis using the observations that have either the default peer group only (control) or additionally join one of the adjacent peer groups (treated), i.e. the previous-month or the next-month peer group. We then extend the analysis to the following: using quality posts only as the outcome, investigating potential heterogeneity by comparing the previous-month peer group with the next-month peer group to shed light on the underlying mechanism, breaking information exchange into posts and responses, and finally generalising to the case of multiple (≥ 2) peer groups.

The instrumental variables estimation yields three main findings. First, joining an additional peer group (or additional peer groups) significantly reduces users' information exchange in the default peer group. Second, the sum of the information exchange generated by the users having two (or multiple) peer groups is less than the sum generated by the users having only one default peer group. Third, the reduction in the information exchange in the default peer group is smaller when the additional peer group is the previous-month group comparing to the next-month group.

[2] A similar counterpart website to ours is the BabyCenter. It is available in the UK, US, Arabia, Australia, Brazil, Canada, Germany, India, and in Spanish language (without countries specified). Another example is Mumsnet. It is currently the UK's biggest network for parents. The common goal of these online communities is to "make parents' lives easier by pooling knowledge, advice and support", as stated on the website.

[3] There are also non-random motives to join additional groups, for example, users want to share information with a greater number of users, access information from more users, or compare information across different peer groups.

[4] Fig. A1 in the Appendix presents the 20 most frequent words mentioned in the information exchange. Apart from the most frequent words "pregnant, baby, and mom(s)", other frequently discussed topics relate to "last menstrual period, four-dimensional ultrasound, belly size, fetal heart, fetal pole, and gender". The verbs such as "whether (be), take a look, and help" are clear demonstration of asking for confirmation, consultation, and help from peers.

We interpret our first finding as a substitution effect between the additional peer group(s) and the default peer group, which is in line with the limited attention literature (e.g. Kahneman, 1973; Falkinger, 2008). Our second finding, if assuming linear production function of information exchange, suggests that there could be further "detrimental" effects besides the substitution effect. One possible explanation is that having additional peer groups weakens the group identity of any group, which can reduce contribution or encourage "lurking" behaviour in each peer group.[5] We interpret our third finding as evidence for information spillovers from more advanced peer group to the less advanced ones. However, the spillover effect is dominated by the substitution effect as the net effect is negative.

Our findings have important policy implications. From the perspective of the online community, our results suggest that organisations and communities with similar features be cautious when designing the group-based structure. Allowing users to join multiple peer groups may cause them to partially shift their activities from the focal group to other groups. Moreover, it could dampen their overall engagement and lower the total amount of "user-generated information", which is in opposition to the objective function of the community. From the perspective of the users, they benefits from active information exchange which serves as a public good in each peer group. A decline in the activeness in the peer group may cause negative experience to them.[6] Our findings have direct implications for online maternity communities that adopt the same concept and group-based structure in other countries in Europe and the US and possibly organisations with a vertical group-based structure as well. It might also be suggestive for information-sharing networks where individuals group together with common attributes (e.g. interests and job profiles) for a common goal (e.g. education and self-development), although such networks are more likely to have a horizontal structure.

Our paper mainly contributes to two strands of literature. First, we contribute to the vibrant literature on peers effects. Peers and peer groups are found to play an important role in productivity at work (Mas and Moretti, 2009; Borjas and Doran, 2015; Cornelissen et al., 2017), education (Sacerdote, 2001; Carrell et al., 2009; Elsner et al., 2020), welfare use (Åslund and Fredriksson, 2009), retirement planning (Duflo and Saez, 2002; Brown and Laschever, 2012), prosocial behaviour (Gächter et al., 2013; Bruhin et al., 2020), and sports competition (Yamane and Hayashi, 2015; Jiang, 2020).[7] Among these, Mas and Moretti (2009), Borjas and Doran (2015), Sacerdote (2001) and Kimbrough et al. (2020) demonstrate spillovers from high productivity workers, researchers, and students to their low productivity peers, respectively. In particular, using a lab experiment, Kimbrough et al. (2020) highlight the importance of peer-to-peer teaching in learning and suggest that grouping students of similar abilities may hurt low-ability individuals as they could benefit from interacting with high-ability peers.

Second, we contribute to the rising but yet scarce literature of multiple group membership that mostly focuses on working contexts. The evidence of having multiple groups on productivity is mixed. While some studies find positive effects (O'leary et al., 2011; Zhu et al., 2014), more studies find negative effects on individual and/or group outcomes (Cummings and Haas, 2012; Pluut et al., 2014; Mortensen and Haas, 2018; Crawford et al., 2019). O'leary et al. (2011) and Bertolotti et al. (2015) suggest that the relationship between the number of groups and productivity is an inverted-U shape. That is, individuals benefit from learning in multiple groups when the number of additional groups is small. However, as the number increases, it starts to fragment attention and introduce lags in productivity.[8]

There are several distinctions between our study and the existing literature on multiple group membership. First of all, to our knowledge, we are the first study that utilises a quasi-experimental setting to provide causal evidence beyond the existing correlational studies. Second, while this literature focuses on relatively small working teams, we look at significantly larger peer groups in a non-working context. Last but not least, the peer groups in our setup are online-based. With the development of digital technologies, more and more traditional offline working and social units have found their online counterparts, which calls for more research in online settings.

The remainder of the paper is organised as follows, Section 2 describes the observational data, Section 3 illustrates the identification challenges and strategy, Section 4 discusses behavioural predictions, Section 5 presents the empirical analysis, and Section 6 concludes.

2. Data

We collect our data from one of the largest Chinese online maternity and parenting communities. We first introduce the online community. Subsequently, we explain the assignment of the default peer group and the occurrence of multiple peer groups. Finally, we present the descriptive statistics of the peer groups in our sample.

[5] Lurking is very common in online communities. It is related to the free riding behaviour in the sense that lurking users solely receive information without generating any information. Research has shown that stronger feelings about group identity increase the level of contribution in the public goods games (Charness et al., 2014).

[6] A continuous information flow is also important for the group as a whole. Given the fast turnover of information exchange in online settings, some of the information can be repetitive or similar. For example, different users may ask similar questions on different days without being aware that the questions have been already asked and answered before. Despite efficiency loss to some extent, having peers answering similar questions repeatedly from time to time can reinforce group cohesiveness.

[7] Studies also show that introducing group identity could alter individual preferences (Charness and Chen, 2020; Goette et al., 2006; Charness et al., 2007; Chen and Li, 2009; Ghiglino et al., 2021).

[8] The literature does not conclude with a definitive consensus of the number of groups where this tipping point occurs that highly depends on the context and factors such as the social ties among the group members (van de Brake et al., 2020), the variety of the groups (O'leary et al., 2011), the collaborative technology (Bertolotti et al., 2015), and the task complexity (Crawford et al., 2019).

Fig. 1. Creation and enrollment of peer groups.
Notes: Peer groups are labelled in the format of *Month Year* – the month and the year of the EDD. Users whose EDDs are in the *Month* and *Year* are assigned to the peer group *Month Year* by default. Peer groups are ordered by month along the timeline.

2.1. The online community

At its root, the online maternity and parenting community is built upon dedicated peer groups (aka *birth clubs*) at monthly frequency. Each peer group is essentially an Internet Forum with an allocated URL within the domain of the community platform.[9] Users of such a peer group can hold conversations (aka *threads*) by generating posts and responses. On a daily basis, the users in each group may generate hundreds of threads with different conversations/topics, which yields tens of thousands threads throughout the lifetime of the peer groups. All threads are listed in reverse-chronological order, and can be accessed to and followed up by all users at any time. The threads/conversations do not build on each other and some of the popular topics may appear again and again over time.[10] In Section B.3 we show two screenshots of the posts to give an idea of how the interface looks like.

The community also provide other general-topic groups (e.g. sharing food recipes) besides the peer groups. In addition to the default peer group, users can manually join any of these groups after registration. However, most of the activities take place in the peer groups.[11]

The company that operates this online community is a for-profit organisation. It conducts advertising, electronic commerce, content monetisation, and other businesses for revenue sources. However, to sustain its business model and to secure future funding from investors, the priority is to engage customers through its core services by providing a platform where users can share positive and meaningful experiences about maternal caring and parenting. Therefore, user-generated information exchange is key for the sustainability of this online community.

2.2. Assignment of default peer groups

Pregnant users are assigned into peer groups based on the month of their estimated due date (EDD). Grouping users who are at the same stage of pregnancy helps them exchange information and knowledge more efficiently in three ways. First, it keeps certain time-sensitive information timely. Second, since users in the peer groups are both information givers and recipients, they have a high level of responsibility and trust among themselves to ensure the quality of information. Last but not least, it keeps the number of users in each peer group more or less balanced.

Peer groups are named in the format of *Month Year* – the month and the year of the EDD. We illustrate the three core peer groups in our sample in Fig. 1. For example, users who are expecting to give birth on any day in March 2018 were assigned to the peer group *March 2018*. All the peer groups are created ten months before the delivery month, e.g. the peer group *March 2018* was created in May 2017. Each peer group is active since the creation till the end of the corresponding delivery month.

Notice that within the three-month span of expected delivery, the assignment of the default peer group can be seen as quasi-random, as users do not have perfect control over pregnancy within such a short period. In other words, apart from the pregnancy status, users' other individual characteristics do not significantly differ across peer groups.

2.3. Multiple peer groups

The feature of peer groups is designed such that each user has only one peer group by default. However, we observe that a non-trivial fraction of users join more than one peer groups in practice. After the assignment of the default peer group,

---

[9] For example, http://domain-name/community/club201803/ is for the peer group of March 2018.

[10] There are a few moderators monitoring the information exchange in the peer groups. They do not engage in any information exchange in the peer groups. The moderators only intervene when there are issues reported by the users such as verbal violence or spreading inappropriate information by removing such posts and responses.

[11] We use the statistics in Table 1 to calculate the share of activities. We calculate the total activities in peer groups by multiplying column (1), average number of posts and number of responses, by column (6), number of users, and denote it as A. Similarly, we calculate the total activities in non-peer groups by multiplying column (1), average number of posts and responses, by column (6), number of users, and denote it as B. Then the share of the activities in peer groups out of the whole activities is given by $\frac{A}{A+B} = 68.7\%$. We also calculate the share at individual level (i.e. individual peer group activities divided by the whole activities), which gives an even higher share of 76.2%.

Table 1
Descriptive statistics of users in the three core peer groups.

Variables	(1) mean	(2) sd	(3) p10	(4) p50	(5) p90	(6) # Obs
Day of Estimated Due Date (EDD) (1st–30th/31st)	15.10	8.626	3	15	27	24,705
Peer Group Enrollment (share)						
Default PG_T	100%	0	1	1	1	24,705
Previous month PG_{T-1}	10.9%	0	1	1	1	2692
Next month PG_{T+1}	6%	0	1	1	1	1485
Number of Posts in						
Default PG_T	2.079	1.992	1	1	4	24,705
Previous month PG_{T-1}	0.290	0.825	0	0	1	2692
Next month PG_{T+1}	0.272	0.919	0	0	1	1485
Number of Responses in						
Default PG_T	3.679	7.185	0	1	11	24,705
Previous month PG_{T-1}	1.105	2.765	0	0	3	2692
Next month PG_{T+1}	0.882	2.558	0	0	2	1485
Non-Peer Groups						
Number of non-PGs joined	9.865	3.371	4	10	13	24,705
Number of Post and Responses in non-PGs	2.718	5.261	0	1	8	24,705

users may manually join other peer groups. To join an additional peer group, a user needs to manually search for the peer group on the platform and click on the "join" button to confirm the official peer group membership. There is no restriction to join, and it does not involve any approval process.[12]

When a user logs in to her account, she will see an interface that only displays the peer group(s) and the general groups that she has officially joined.[13] Moreover, after a user has officially joined a peer group, she will receive notifications of new posts in that peer group. In other words, if a user does not officially join an additional peer group, she will not see that peer group on her interface nor will she receive any notifications of new posts in that peer group.

There could be various motives to join additional peer groups: users may want to exchange information with a greater number of users, with users in different pregnancy statuses, or compare information across different peer groups. Another important reason is related to the EDD. The standard deviation of EDD is about ±15 days (Hoffman et al., 2008). Users seem to be aware of this fact and internalise it as we observe a clear pattern that users whose EDD falling in the beginning (end) of the month tend to additionally join the peer group of the previous (next) month. We provide more details of this pattern later on in Fig. 3 in Section 3.1.

2.4. Descriptive statistics

Our data include three core peer groups with a complete cycle, i.e. from the creation of the peer group to the end of the delivery month.[14] They are *March 2018, April 2018*, and *May 2018*. We have information about users' EDD, the enrollment of both peer groups and non-peer groups, the number of each user's posts and responses in the default peer group, the additional peer groups and non-peer groups. We do not observe many demographic characteristics apart from that all users are women at fertility age.[15]

We present the descriptive statistics of the sample that we use for the main empirical analysis. It includes the observations that join only the default peer group as the "control" group and the observations that additionally join one of the adjacent peer groups as the "treated" group. Table 1 presents the summary of the three core peer groups pooled together.[16] There are 24,705 unique users – who are either *single-default-peer-group* users or *two-peer-group* users who additionally joined one of the adjacent peer groups – in these three core peer groups. The day of EDD (1st–30th/31st) follows ap-

[12] In theory, the users could also enter a non-default peer group and browse the posts and responses in that group without officially joining it. Unfortunately, we could not observe such informal joining behaviour in the data. This means, we could only capture all the official multiple-peer-group users but not the informal ones. As a result, our empirical comparison between the single-peer-group users and the multiple-peer-group users is restricted to those who officially joined the peer groups. Omitting the informal joining behaviour would lead to the following consequence: we treat those informal multiple-peer-group users as the official single-peer group users, while the status of the official multiple-peer-group users is unaffected. If these informal multiple-peer-group users are more active than the official single-peer-group users, then our estimates would be downward biased, and vice versa.

[13] We show an interface screenshot of a user who has joined two peer groups in Fig. B.4 in the Appendix B.3.

[14] We exclude users who are not giving birth for the first time and users who have never generated any post or response throughout the entire pregnancy. Additionally, we trim the top 5% of our data in terms of the number of posts and responses to exclude outliers, which is a standard practice with communication data (Turkiewicz, 2018).

[15] Age and education might affect one's activeness in the peer groups and adding them would increase the precision. Unfortunately, we do not have the info of these two variables as they are not required at the registration.

[16] Słoczyński (2020) emphasises the concern about differential weights on each group in case there is heterogeneity across groups, if group fixed effects are included in a linear IV estimator. We present the descriptive statistics of each peer group in Table A1 in the Appendix. Apart from the natural fluctuation in the number of observations in each peer group, there is no significant difference in the instrument variable and the endogenous variables across peer groups.

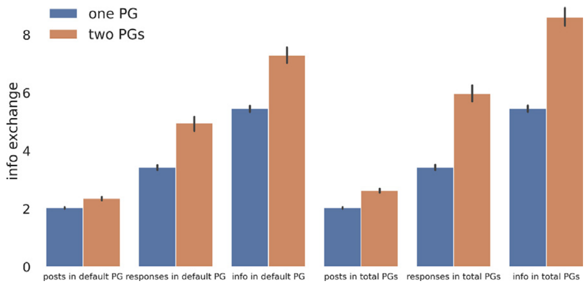

Fig. 2. Information exchange by number of peer groups.

proximately a uniform distribution with a mean of 15.1 and a standard deviation of 8.6. We first look at the peer group enrollment. By design, all users have a default peer group, which we denote as PG_T. There are 2692 users (10.9%) who additionally join the peer group of the month just before their EDD, which we denote as PG_{T-1}; and there are 1485 users (6%) who additionally join the peer group of the month just after their EDD, which we denote as PG_{T+1}.

Next we look at the information exchange – the number of posts and responses – in the peer groups. On average, 2.08 posts per user (or a total of 51,362 posts) are generated in the default peer groups, 0.29 (or a total of 781 posts) in the previous month peer groups (if joined), and 0.27 (or a total of 404 posts) in the next month peer groups (if joined). Users are more active in generating responses than posts. On average, 3.68 responses per user (or a total of 90,890 responses) are generated in the default peer groups, 1.1 (or a total of 2975 responses) in the previous month peer groups (if joined); and 0.88 (or a total of 1310 responses) in the next month peer groups (if joined).

Besides the peer groups users can also join non-peer groups that are open to all users without the reference of EDD.[17] On average, users have 9.9 non-peer groups and generate about 2.7 posts and response (or a total of 67,149) in non-peer groups.

Notice that we could not capture lurking behaviour in our data. That is, users do not generate any posts or responses but still have access to other users' information exchange. Therefore, our information exchange account is only restricted to active posting and responding. Any passive form of information absorption is not captured in our analysis.

Finally, we present a descriptive graph in Fig. 2 to show the number of posts, responses, and total info exchange per user in the default peer group (left) and total peer groups (right), by the number of peer groups: one peer group in blue and two peer groups in brown. This descriptive graph would correspond to the OLS regressions in Columns (2) and (4) in Table 5. Joining an additional peer group is associated with more posts, responses, and total information exchange in the default peer group as well as in total peer groups.

3. Identification

The empirical challenges lie in the endogenous choice of which and how many peer groups to join. In principle, users can manually join any peer groups that are open in the community.[18] Therefore, simply comparing *single-default-peer-group* users with *two-peer-group* users is subject to the omitted-variable bias. For instance, users who are more active in general or need more information are more likely to join an additional peer group. If this is the case, the naive comparison will display a positive correlation between joining additional peer groups and users' overall activeness. We first demonstrate our identification strategy, and subsequently, we present the placebo checks.

3.1. Identification strategy

Our identification strategy relies on the community's initial peer group assignment described above and the fact that the day of EDD is a natural lottery. Two desirable features of the community's setting allow us to overcome the two empirical challenges: i) which peer group to join and ii) how many peer groups to join.[19] First, all users are automatically assigned to a default peer group based on the estimated delivery month, and therefore, the assignment of default peer group can

[17] The non-peer groups have particular themes such as "Marriage","Sentiment", "Cooking" etc. They are not in the format of *Month Year* and the enrollment is completely unrelated to the EDD. Therefore, users in such groups are much more heterogeneous in their maternal status. In our sample, on average a user spends 74.3% of her activities in the default peer group, indicating that the default peer group is the principal group for the users in this online community.

[18] As mentioned before, the range of such peer groups featured on the homepage is one year before and ten months after the current month. This asymmetry does not affect our identification. In the main empirical analysis we only look at additional peer group in the adjacent months. When we extend the analysis of multiple peer groups without any restriction, the results remain robust.

[19] It is important to notice that since we are only comparing two peer groups vs. one (default) peer group, our identification challenge is the endogenous group enrollment rather than the *reflection problem* in Manski (1993). In other words, we are interested in the $\hat{\beta}$ in $Y_i = \beta(\text{Two PGs})$ instead of $Y_i = \beta\bar{Y}_{-i}$. Notice that such comparison captures the direct interaction effects while assuming the indirect interactions equal for all users.

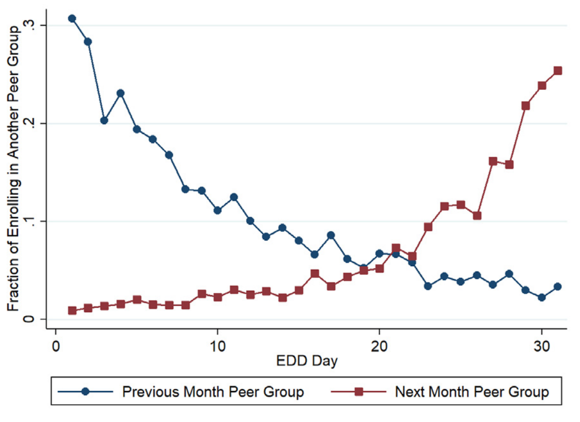

Fig. 3. Fractions of users joining the adjacent peer groups, by EDD day.
Notes: The figure shows the fraction of users joining the previous-month peer group and the next-month peer group, respectively, on each day of the month in the three core peer groups (March, April, and May 2018).

be seen as quasi-random within the narrow period we study. Second, whether users join an additional peer group partially depends on the day of EDD.

Fig. 3 shows the fraction of the users joining an additional peer group of an adjacent month for each day of the month. The day of EDD directly affects the propensity of joining an adjacent peer group. The fraction of the users additionally joining the previous-month peer group is above 30% for users whose EDD is on the 1st day of the month, and it gradually decreases to about 5% for the users whose EDD is on the last day of the month. Almost like a mirror image, the fraction of the users additionally joining the next-month peer group is barely 1% for the users whose EDD is on the 1st day of the month, and it gradually increases to above 25% for users whose EDD is on the last day of the month.

The day of the EDD should not be correlated with users' other characteristics or affect the information exchange directly as it can be seen as a natural lottery ticket: each pregnant user draws a number from 1 to 30/31. Those who draw a small number may anticipate that a slightly earlier delivery would end up delivering in the month before, and therefore, are more likely to additionally join the peer group of the previous month. Similarly, those who draw a large number may anticipate that a slightly later delivery would end up delivering in the month after, and therefore, are more likely to additionally join the peer group of the next month.[20] A slightly later (earlier) delivery for an EDD in the beginning (end) of the month, or a slightly earlier or later delivery for an EDD in the middle of the month would not change the month of delivery, and therefore, is less likely to trigger the enrollment of an additional peer group.[21]

Therefore, the day of EDD is a valid instrument for having multiple peer groups as i) it directly affects the propensity of joining one of the adjacent peer groups; and ii) it is a natural lottery.

Fig. 4 shows the first stage which corresponds to Fig. 1 in Angrist and Chen (2011). As we can see, the connected line is flat at the beginning, i.e. around the middle of the month, and becomes more steep towards the end, i.e. either at the beginning or the end of the month. Overall, the relationship is quite linear.

Our identification assumption is that the day of EDD has an effect on users' information exchange only through its effects on joining an additional peer group. Notice that using the instrumental variable estimator we are identifying the local average treatment effect (LATE), which is the average treatment effect (ATE) among the compliers (Imbens and Angrist, 1994).[22]

3.2. Placebo checks

To further support the assumption that the day of EDD does not correlate with any unobserved characteristics, we perform a placebo check using two placebo outcomes: the number of non-peer groups joined and the amount of information

[20] The monotonicity assumption claims that "while the instrument may have no effect on some people, all of those who are affected are affected in the same way." (Imbens and Angrist, 1994), and "A failure of monotonicity means the instrument pushes some people into treatment while pushing others out." (Angrist, J.D. and Pischke, J.S., 2009). Just like the draft-lottery in Angrist (1990), the day of the EDD satisfies the monotonicity assumption: although having an EDD close to the two ends of the month may have had no effect on the probability of joining an additional peer group for some users, there is no one who was actually kept out of joining an additional peer group by having EDD close to the two ends of the months. This holds by construction as peer groups are open to everyone.

[21] Notice that users do not have any incentives to manipulate their EDD in order to join an additional peer group, i.e. users who want to join an additional peer group report their day of their EDD either at the beginning of the month or end of the month, since nothing prevents users from joining additional peer groups. We also show that the number of users who expect to deliver on each day of the month follows approximately a uniform distribution in Fig. A2 in the Appendix.

[22] To give an idea of how many compiers there are, how representative the LATE might be, and how similar they are to always- and never-takers, we show the share of compliers and the average outcomes of different types of uses in Table A2 in the Appendix.

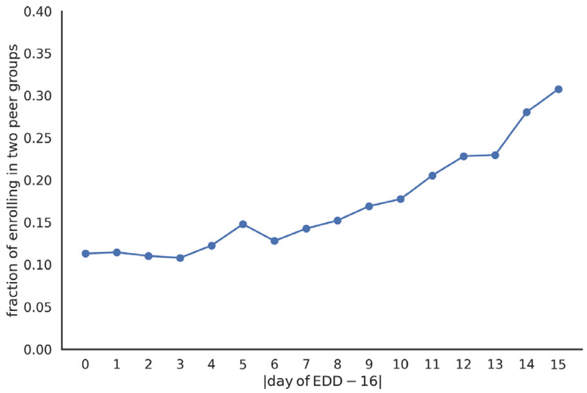

Fig. 4. Fractions of users joining two peer groups, by the instrument |EDD day − 16|.
Notes: The figure shows the fraction of users additionally joining either the previous-month peer group or the next-month peer group, respectively, on each value of the instrument in the three core peer groups (March, April, and May 2018).

Table 2
The effects of the day of EDD on placebo outcomes.

Variables	(1) Number of NPGs OLS	(2) log(info) in NPGs OLS	(3) Number of NPGs Poisson	(4) Info in NPGs Poisson
|Day of EDD-16|	7.06e−05	−0.00167	7.15e−06	−0.000642
	(0.00485)	(0.00131)	(0.000492)	(0.00280)
Constant	9.909***	0.809***	2.293***	0.999***
	(0.0514)	(0.0139)	(0.00519)	(0.0294)
Peer group fixed effects	Yes	Yes	Yes	Yes
Observations	24,705	24,705	24,705	24,705
R-squared	0.0032	0.0006	0.0007	0.0001

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

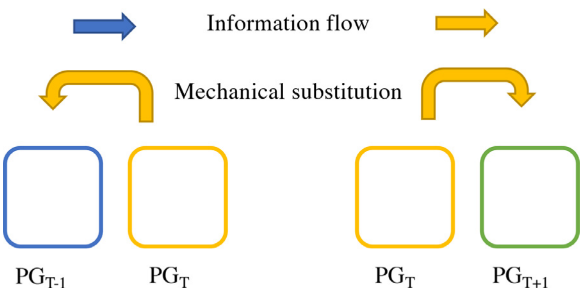

Fig. 5. Two potential effects of joining an additional peer group.

exchange in non-peer groups. Unlike the peer groups that are exclusively designed for pregnant users, non-peer groups are open to all users (not yet pregnant, pregnant, and mothers of young kids) in this community. These two outcomes are informative about users general activeness. Since the enrollment of non-peer groups is completely unrelated to the EDD, we should not observe any correlation between the day of EDD and the enrollment of non-peer groups or the activities in those groups. Table 2 confirms this in both OLS and Poisson specifications.

4. Behavioural predictions

We make behavioural predictions by focusing on two mechanisms: the mechanical substitution and the asymmetric information flow driven by time-sensitive information. We illustrate these two mechanisms in Fig. 5.

On the one hand, additional groups could increase the information load and switching costs, and decreases attention to new information and time to encode information (Margolis, 2020; O'leary et al., 2011). The limited attention literature (e.g. Kahneman, 1973; Falkinger, 2008) emphasises that processing massive amount of information is often beyond human beings' capacity of attention. In our context, users who joined an additional peer group – that doubles the number of peers and the amount of available information – face a trade-off in allocating their time and attention in exchanging information between peer groups. Under the mechanical substitution, the two-peer-groups users simply spread their activities into two

Table 3
Behavioural predictions.

Mechanical substitution	Asymmetric information flow
Info in $PG_T \downarrow$	New info ($PG_T \uparrow$) : $(PG_{T-1} \Rightarrow PG_T) > (PG_T \Leftarrow PG_{T+1})$
Info in total PGs $\equiv$	Strategic shift ($PG_T \downarrow$) : $(PG_T \Rightarrow PG_{T-1}) > (PG_T \Rightarrow PG_{T+1})$

peer groups without differentiating between peer groups. Consequently, it would cause the reduction in the information exchange in either peer group and the reduction should be the same for joining either the earlier group or the latter group.[23] Under the assumption of a linear production function of information generation, the sum of information exchange in the total peer groups would be the same as the one from the single-default-peer group users.

On the other hand, the social network literature (e.g. Jackson, 2010) suggests that exposure to a wider social network promotes information diffusion and knowledge spillovers. In our setup, the peer groups are created along the timeline and the information sets expands with the pregnancy status.[24] By construction, at any given point of time, users in PG_{T-1} are on average two months ahead of PG_{T+1} in the pregnancy status. Let $i_{T,T-1}$ and $j_{T,T+1}$ denote two users whose default peer group is PG_T, the former additionally joined PG_{T-1} while the latter additionally joined PG_{T+1}, respectively. Given the time difference, user $i_{T,T-1}$ is exposed to peers who are more advanced in the pregnancy status and has access to time-sensitive information in advance comparing to user $j_{T,T+1}$.

This could have two potential consequences. First, user $i_{T,T-1}$ will have access to something "new" comparing to user $j_{T,T+1}$. Such "new" information spillover may serve as an activator of the information exchange in the default peer group and could partially mitigate the reduction of information exchange due to the mechanical substitution. Second, user $i_{T,T-1}$ may have more incentives to shift her activities to PG_{T-1}, resulting in less information in the default peer group. Such strategic shift, although with a different motive, would go in the same direction as the mechanical substitution. Therefore, we would expect a further reduction in the default peer group for user $i_{T,T-1}$.

Table 3 summarises our behavioural predictions of each mechanism for the two peer-group users. Under the mechanical substitution, the information in the default peer group will decrease but remains unchanged in total peer groups for both user $i_{T,T-1}$ and $i_{T,T+1}$. The asymmetric information flow would have two potential counteracting spillover effects which are different for user $i_{T,T-1}$ and $i_{T,T+1}$: one is receiving something new from the other group as an activator which could boost the information exchange in the default group (more from PG_{T-1} to PG_T than from PG_{T+1} to PG_T), and the other is causing a potential strategic shift to the other group which could further reduce the information exchange in the default group (more from PG_T to PG_{T-1} than from PG_T to PG_{T+1}).

5. Econometric analysis

We first show the main results in Section 5.1 using the observations that joined one of the adjacent peer groups. We show the effect of joining two peer groups on the information exchange in i) the default peer group, and ii) total peer groups. The baseline compared with is the amount of information exchange generated by the single-default-peer-group users. We then extend the analysis to the following: using quality posts only as the outcome in Section 5.2, investigating potential mechanism by comparing the previous-month peer group with the next-month peer group in Section 5.3, breaking information exchange into posts and responses in Section 5.4, and finally generalising to the case of multiple peer groups in Section 5.5.

5.1. Main results

We estimate the econometric model in Eq. (1).[29] The outcome variable Y_i is the log-transformed information exchange generated by user i in her peer group(s) during the entire cycle. The indicator $\mathbb{1}[\text{Two PGs}_i]$ is one if user i has two peer groups instead of one default peer group.[25] We also include fixed effects of the default peer group – equivalent to the month fixed effects – which are captured by the term α_{pg}.

$$\mathrm{Y}_i = \beta_1 \mathbb{1}[\text{Two PGs}_i] + \alpha_{pg} + \epsilon_i \tag{1}$$

Column (1) in Table 4 reports the first stage regressions in a linear probability model. We use |Day of EDD-16| to predict $\widehat{\text{Two PGs}}$, as the farther the day of EDD is away from the middle of the month, the more likely that a user joins an additional

[23] This is based on the assumption that total time is fixed as in the literature of multiple team memberships.

[24] To give an idea of how the topics in the peer groups evolve during the course of pregnancy, we show the word clouds by trimester in Section B.2. We follow the instruction of text analysis in Ferrario and Stantcheva (2022).

[29] The results are robust when adding controls for the number of non-peer groups (NPGs) joined by user and the log-transformed information exchange generated by user in the non-peer groups (see Table A4). The reduced forms of both models can be found in Table A3. As a further robustness check, we also estimate in a Poisson IV model and the results are qualitatively unchanged as shown in Table A5 in the Appendix.

[25] The results are robust when adding controls for the number of non-peer groups (NPGs) joined by user i and the log-transformed information exchange generated by user i in the non-peer groups (see Table A4). As a further robustness check, we also estimate in a Poisson IV model and the results are qualitatively unchanged as shown in Table A5 in the Appendix.

Table 4
Information exchange and two peer groups.

Variables	(1) IV 1st stage Two PGs	(2) OLS/ IV 2nd stages log(info) in Default PG OLS	(3) log(info) in Default PG IV	(4) log(info) in Total PGs OLS	(5) log(info) in Total PGs IV
$\widehat{\text{Two PGs}}$		0.232***	−0.719***	0.391***	−0.410***
		(0.0146)	(0.104)	(0.0147)	(0.102)
| Day of EDD-16|	0.0124***				
	(0.000554)				
Constant	0.0969***	1.474***	1.653***	1.471***	1.621***
	(0.00554)	(0.00946)	(0.0219)	(0.00946)	(0.0215)
Peer group fixed effects	Yes	Yes	Yes	Yes	Yes
F-tests of instrument	–	–	504.19	–	504.19
Observations	24,705	24,705	24,705	24,705	24,705

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

(adjacent) peer group. The coefficient of |Day of EDD-16| is positive and significantly different from zero at 1% level. A one-day deviation from the middle of the month increases the probability of having two peer groups by 1.24%.

Columns (2)–(5) in Table 4 show the estimated effects of having two peer groups on users' information exchange. Columns (2) and (3) show the information exchange in their default peer group, and Columns (4) and (5) in total peer groups (one group if having default PG only; two groups if having an additional PG). We estimate each effect in OLS and IV regressions, respectively. Having two peer groups is associated with 26% more information in the default peer group and 48% more information in total peer groups in the OLS specification (Columns (2) and (4)).[26] However, once we instrument the additional peer group, the coefficients become significantly negative ($p < 0.01$). In the IV specification (Columns (3) and (5)), having two peer groups leads to about 51% less information in the default peer group and 33% less information in total peer groups. Note that the null hypothesis of a weak instrument has been rejected with the *F*-test, which is consistent with the first stage results in Column (1). The direction of the biases in the OLS estimates for the variable of Two PGs is upward, which can be explained by that the unobserved activeness of users is positively correlated with both information exchange and the tendency of joining an additional peer group. Therefore, after the IV correction, the causal effect of joining an additional peer group is an decrease in users' information exchange in the default peer group as well as in the total peer groups.

5.2. Quality posts with at least one response

To explore the quality of information exchange, we look at the number of responses which could be a proxy measure of post quality. It has a similar flavour of the number of "thumbs up" which unfortunately is not a feature on this platform. We impose a condition on the posts of having at least one response. With the condition in place, the number of posts decreases from 51,373 to 42,972 (16.4% reduction) for default PG and decreases from 52,558 to 44,037 (16.2% reduction) for total PGs. Notice that 13.8% of the users have zero posts in their default PG (or 13.6% of the users if all the three peer groups are considered).

While all the coefficients in Table 5 have smaller magnitude, they remain qualitatively robust compared to the results using all the information exchange in the baseline results in Table 4.

5.3. Previous-month vs. next-month peer group

Subsequently, we look at potential heterogeneous effects of having the previous-month peer group, PG_{T-1}, versus the next-month peer group, PG_{T+1}, as the additional peer group.

We estimate the econometric model in Eq. (2). The outcome variable Y_i is the log-transformed information exchange generated by user i in her peer group(s) during the entire cycle. In this equation, we have two endogenous variables: joining the PG_{T-1} and joining the PG_{T+1} additionally. Thus, we use $f_1(EDD) = EDD$ and $f_2(EDD) = EDD^2$ as the instruments for the PG_{T-1} enrollment and PG_{T+1} enrollment, respectively. We also use an alternative functional form of the EDD as the instruments: $g_1(EDD) = |EDD - 16|$, if $EDD \leq 16$; $g_1(EDD) = 0$, if $EDD > 16$; and $g_2(EDD) = 0$, if $EDD \leq 16$; $g_2(EDD) = |EDD - 16|$, if $EDD > 16$.[27] As in the baseline analysis in Section 5.1, the baseline compared with is the information exchange generated by the single-default-peer-group users.

$$Y_i = \kappa_1 \mathbb{1}[\text{PG}_{T-1,i}] + \kappa_2 \mathbb{1}[\text{PG}_{T+1,i}] + \alpha_{pg} + \mu_i \tag{2}$$

Table 5
Information exchange: quality posts with at least one response.

Variables	(1) IV 1st stage Two PGs	(2) OLS/ IV 2nd stages log(posts) in Default PG OLS	(3) log(posts) in Default PG IV	(4) log(posts) in Total PGs OLS	(5) log(posts) in Total PGs IV
$\widehat{\text{Two PGs}}$		0.112***	−0.197***	0.183***	−0.0210
		(0.00926)	(0.0625)	(0.00956)	(0.0625)
|Day of EDD-16|	0.0124***				
	(0.000554)				
Constant	0.0969***	0.837***	0.895***	0.840***	0.878***
	(0.00554)	(0.00611)	(0.0132)	(0.00613)	(0.0131)
Peer group fixed effects	Yes	Yes	Yes	Yes	Yes
F-tests of instrument	–	–	504.19	–	504.19
Observations	24,705	24,705	24,705	24,705	24,705

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 6
Heterogeneous peer groups and information exchange.

Variables log(info) in	(1) Default OLS	(2a) Default IV_1	(2b) Default IV_2	(3) Total OLS	(4a) Total IV_1	(4b) Total IV_2
PG_{T-1} (κ_1)	0.311***	−0.558***	−0.560***	0.478***	−0.236**	−0.254**
	(0.0183)	(0.0990)	(0.103)	(0.0180)	(0.0973)	(0.102)
PG_{T+1} (κ_2)	0.0902***	−1.027***	−1.030***	0.232***	−0.695***	−0.718***
	(0.0214)	(0.126)	(0.132)	(0.0222)	(0.124)	(0.130)
Constant	1.472***	1.648***	1.649***	1.468***	1.613***	1.617***
	(0.00946)	(0.0209)	(0.0217)	(0.00945)	(0.0204)	(0.0213)
Peer group FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	24,705	24,705	24,705	24,705	24,705	24,705
R-squared	0.014	0.724	0.724	0.034	0.746	0.745
F-tests of instrument	–	254.52	252.17	–	254.52	252.17
χ^2 Test $\kappa_1 = \kappa_2$	–	$p < 0.001$	$p < 0.001$	–	$p < 0.001$	$p < 0.001$

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.
IV_1: We use $f_1(EDD) = EDD$ and $f_2(EDD) = EDD^2$ for PG_{T-1} and PG_{T+1}, respectively.
IV_2: We use $g_1(EDD) = |EDD - 16|$, if $EDD \leq 16$; $g_1(EDD) = 0$, if $EDD > 16$; and $g_2(EDD) = 0$, if $EDD \leq 16$; $g_2(EDD) = |EDD - 16|$, if $EDD > 16$ for PG_{T-1} and PG_{T+1}, respectively.

Table 6 shows the results. In the OLS specifications, joining either PG_{T-1} or PG_{T+1} is positively correlated with the information exchange in users' default peer group (Columns (1)) as well as in total peer groups (Columns (3)). The two different functional forms of instruments are both strong and deliver very similar IV estimates in Columns (2a/2b) and (4a/4b). After the IV correction, having PG_{T-1} leads to 43% less information ($p < 0.01$) and having PG_{T+1} leads to almost 64% less information ($p < 0.01$) in the default peer group. These results are consistent with the main results in Table 4 where we pooled the PG_{T-1} and PG_{T+1} together. Moreover, the reduction is significantly smaller when users additionally joining PG_{T-1} than in PG_{T+1} (χ^2 Test $\kappa_1 = \kappa_2$, $p < 0.001$). When we aggregate all the information exchange of the two-peer-group users in both peer groups they joined, having PG_{T-1} leads to almost 20% less information ($p < 0.05$) and having PG_{T+1} leads to almost 50% less information ($p < 0.01$) in total. Again, the reduction is significantly smaller when users additionally joining PG_{T-1} than in PG_{T+1} (χ^2 Test $\kappa_1 = \kappa_2$, $p < 0.001$).

While we cannot say much about the existence of the strategic shift as we only observe the net effect of these two counteracting effects due to the asymmetric information flow, our results support the prediction of receiving new information from the more advanced to the less advanced peer groups. Despite that joining an additional peer group leads to an reduction in both default and total peer groups, having access to a more advanced peer group (comparing to a less advanced one) mitigates the reduction.

5.4. Breaking information into posts and responses

Next, we break the information exchange into posts and responses. The main motivation of looking at the posts and responses separately is the first intuition that users might be more likely to ask questions in the posts and answer questions in the responses. To verify this intuition, we compute the share of posts that contain any question indicators. The details of

[26] Since the dependent variable is log-transformed, the estimated effects = $(e^{\beta} - 1) \times 100\%$.

[27] We thank Bo Honore for suggesting the alternative instruments. The first stages of both sets of instruments can be found in Table A6 in the Appendix.

Table 7
Separating posts and responses: IV regressions.

Panel A: Two peer groups				
Variables	(1) Posts in default PG	(2) Posts in total PG	(3) Responses in default PGs	(4) Responses in total PGs
Two PGs	−0.144***	0.0302	−0.987***	−0.668***
	(0.0528)	(0.0531)	(0.130)	(0.128)
Constant	1.031***	1.014***	1.121***	1.086***
	(0.0112)	(0.0112)	(0.0274)	(0.0270)
Peer group fixed effects	Yes	Yes	Yes	Yes
F-tests of instrument	504.19	504.19	504.19	504.19
Observations	24,705	24,705	24,705	24,705
Panel B: PG_{T-1} vs. PG_{T+1}				
Variables	Posts in default PG	Posts in total PG	Responses in default PGs	Responses in total PGs
PG_{T-1}	−0.112**	0.0748	−0.770***	−0.445***
	(0.0507)	(0.0511)	(0.124)	(0.122)
PG_{T+1}	−0.255***	−0.0794	−1.375***	−1.014***
	(0.0631)	(0.0636)	(0.158)	(0.156)
Constant	1.033***	1.014***	1.113***	1.075***
	(0.0107)	(0.0107)	(0.0261)	(0.0256)
Peer group fixed effects	Yes	Yes	Yes	Yes
F-tests of instrument	254.52	254.52	254.52	254.52
Observations	24,705	24,705	24,705	24,705

Notes: In Panel A, we use |EDD-16| as the instrument for two PGs; In Panel B, we use *EDD* and EDD^2 as the instruments for PG_{T-1} and PG_{T+1}, respectively.
Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 8
Information exchange and multiple peer groups (#PGs $\geq$ 2).

	(1) IV 1st stage	(2) OLS/ IV 2nd stages	(3)	(4)	(5)
Variables	Multiple PGs	log(info) in Default PG OLS	log(info) in Default PG IV	log(info) in Total PGs OLS	log(info) in Total PGs IV
$\widehat{\text{Multiple}}$ PGs ($\geq$ 2)		0.228***	−0.916***	0.309***	−0.384***
		(0.00910)	(0.135)	(0.00918)	(0.124)
|Day of EDD-16|	0.00896***				
	(0.000570)				
Constant	0.408***	1.522***	2.064***	1.519***	1.847***
	(0.00608)	(0.00887)	(0.0649)	(0.00890)	(0.0593)
Peer group fixed effects	Yes	Yes	Yes	Yes	Yes
F-tests of instrument	–	–	246.73	–	246.73
Observations	38,919	38,919	38,919	38,919	38,919

Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

the computation can be found in Section B.1. Among all the posts in our sample, 42.8% contain at least one of the question indicators, which suggests that users do not necessarily ask questions in their posts.

Despite the fact above, it would be interesting to run the same analysis using the number of the posts and the number of responses as two different outcome variables. The results are shown in Table 7. The coefficients of two PGs in Panel A are consistent with the ones of pooling the posts and responses in Table 4. If we nevertheless assume that posts equal to asking questions and responses equal to answering questions, we would expect that users who additionally join PG_{T-1} would respond more and ask less in their default group *PG* than those who additionally join PG_{T+1}. The results in Panel B in Table 7, however, do not confirm this prediction which again casts doubt on the assumption that posts and responses have distinct patterns of asking and answering questions.

5.5. The case of multiple peer groups

Finally, we generalise the analysis from two peer groups to multiple peer groups. We relax the restriction on the number as well as the "distance" of additional peer groups from the default peer group. We include users that either have the default peer group only (control) or additionally join any peer group(s) (treated), which yields 38,919 observations.

Column (1) in Table 8 reports the first stage regressions in a linear probability model. The variable |Day of EDD-16| is still a good predictor for $\widehat{\text{Multiple}}$ PGs ($p < 0.01$). Columns (2)–(5) in Table 8 show the estimated effects of having multiple peer groups on users' information exchange in their default peer group and in total peer groups, respectively. Similar to the case of exactly two peer groups, having multiple peer groups is associated with 26% more information in the default peer group and 36% more information in total peer groups in the OLS specification (Columns (2) and (4)). However, once we instrument the multiple peer groups, the coefficients become significantly negative ($p < 0.01$). In the IV specification (Columns (3) and (5)), having multiple peer groups leads to about 60% less information in the default peer group and 32% less information in total peer groups. The *F*-test confirms that the instrument is still very strong in the case of multiple peer groups.

In both cases —having one additional adjacent peer group and having any additional peer group(s)— we find a reduction in the information exchange in user's default peer group. This is in line with our prediction for the substitution effect between the additional peer group(s) and the default peer group. However, when we aggregate all the information exchange of the two/multiple-peer-group users in total peer groups they joined, it is still lower than from the single-default-peer-group users. If we assume a linear production function of information exchange, the overall reduction suggests that there could be further "detrimental" effects besides the substitution effect. One possible explanation is that having additional peer groups weakens the group identity of any group, which can trigger more "lurking" behaviour and reduce contribution in each peer group.[28]

6. Conclusion

In this paper we combine a setup of quasi-randomly assigned groups with an instrumental variable approach to investigate the causal effects of having multiple peer groups (vs a single default peer group) on the information exchange in an online community. Our results show that individuals who have multiple peer groups generate less information exchange not only in their default peer group but also in all peer groups they joined than those who have only one default peer group. Furthermore, the reduction in the information exchange in the default peer group is smaller when the additional peer group is more advanced than the default peer group.

We interpret the overall reduction in the default peer group as substituting information between peer groups. The heterogeneous effects from the additional peer group — previous-month vs next month — suggest that the users joining the earlier group benefit from receiving new information due to the asymmetric information flow from the more advanced to the less advanced peer groups. Moreover, given that the total information exchange of the two/multiple-peer-group users' is less than of the single-default-peer-group users, there can be other effects besides the substitution and spillover effects. For instance, having multiple peer groups may weaken the group identity of the default peer group as well as the additional peer groups. This could alter individual preference and behaviour in their groups (Charness and Chen, 2020), a demonstration of which could be increased lurking behaviour.

Our results suggest that organisations and communities with similar features be cautious when designing the group-based structure. Allowing individuals to join multiple peer groups may dampen their total engagement and lower the amount of information exchange, which is in opposition to the original intention of the community. This can also go beyond our context, especially in the domain of social media. Today, our online social media status is defined by numbers: number of "friends", "followers", "likes", and "subscribers". Users are often encouraged to create multiple social media accounts, make more friends and links, attract more followers etc. Moreover, new online platforms are emerging and expanding their user pools, users are increasingly facing the decision of joining additional platforms. The face value of the costs of doing so is extremely low if not for free. However, as the numbers of groups and links rocket, to process all the information is clearly beyond our capacities, as Simon (1957) pointed out:"A world in which the scarce factor is information may be exactly the wrong one for a world in which the scarce factor is attention."

Finally, this paper sets out a research agenda to further evaluate group-based structure and in particular, examine multiple group membership as a determinant of information exchange, social interaction, as well as other productivity and performance measures in online settings. Not only traditional social events and activities but also workplace have been increasingly transformed into digital versions. There are certainly differences between online and offline networks for multiple group membership to operate. Whereas both virtual and real communities could function as platforms for information exchange and social interaction, they differ in important ways. For example, online social communities enable people to overcome physical and time constraints embedded within their offline counterparts. Moreover, the option of maintaining anonymity in online settings provides users more convenience when discussing sensitive and controversial topics. The online peer groups studied here is just one of the examples and further qualitative text analysis could complement the results of our quantitative analysis. Moreover, while the current study stands more in the shoe of the online community that mainly seeks high volume of information flow, future studies could also take the perspective from the user side and investigate how multiple group membership may affect individual well-being.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

[28] Lurking is very common in online communities. It is related to the free riding behaviour in the sense that lurking users have access to all information without generating their own information. Research has shown that stronger feelings about group identity increase the level of contribution in the public goods games (Charness et al., 2014).

Appendix A. Appendix

A1. Additional descriptive statistics

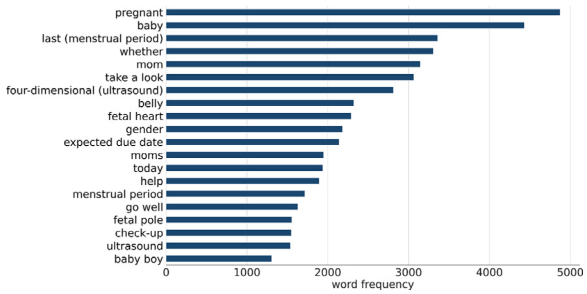

Fig. A1. Word frequency in the information exchange.
Notes: This figure shows the most frequently mentioned words in the three core peer groups (March, April, and May 2018). It is based on 68,079 anonymous posts with at least one responses.

Table A1
Descriptive statistics of each peer group.

Variables	March 2018	April 2018	May 2018
Day of estimated due date (EDD) (1st–30th/31st)	15.18	15.00	15.09
	8.721	8.414	8.729
Number of posts per user in default peer group	2.045	2.097	2.106
	(1.953)	(1.980)	(2.055)
Number of responses per user in default peer group	3.598	3.779	3.677
	(7.017)	(7.217)	(7.371)
Fraction of two PGs	15.30%	18.71%	17.03%
Observations	9638	7940	7127

Notes: This table shows the descriptive statistics in each of the three core peer groups (March, April, and May 2018) separately.

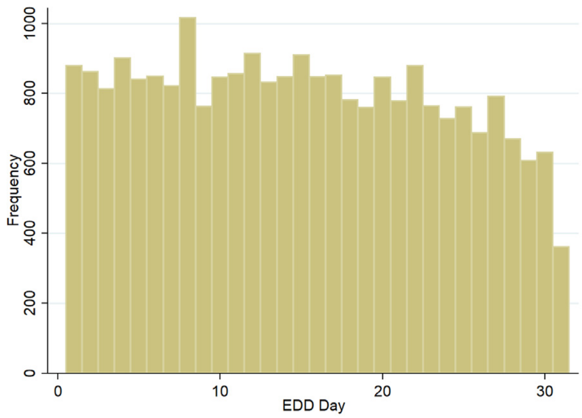

Fig. A2. Number of users by the day of EDD.
Notes: The figure shows the number of users on each day of the month in the three core peer groups (March, April, and May 2018). The numbers are approximately equally spread over the month despite a slight downward trend towards the end of the month and that one of our peer groups (the month of April) does not have 31 days. There is also a spike on the 8th which is perceived as a lucky number in China and might be related to planned cesarean cut. We perform the same empirical analysis excluding users whose EDD is on the 8th of the month, the results remain robust.

Table A2
Counting and characterising compliers.

(1) $P[D=1]$ 0.169	(2) $P[D_1 > D_0]$ 0.096	(3) $P[z=1]$ 0.454	(4) $P[D_1 > D_0 \mid D=1]$ 0.256	(5) $P[D_1 > D_0 \mid D=0]$ 0.063
	Always-takers when treated	Never-takers when not treated	Compliers when treated	Compliers when untreated
		Information exchange in Default PG		
Variables				
Posts	2.421	1.967	2.292	2.069
	(2.266)	(1.865)	(2.153)	(1.997)
Responses	5.709	3.059	4.425	3.689
	(8.984)	(6.515)	(7.548)	(7.220)
Total info	8.131	5.026	6.717	5.758
	(10.215)	(7.457)	(8.687)	(8.196)
		Information exchange in Total PG		
Variables				
Posts	2.575	1.967	2.664	2.069
	(2.361)	(1.865)	(2.431)	(1.997)
Responses	6.410	3.059	5.674	3.689
	(9.522)	(6.515)	(8.798)	(7.220)
Total info	8.985	5.026	8.338	5.758
	(10.750)	(7.457)	(10.058)	(8.196)

Notes: Average outcomes of always-takers when treated, never-takers when not treated, compliers when treated, and compliers when untreated. Notice that in Columns (3) and (5), since these users only have one PG, the information exchange in Total PG is equal to the one in the default PG.

A2. Reduced forms and robustness checks

Table A3
Reduced forms: two peer groups and information exchange.

Variables	(1) log(info) in Default PG	(2) log(info) in Total PGs	(3) log(info) in Default PG	(4) log(info) Total PGs
|Day of EDD-16|	-0.00894^{***}	-0.00510^{***}	-0.00837^{***}	-0.00451^{***}
	(0.00118)	(0.00121)	(0.00109)	(0.00111)
Number of non-PGs			-0.00218	-0.00179
			(0.00151)	(0.00153)
log(info) in non-PGs			0.343^{***}	0.352^{***}
			(0.00570)	(0.00575)
Constant	1.583^{***}	1.581^{***}	1.327^{***}	1.314^{***}
	(0.0127)	(0.0128)	(0.0189)	(0.0191)
Peer group fixed effects	Yes	Yes	Yes	Yes
Observations	24,705	24,705	24,705	24,705
R-squared	0.003	0.001	0.143	0.146

Notes: The reduced forms in Columns (1) and (2) are for the specifications in Table 4; The reduced forms in Columns (3) and (4) are for the specifications in Table A4.
Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A4
Two peer groups and information exchange with NPG controls.

	(1) IV 1st stage	(2) OLS/ IV 2nd stages	(3)	(4)	(5)
Variables	Two PGs	log(info) in Default PG OLS	log(info) in Default PG IV	log(info) in Total PGs OLS	log(info) in Total PGs IV
$\widehat{\text{Two PGs}}$		0.175***	−0.671***	0.333***	−0.361***
		(0.0138)	(0.0949)	(0.0137)	(0.0931)
| Day of EDD-16|	0.0125***				
	(0.000551)				
Number of non-PGs	0.00522***	−0.00312**	0.00131	−0.00354**	9.88e−05
	(0.000716)	(0.00151)	(0.00170)	(0.00151)	(0.00166)
log(info) in non-PGs	0.0270***	0.339***	0.361***	0.344***	0.362***
	(0.0272)	(0.00571)	(0.00662)	(0.00570)	(0.00646)
Constant	0.0233***	1.246***	1.343***	1.243***	1.322***
	(0.00871)	(0.0170)	(0.0214)	(0.0170)	(0.0209)
Peer group fixed effects	Yes	Yes	Yes	Yes	Yes
F-tests of instrument	–	–	512.85	–	512.85
Observations	24,705	24,705	24,705	24,705	24,705

Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A5
Two peer groups and information exchange in a Poisson IV model.

Variables	(1) Information in Default PG	(2) Information in Total PGs	(3) Information in Default PG	(4) Information in Total PGs
Two PGs	−3.187	−0.862***	−2.513*	−0.805***
	(2.954)	(0.328)	(1.432)	(0.293)
Number of non-PGs			0.00917***	0.00722**
			(0.00312)	(0.00300)
Information in non-PGs			0.0404***	0.0403***
			(0.00144)	(0.00135)
Constant	1.969***	1.920***	1.735***	1.707***
	(0.0322)	(0.0329)	(0.0383)	(0.0377)
Peer group fixed effects	Yes	Yes	Yes	Yes
Observations	24,705	24,705	24,705	24,705

Robust standard errors in parentheses *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A6
First stages of the heterogeneity analysis.

	(1) First set of IV (f_1, f_2)	(2)	(3) Second set of IV (g_1, g_2)	(4)
Variables	Additional PG PG_{T-1}	Additional PG PG_{T+1}	Additional PG PG_{T-1}	Additional PG PG_{T+1}
f_1	−0.0206***	−0.00665***		
	(0.00103)	(0.000659)		
f_2	0.000411***	0.000413***		
	(2.90e−05)	(2.49e−05)		
g_1			0.0143***	0.000131
			(0.000610)	(0.000282)
g_2			−0.00200***	0.0119***
			(0.000378)	(0.000495)
Constant	0.321***	0.0323***	0.0884***	0.0111***
	(0.00881)	(0.00378)	(0.00466)	(0.00327)
Peer group FE	Yes	Yes	Yes	Yes
Observations	24,705	24,705	24,705	24,705

Notes: The first set of instruments is $f_1(EDD) = EDD$ and $f_2(EDD) = EDD^2$ for PG_{T-1} and PG_{T+1}, respectively. The second set of instruments is $g_1(EDD) = |EDD - 16|$, if $EDD \leq 16$; $g_1(EDD) = 0$, if $EDD > 16$; and $g_2(EDD) = 0$, if $EDD \leq 16$; $g_2(EDD) = |EDD - 16|$, if $EDD > 16$.

Appendix B. Miscellaneous

B1. Counting posts containing question indicators

In our baseline empirical analysis, we consider both posts and responses as information exchange and pool them together as the outcome variable. However, our first intuition is that a post mostly likely raises questions and a response delivers answers. To verify whether this is the case, we compute the share of posts that contain any question indicators including interrogative pronouns, modal particles, help and request, and question marks. The list of question indicators is ‘谁’,‘何’,‘么’,‘哪’,‘几’,‘多少’,‘怎’,‘吗’,‘呢’,‘吧’,‘啊’, ‘嘛’,‘呀’,‘咋’,‘啥’,‘有没有’‘’有木有’,‘帮’,‘请’,‘看看’,‘求助’,‘求解’, ‘是不’,‘能不’,‘好不’,‘否’,‘有宝妈’,‘多久’,‘急’, and ‘?’.

Among all the posts in our sample, 42.8% contain at least one of the question indicators, which confirms that users do not necessarily ask questions in their posts and justifies pooling both posts and responses as information exchange in the main empirical analysis.

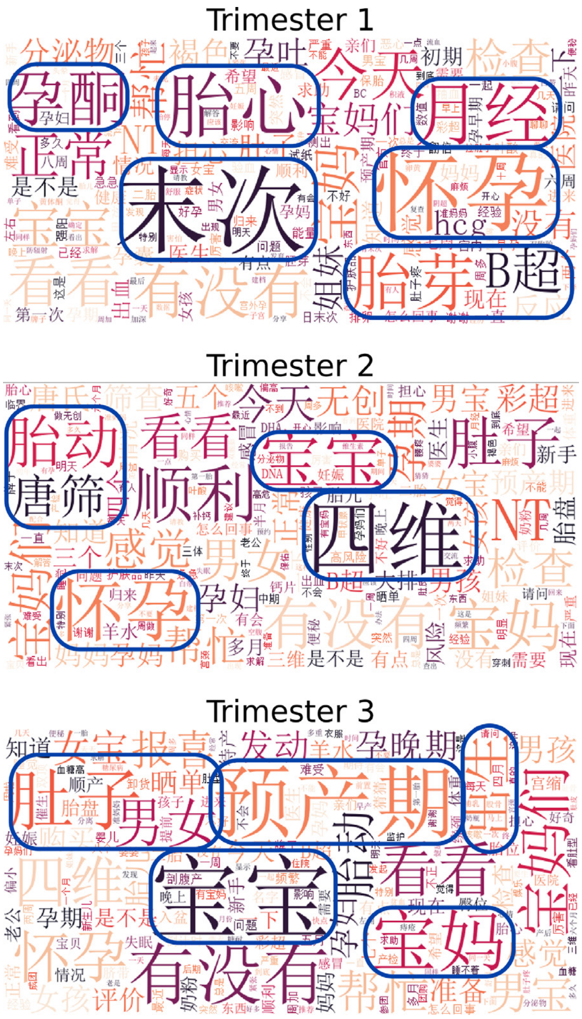

Fig. B1. World cloud by trimester.

B2. Topic analysis by trimester

To explore whether there are any variations in the pattern of information exchange along the course of pregnancy, we split the entire pregnancy by three periods, namely, the three trimesters. The first/second/third trimester is 0–13 weeks/14–26 weeks/27–40 weeks. Each of the trimester is marked by specific fetal developments and experiences different symptoms. We generate the word clouds by trimester in the original language in Fig. B.1. To give an idea how the main topics change over time, we circled the most frequent nouns in each trimester. In Trimester 1, the circled words are: pregnancy, last menstrual period (LMP), progesterone, fetal heart, menstruation, fetus, B-mode ultrasound; In Trimester 2, the circled words are: pregnancy, baby, four dimensional ultrasound, fetal movement, Down's screening; In Trimester 3, the circled words are: baby, estimated due date, belly, gender, mother(s), birth.

Fig. B2. A screenshot of a post with sharing content.
Notes: A screenshot of a post sharing her experience at week 7. This user was in the peer group April 2018, she shared her B-mode ultrasound results and what the doctor told her during the visit. We added the translation next by to the original language. Notice that the name of the user on the document has been removed by the user herself.

B3. Examples of posts and user interface

We show examples of two different types of post and their responses in Figs. B.2 and B.3 below. The first type of post is sharing experience without asking specific questions; and the responses of such posts may consist of follow up questions or simply show appreciation. The second type of post is raising a question (here about frequent urination); and the responses of such posts can consist of answers or follow up questions. There can be (and often is the case) multiple responses to the same post, some containing more substantial information than others, which of course depends on the types of post.

Fig. B3. A screenshot of a post with a clear question.
Notes: A screenshot of a post with a clear question in week 12. This user was in the peer group April 2018, she asked a question about frequent and urgent urination. We added the translation next by to the original language.

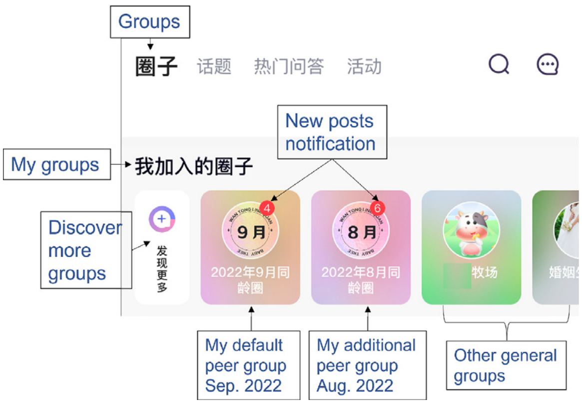

Fig. B4. A screenshot of a user's interface with two peer groups.
Notes: This user has joined two peer groups. In her interface, she sees the two peer groups with notifications of new posts, and other general groups that she has joined.

References

Angrist, J.D., 1990. Lifetime earnings and the vietnam era draft lottery: evidence from social security administrative records. Am. Econ. Rev. 313–336.
Angrist, J.D. and Pischke, J.S., 2009. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
Angrist, J.D., Chen, S.H., 2011. Schooling and the vietnam-era GI bill: evidence from the draft lottery. Am. Econ. J. 3 (2), 96–118.
Åslund, O., Fredriksson, P., 2009. Peer effects in welfare dependence quasi-experimental evidence. J. Hum. Resour. 44 (3), 798–825.
Bertolotti, F., Mattarelli, E., Vignoli, M., Macrì, D.M., 2015. Exploring the relationship between multiple team membership and team performance: the role of social networks and collaborative technology. Res. Policy 44 (4), 911–924.
Borjas, G.J., Doran, K.B., 2015. Which peers matter? The relative impacts of collaborators, colleagues, and competitors. Rev. Econ. Stat. 97 (5), 1104–1117.
van de Brake, H.J., Walter, F., Rink, F.A., Essens, P.J., van der Vegt, G.S., 2020. Multiple team membership and job performance: the role of employees' information-sharing networks. J. Occup. Organ. Psychol. 93 (4), 967–987.
Brown, K.M., Laschever, R.A., 2012. When they're sixty-four: peer effects and the timing of retirement. Am. Econ. J. 4 (3), 90–115.
Bruhin, A., Goette, L., Haenni, S., Jiang, L., 2020. Spillovers of prosocial motivation: evidence from an intervention study on blood donors. J. Health Econ. 70, 102244.
Carrell, S.E., Fullerton, R.L., West, J.E., 2009. Does your cohort matter? Measuring peer effects in college achievement. J. Labor Econ. 27 (3), 439–464.
Charness, G., Chen, Y., 2020. Social identity, group behavior, and teams. Annu. Rev. Econ. 12, 691–713.
Charness, G., Cobo-Reyes, R., Jiménez, N., 2014. Identities, selection, and contributions in a public-goods game. Games Econ. Behav. 87, 322–338.
Charness, G., Rigotti, L., Rustichini, A., 2007. Individual behavior and group membership. Am. Econ. Rev. 97 (4), 1340–1352.
Chen, Y., Li, S.X., 2009. Group identity and social preferences. Am. Econ. Rev. 99 (1), 431–457.
Cohen, S., Syme, S.L., 1985. Issues in the study and application of social support. Social Support Health 3, 3–22.
Cornelissen, T., Dustmann, C., Schönberg, U., 2017. Peer effects in the workplace. Am. Econ. Rev. 107 (2), 425–456.
Crawford, E.R., Reeves, C.J., Stewart, G.L., Astrove, S.L., 2019. To link or not to link? Multiple team membership and unit performance. J. Appl. Psychol. 104 (3), 341.
Cummings, J.N., Haas, M.R., 2012. So many teams, so little time: time allocation matters in geographically dispersed teams. J. Organ. Behav. 33 (3), 316–341.
Duflo, E., Saez, E., 2002. Participation and investment decisions in a retirement plan: the influence of colleagues' choices. J. Public Econ. 85 (1), 121–148.
Elsner, B., Isphording, I.E., Zölitz, U., 2021. Achievement rank affects performance and major choices in college. The Economic Journal 131 (640), 3182–3206.
Evans, W.N., Oates, W.E., Schwab, R.M., 1992. Measuring peer group effects: a study of teenage behavior. J. Polit. Econ. 100 (5), 966–991.
Falkinger, J., 2008. Limited attention as a scarce resource in information-rich economies. Econ. J. 118 (532), 1596–1620.
Ferrario, B., Stantcheva, S., 2022. Eliciting people's first-order concerns: text analysis of open-ended survey questions. In: AEA Papers and Proceedings, vol. 112, pp. 163–169.
Gächter, S., Nosenzo, D., Sefton, M., 2013. Peer effects in pro-social behavior: social norms or social preferences? J. Eur. Econ. Assoc. 11 (3), 548–573.
Ghiglino, C., Juárez-Luna, D., Müller, A., 2021. Class altruism and redistribution. Econ. J. 131, 3274–3295.
Goette, L., Huffman, D., Meier, S., 2006. The impact of group membership on cooperation and norm enforcement: evidence using random assignment to real social groups. Am. Econ. Rev. 96 (2), 212–216.
Granovetter, M., 2018. Economic action and social structure: the problem of embeddedness. In: The Sociology of Economic Life. Routledge, pp. 22–45.
Heckman, J., 1990. Varieties of selection bias. Am. Econ. Rev. 80 (2), 313.
Hoffman, C.S., Messer, L.C., Mendola, P., Savitz, D.A., Herring, A.H., Hartmann, K.E., 2008. Comparison of gestational age at birth based on last menstrual period and ultrasound during the first trimester. Paediatr. Perinat. Epidemiol. 22 (6), 587–596.
Imbens, G.W., Angrist, J.D., 1994. Identification and estimation of local average treatment effects. Econometrica 62 (2), 467–475.
Jackson, M.O., 2010. Social and Economic Networks. Princeton University Press.
Jiang, L., 2020. Splash with a teammate: peer effects in high-stakes tournaments. J. Econ. Behav. Organ. 171, 165–188.
Jiang, L., Zhu, Z., forthcoming. *Maternal mental health and social support from online communities during pregnancy*. Health & Social Care in the Community.
Kahneman, D., 1973. Attention and Effort, vol. 1063. Citeseer.
Kimbrough, E.O., McGee, A.D., Shigeoka, H., 2020. How do peers impact learning? An experimental investigation of peer-to-peer teaching and ability tracking. J. Hum. Resour. 0918–9770R2.
Manski, C.F., 1993. Identification of endogenous social effects: the reflection problem. Rev. Econ. Stud. 60 (3), 531–542.
Margolis, J., 2020. Multiple team membership: an integrative review. Small Group Res. 51 (1), 48–86.
Mas, A., Moretti, E., 2009. Peers at work. Am. Econ. Rev. 99 (1), 112–145.
Milgrom, P. R., Roberts, J. D., 1992. Economics, organization and management.
Mortensen, M., Haas, M.R., 2018. Perspective–rethinking teams: from bounded membership to dynamic participation. Organ. Sci. 29 (2), 341–355.
O'leary, M.B., Mortensen, M., Woolley, A.W., 2011. Multiple team membership: a theoretical model of its effects on productivity and learning for individuals and teams. Acad. Manag. Rev. 36 (3), 461–478.
Pluut, H., Flestea, A.M., Curşeu, P.L., 2014. Multiple team membership: a demand or resource for employees? Group Dyn. 18 (4), 333.
Sacerdote, B., 2001. Peer effects with random assignment: results for Dartmouth roommates. Q. J. Econ. 116 (2), 681–704.
Simon, H. A., 1957. Models of man; social and rational.
Słoczyński, T., 2020. Interpreting OLS estimands when treatment effects are heterogeneous: smaller groups get larger weights. Rev. Econ. Stat. 1–27.
Turkiewicz, K.L., 2018. Data trimming. In: Allen, M. (Ed.), The SAGE Encyclopedia of Communication Research Methods. SAGE Publications, Inc, Thousand Oaks, pp. 347–348.
Yamane, S., Hayashi, R., 2015. Peer effects among swimmers. Scand. J. Econ. 117 (4), 1230–1255.
Zhu, H., Kraut, R.E., Kittur, A., 2014. The impact of membership overlap on the survival of online communities. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 281–290.