Who and what gets recognized in peer recognition

Meagan Sundstrom,[1,†] L. N. Simpfendoerfer,[1] Annie Tan,[1]
Ashley B. Heim,[2] and N. G. Holmes[1,*]

[1]*Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, New York 14853, USA*
[2]*Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853, USA*

Previous work has identified that recognition from others is an important predictor of students' participation, persistence, and career intentions in physics. However, research has also found a gender bias in peer recognition in which student nominations of strong peers in their physics course disproportionately favor men over women. In this study, we draw on methods from social network analysis and find a consistent gender bias in which men disproportionately undernominate women as strong in their physics course in two offerings of both a lecture course (for science and engineering, but not physics, majors) and a distinct lab course (for science, engineering, and physics majors). We also find in one offering of the lecture course that women disproportionately undernominate men, contrary to what previous research would predict. We expand on prior work by also probing two data sources related to who and what gets recognized in peer recognition: students' interactions with their peers (who gets recognized) and students' written explanations of their nominations of strong peers (what gets recognized). Results suggest that the nature of the observed gender bias in peer recognition varies between the instructional contexts of lecture and lab. In the lecture course, the gender bias is related to who gets recognized: both men and women disproportionately overnominate their interaction ties to students of their same gender as strong in the course. In the lab course, the gender bias is also related to what gets recognized: men nominate men more than women because of skills related to interactions, such as being helpful. These findings illuminate the different ways in which students form perceptions of their peers and add nuance to our understanding of the nature of gender bias in peer recognition.

I. INTRODUCTION

Gaining recognition as a physicist is important for students' participation and persistence in their physics course [1–8]. Recognition is particularly important for the participation of historically underrepresented groups in physics, such as women [3,7,9,10]. However, research has found that men both perceive higher recognition from others [2,7,11,12] and receive more recognition from their physics peers [13,14] than women. To better understand these effects, we investigated the nature of student recognition of strong peers with a focus on the gender bias in such peer recognition. Specifically, we probe two questions: whether and how gender bias in students' nominations is related to patterns of peer interactions (*who* gets recognized), and whether and how gender bias in students' nominations is related to their written explanations of these nominations (*what* gets recognized). Throughout the paper, we use the phrase *gender bias* to refer to a distinguishable difference between the amount of peer recognition received by men (women) and the amount of peer recognition we would expect men (women) to receive if recognition were distributed equitably across men and women.

A. Recognition in physics courses

A student's sense of physics identity—the degree to which they believe they are a "physics person" [15]—has been shown to predict their participation, persistence, and career intentions in physics [2,4,7]. Researchers have modeled physics identity as containing three dimensions: performance and competence, interest, and recognition [3,4]. Previous studies demonstrate that *recognition* is the most important of these three dimensions in relating to and predicting student outcomes [1–8]. Recognition is the extent to which meaningful others (e.g., peers, teachers, and family) perceive an individual as a physics person. When a student receives more recognition from others, they are more likely to see themselves as a physics person and, therefore, develop a stronger physics identity [5,16].

*Corresponding author: ngholmes@cornell.edu
†mas899@cornell.edu

Recognition from others, however, is often shaped by sociohistorical norms and stereotypes, such as those that position men as more suitable to the field of physics than women [3,7,9,17–26]. Perhaps as a result of such stereotypes, a handful of research studies demonstrate that men report higher perceived recognition (the extent to which they feel recognized by others) in their physics classes than women [2,7,11,12]. This difference may put men in a better position than women to develop their physics identity, contributing to the underrepresentation of women in physics documented in, for example, Refs. [27–29].

B. Gender bias in peer recognition

Other studies have probed "actual" peer recognition, rather than perceived recognition, by asking students to nominate peers they believe are strong in their science course [13,14,30,31]. We use the term actual recognition to mean a measure of how much others recognize an individual as a physics person, though the others may not have indicated that recognition to the individual, necessarily. These studies largely draw on quantitative methods of social network analysis to determine the extent to which a gender bias exists in students' nominations of strong peers [14,30,31], finding mixed results. Grunspan and colleagues [30], for example, examined three offerings of an introductory biology course (the second in the course sequence) for first-year students. They observed that men disproportionately undernominated women, while women proportionately nominated men and women, as strong in the course material in all three offerings. Bloodhart and colleagues [13] observed a similar gender bias in peer recognition across many introductory physics courses for first-year students, but found that women also disproportionately undernominated women as strong in the course material. In the same study, researchers found that men proportionately nominated both men and women, but women disproportionately overnominated other women, as strong in introductory life sciences courses for first-year students. Salehi and colleagues [31], however, found no gender bias in either men's or women's nominations of strong peers across two offerings of a mechanical engineering course taken by second and third-year students.

Our previous work [14] examined peer recognition in three different remote physics courses and added nuance to these studies. We observed a gender bias in peer recognition favoring men (in which men disproportionately undernominated women, but women proportionately nominated men and women) in two introductory physics courses aimed at first-year students, but no gender bias favoring men in an introductory physics course comprised mostly of second-year students (though women disproportionately overnominated other women in this course). Comparing across all four studies [13,14,30,31], the presence or absence of a gender bias in peer recognition seems to vary by course level more than any other aspect

of the instructional context: researchers find a gender bias in peer recognition in science courses for first-year, but not beyond-first-year, students. Gender bias in peer recognition also seems related to student outspokenness (i.e., verbally participating in class). In the two of these four studies that measure outspokenness, a gender bias in students' nominations of strong peers (favoring men) is present when there is also a gender disparity in who is outspoken (i.e., when men participate more than women), and there is no gender bias in peer recognition when there is no gender disparity in who is outspoken [14,30].

In physics, patterns of peer recognition also seem to vary between lecture and lab contexts, with our previous study of physics courses [14] finding more evidence of gender bias in peer recognition in the context of lecture material than lab material. This result may be attributable to these two instructional contexts covering distinct content and aiming to develop different sets of skills [32–37]. Indeed, research has shown that students believe lecture skills include knowing mathematics, while lab skills involve handling equipment and using technical skills [18,19]. However, the difference in peer recognition across instructional contexts may also be attributable to pedagogy: lectures typically contain many students who focus on the instructor and labs typically contain a small number of students who collaborate on tasks. This variation in how much visibility students have in front of their peers may impact patterns of peer recognition, especially during remote instruction as in our prior work [14]. In the current study, therefore, we determine the extent to which gender bias in peer recognition exists when lecture and lab material are taught with a similar pedagogical style and in person. Specifically, we analyze physics courses for first-year students where the lab and lecture material comprise distinct courses (i.e., students co-enroll in one lab course and one lecture course and receive a separate grade in each course), and both of these courses contain a lecture session where all students focus on the instructor and a small-group session where students solve problems or conduct experiments with their peers.

C. What is the nature of this gender bias?

While the quantitative studies mentioned above importantly determine the extent to which a gender bias in peer recognition exists, they do not probe the nature of this gender bias. Toward this end, separate threads of research have started to unpack two possible mechanisms underlying recognition: peer interactions (i.e., students learn about their peers' skills during interactions with these peers) and students' reasons for recognizing others as strong in their physics course (i.e., students recognize their strong physics peers for different skills). Of course, there may also be other explanations for the gender bias in peer recognition, such as sociohistorical gender stereotypes

alone, but, in this study, we seek to build on the existing research threads.

1. Who gets recognized: Peer interactions and indirect observations of peers

Previous work suggests that one mechanism through which recognition forms is interactions with others [15,38,39]. In their original conception of the identity framework, Gee states that "the modern need for recognition places a particular importance on discourse and dialogue...Individuals must win recognition for them through exchange with others" [15] (p. 112). We interpret this to mean that an individual demonstrates their knowledge, skills, and personality traits through conversations with others, who then form perceptions of that individual as a certain kind of person. One study, for example, conducted interviews with undergraduate students to understand their experiences in a remote summer research program [39]. The authors found that students' research group members and advisors started to recognize the students as physicists and researchers during interactions with one another: "Other recognition was supported by conversations between the mentee and other group members" [39] (p. 10).

Similar work describes peer interactions as a way for students to determine who of their peers is strong in physics. In one study, researchers performed a longitudinal case study of a woman in physics named Cassidy [38]. At the beginning of her undergraduate physics studies, Cassidy recognized her more senior peer tutor, one of the only other physics students with whom she interacted, as a "smart" physics student because they showed her an unnecessarily complicated solution to a physics problem. About a year later, however, Cassidy became a more outgoing member of the physics community who frequently collaborated with peers on assignments. These peer interactions facilitated Cassidy's understanding of the "multiplicity of ways to be 'good' at physics," [38] (p. 12) such as bringing in different areas of expertise to a peer collaboration. She then recognized many of her peers as being strong physics students, rather than only her peer tutor. This and other studies [15,39], therefore, suggest that interactions are likely a mechanism for forming peer recognition: interactions facilitate students' understanding of their peers' knowledge and skill sets which informs who gets recognized. This mechanism of forming peer recognition may also relate to gender bias in peer recognition because previous work has found that students tend to interact with peers of their same gender [40,41].

Peer interactions, however, are not the only means through which students determine who gets recognized. Grunspan and colleagues [30], for example, demonstrate that outspokenness—frequent verbal participation in front of many others—also relates to which students receive peer recognition. They found that students who actively participated in lecture tended to receive more nominations from peers as strong in their biology course despite these students never directly interacting with one another. In addition to direct interactions with peers, therefore, students may determine who they recognize as a strong peer by indirectly observing their peers.

In the current study, we examine the relationship between peer interactions, indirect observations of peers, and peer recognition by quantitatively comparing students' self-reported peer interactions to their nominations of strong peers. We also compare this relationship across men's versus women's nominations to determine whether patterns of interactions relate to the nature of gender bias in peer recognition (we could not measure whether patterns of indirect observations relate to the nature of gender bias in peer recognition, see Sec. II C 3).

2. What gets recognized: Skill sets associated with being a physicist

Other studies have explicitly probed the knowledge, skill sets, and traits for which students recognize strong peers in their physics or other science courses [18,19,42–51]. Doucette and colleagues, for example, asked undergraduate physics students to describe their ideal lab partner [49]. The authors identified 13 characteristics from the responses that students recognized in a "good lab partner," including knowledgeable, hardworking, communicative, helpful, and efficient. In another study, Irving and Sayre interviewed upper-level physics students and asked them what they think it means to be a physicist [47]. The participants noted a wide array of skills or traits that they recognized in a physicist, including intuition for learning physics, interest in physics, solving physics problems, designing experiments, and collecting and interpreting experimental data. Some studies also relate the identified skills to gender. Danielsson [18], for instance, found that undergraduate students associate natural ability, tinkering with lab equipment, and mathematical competence with men and diligence and note taking with women.

In the current study, we expand on this body of work by (i) collecting and analyzing a large number of student explanations of their nominations of strong peers and (ii) comparing the frequencies of explanations written by men and women when nominating men versus women to determine whether and how the explanations relate to gender bias in the nominations.

D. Current study

In summary, research has demonstrated that whether a gender bias exists in peer recognition varies across courses and instructional contexts. Prior work also suggests that peer interactions and differences in what skills sets are associated with being strong in physics might help to explain the nature of this gender bias. To probe these two possible mechanisms underlying who and what gets recognized in peer recognition, we conducted a mixed-methods

study of in-person physics courses to answer the following research questions:

1. To what extent does a gender bias exist in students' recognition of strong physics peers within distinct lab and lecture courses?
2. **Who gets recognized:** In distinct lab and lecture physics courses, how (if at all) is gender bias in peer recognition related to patterns of peer interactions?
3. **What gets recognized:** In distinct lab and lecture physics courses, how (if at all) is gender bias in peer recognition related to the skill sets students associate with being strong in physics?

We collected students' nominations of strong peers, explanations for these nominations, and self-reported interactions with peers in two offerings of distinct introductory lab and lecture physics courses for first-year science and engineering students at Cornell University. Similar to prior research examining introductory biology and physics courses for first-year students [13,14,30], we find a gender bias in peer recognition in which men disproportionately undernominate women compared to men in all analyzed courses. We also find that women disproportionately undernominate men in one offering of the lecture course. Comparing the nominations of strong peers to peer interactions, we observe in most cases that the overall gender bias in peer recognition is related to gender bias in interaction-based recognition, where students disproportionately overnominate their interaction ties to peers of their same gender. Finally, we find a difference in students' written explanations in the lab course, where men nominate men more than women because of the ways they interacted, such as being helpful, but not in the lecture course, where men and women nominate men and women for similar skill sets.

II. METHODS

In this section, we describe the instructional context of our study and then discuss our data collection and analysis methods.

A. Instructional context

The data come from two in-person offerings (fall and spring) of two distinct introductory physics courses (summarized in Table I), one *lab course* and one *lecture course*, at Cornell University—a large, private, Ph.D.-granting institution in the northeastern United States with a Carnegie classification of very high research activity.

The *lab course* focused on developing experimental skills rather than reinforcing physics concepts (see,

TABLE I. Summary of survey response rates and self-reported student demographics for the four courses we analyzed. All analyzed students in the lecture course are also in the lab course of the corresponding semester. Percentages are relative to the number of students included in the analysis unless specified otherwise. We grouped race or ethnicity by underrepresented racial minority (URM) status, where non-URM students are those solely identifying as White and/or Asian or Asian American and URM students are those identifying as at least one of any other race or ethnicity (including American Indian or Alaska Native, Black or African American, Hispanic or Latinx, and Native Hawaiian or other Pacific Islander). We denote students' demographic information as "unknown" if they preferred not to disclose this information on the survey or if they did not complete the survey.

	Lab course		Lecture course	
	Fall	Spring	Fall	Spring
Survey response rate (% of total enrolled)	95%	98%	99%	95%
Students in analysis	387	646	237	513
Gender				
Men	200 (51.7%)	293 (45.4%)	111 (46.9%)	222 (43.3%)
Women	153 (39.5%)	302 (46.7%)	106 (44.7%)	256 (49.9%)
Nonbinary	4 (1.0%)	2 (0.3%)	2 (0.8%)	2 (0.4%)
Unknown	30 (7.8%)	49 (7.6%)	18 (7.6%)	33 (6.4%)
Race or ethnicity				
Non-URM	291 (75.2%)	390 (60.4%)	174 (73.4%)	307 (59.8%)
URM	59 (15.2%)	164 (25.4%)	39 (16.5%)	140 (27.3%)
Unknown	37 (9.6%)	92 (14.2%)	24 (10.1%)	66 (12.9%)
Major				
Physics or engineering physics	57 (14.7%)	51 (7.9%)	16 (6.8%)	19 (3.7%)
Engineering	260 (67.2%)	475 (73.5%)	178 (75.1%)	403 (78.5%)
Other	37 (9.6%)	35 (5.4%)	20 (8.4%)	28 (5.5%)
Unknown	33 (8.5%)	85 (13.2%)	23 (9.7%)	63 (12.3%)
Year				
First-year	319 (82.5%)	594 (92.0%)	189 (79.8%)	480 (93.6%)
Second-year	52 (13.4%)	11 (1.7%)	37 (15.6%)	9 (1.7%)
Other or unknown	16 (4.1%)	41 (6.3%)	11 (4.6%)	24 (4.7%)

e.g., [34–37]) and covered topics in both mechanics and electromagnetism. For the lab course, students attended one 50 min lecture session (instructed by a faculty member of the physics department) and one 2 hr lab session (instructed by a graduate teaching assistant and often a supporting undergraduate teaching assistant) each week. The lecture sessions of the lab course included active learning pedagogies, such as students answering poll questions in small groups. The course was split into two lecture sections per semester, each with 200–300 students in a large stadium-seating lecture hall. During the lab sessions, which contained 20–25 students each, students conducted open-ended investigations in small groups of two to four and each group submitted lab notes at the end of every session to be graded. Lab groups were formed by the teaching assistants based on student preferences from a group-forming survey and remained the same for the whole semester. In forming the groups, the teaching assistants were advised to avoid lab groups containing an isolated woman. Outside of class, students completed individual lab homework assignments using Jupyter Notebook each week [52]. There were also multiple office hours per week where students could receive individual help on course content from graduate and undergraduate teaching assistants or the main instructor.

Most students in the lab course were simultaneously enrolled in one of two calculus-based mechanics *lecture courses*: one intended for physics majors (the "physics majors" course) and one intended for engineering and other science majors (the "nonmajors" course). In this paper, we only analyze the nonmajors lecture course (200–500 students) because the physics majors course only contained 30–50 students. Students in this lecture course attended three 50 min lecture sessions (instructed by a faculty member of the physics department) and two 50 min discussion sessions (instructed by a graduate teaching assistant and often a supporting undergraduate teaching assistant) each week. This course used active learning pedagogies including a "flipped classroom" model, such that students read relevant sections of the textbook and took a reading quiz before attending lecture. During lecture sessions, which contained half of the enrolled students at a time (there were two lecture sections per semester) and took place in a large stadium-seating lecture hall, students answered conceptual poll questions in small groups. The course also made extensive use of interactive lecture demonstrations. In the discussion sessions, which contained about 20 students each, students completed physics problems in small groups of two to four but this work was not submitted for a grade. Discussion groups were not formed by the teaching assistants, rather students formed their own groups. Students typically worked with the same discussion group every week. Outside of class, students completed individual homework assignments (problem sets) each week. There were multiple office hours per

week where students typically worked together on the homework assignments with the help of graduate and undergraduate teaching assistants.

In this study, all analyzed students in the lecture course were co-enrolled in the lab course. Therefore, it was possible for students to be surrounded by some of the same peers in both courses: the lecture and lab sessions of the lab course and the lecture and discussion sections of the lecture course. Between 20% and 40% of students in the lab course (depending on the semester), however, were not coenrolled in the lecture course we analyze.

B. Data collection

We administered an online network survey as part of a homework assignment in the lab course in the middle of the 15-week semester (see Fig. 1). On the survey, we distinguished peer recognition in the lab and lecture courses because our prior work identified that patterns of peer recognition varied between these instructional contexts [14]. Specifically, we asked students to nominate peers in each course who they believed were knowledgeable about the course material [13,14,30,31] as a measure of their recognition of strong peers. We also asked students to describe why they nominated their peers.

A second set of questions asked students to self-report peers with whom they had meaningful interactions about the instructional material in each course [41,53–56]. As in prior work, "students self-identified what counted as a meaningful interaction" [56] (p. 6). We asked students about whom they interacted with "this week" to capture interactions that students were consistently having with

Please list the students in this physics lab class that you think are particularly strong in the course material.

Name — *Please briefly explain why you chose this student as strong in the course material.*

Please list any students in your physics lecture class that you think are particularly strong in the course material.

Name — *Please briefly explain why you chose this student as strong in the course material.*

Please list any students in this physics lab class that you had a meaningful interaction* with about course material this week.

Name

Please list any students in your physics lecture class that you had a meaningful interaction* with about course material this week.

Name

*A meaningful interaction may mean in class, out of class, in office hours, virtually over zoom, through remote chat or discussions boards, or any other form of communication, even if you were not the main person speaking or contributing.

FIG. 1. Online survey prompts analyzed in this study. Students typed their responses into the text boxes and could enter up to 15 peers' names for each prompt. Students were given access to the course roster.

their peers throughout the semester, while reducing the possibility of recall bias (e.g., by asking them to recall all peers with whom they have interacted throughout the semester). This phrasing may have captured a few one-off interactions that only occurred the week of the survey, however these likely represent a small fraction of the reported interactions.

Each question was in an open response format, where students entered each peer's name in a separate text box and the associated explanation for each peer also in a separate text box. Students could enter up to 15 peers' names for each prompt, though no student provided the maximum number of names for any prompt. Students were also given access to the course rosters to facilitate their remembering and spelling of peers' names. Students could nominate anyone in their course; for example, they were not restricted to naming peers in their specific lecture section.

At least 95% of enrolled students in each course responded to the survey (see Table I). Students occasionally misspelled peers' names and/or reported just a first or a last name. In these cases, the first author manually processed the text to match the names in the survey responses to the course roster when possible. We could not match students if the respondent provided only a first (or last) name and multiple students in the course had that first (or last) name and so these responses were subsequently dropped from the dataset. In each course, we were able to match at least 90% of the nominations to strong peers and self-reported interactions to the course roster.

Our analysis included all students who responded to the survey and/or were listed by at least one peer on a given survey prompt. Our analysis also included only the nominations and self-reported interactions made by students who consented to participate in research (more than 95% of survey responders). If a consenting student wrote the name of a nonconsenting student, we included the survey response, but removed all information (e.g., demographics) about the nonconsenting student. We were able to apply social network analysis methods to our data because both the survey response rate and the name matching rate (from the raw survey responses to the course roster) were at least 90% and very few (<2%) nonconsenting students were removed from analysis, and network methods are reliable for datasets with less than 30% missing data [57].

We also collected students' self-reported gender, race or ethnicity, intended major, and academic year on the survey (see Table I). Most students in the dataset intended to major in engineering and the majority were in their first academic year. Each offering of the lab and lecture course contained roughly equal proportions of men and women. We grouped race or ethnicity by underrepresented racial minority (URM) status, where non-URM students are those solely identifying as White and/or Asian or Asian American and URM students are those identifying as at least one of any other race or ethnicity (including American Indian or Alaska Native, Black or African American, Hispanic or Latinx, and Native Hawaiian or other Pacific Islander). The majority of students (>60%) in each course were non-URM. Because the role of race or ethnicity was not part of our research questions, this categorization provided a limited ability to control for possible effects of race or ethnicity in our evaluation of the role of gender. We acknowledge, however, the limitations of this categorization [58] and encourage future work to probe this variable explicitly.

At the end of the semester, we collected students' discussion and lab section enrollment, lab groups, and final grades in each course.

C. Data analysis

We conducted our data analysis in four stages (summarized in Fig. 2), largely drawing on methods of social network analysis [55,59,60].

1. Recognition network structure

We first analyzed the four recognition networks, one for each offering (fall and spring) of each course (lab and lecture), using student responses to the first two questions on the survey (Fig. 1). Similar to prior work [14,30,31], we converted the nominations of strong peers into directed networks (see Fig. 4) to identify broad patterns of peer recognition. *Nodes* in the network represented students and *edges* (or *ties*) in the network represented all nominations made between students (including direction, from the nominator to the nominee). We distinguish nonbinary students from men, women, and students of unknown gender in the network diagrams (Fig. 4) to visualize these students' positions in each network. However, nonbinary students are not distinguished in the remainder of the analysis because they make up 1% or less of the student population in each course (Table I).
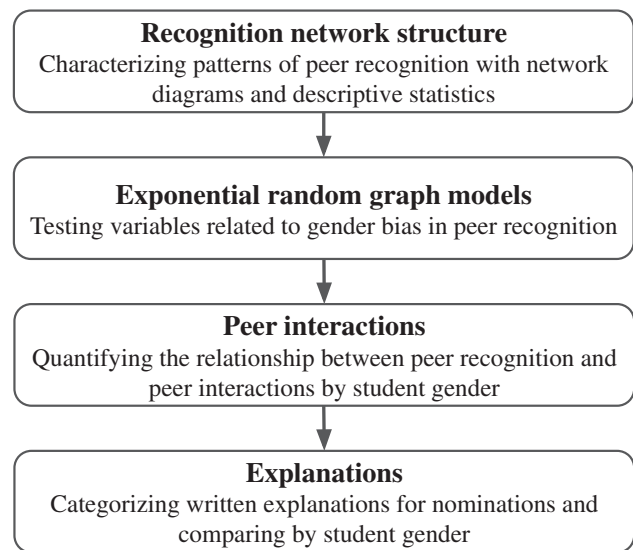

FIG. 2. Flowchart depicting our stages of data analysis.

To characterize the structures of the observed networks, we calculated three network-level statistics—*density*, *indegree centralization*, and *transitivity*—for each network. Density is the number of edges in the network that we observed as a fraction of the number of possible edges in the network. Indegree centralization measures the extent to which the nominations are concentrated around a single student or a small subset of students. This measure is calculated as the sum of differences in indegree (number of received nominations) between the node with the highest indegree (receiving the most nominations) and every other node in the network, divided by the maximum possible sum of differences of indegree for all nodes. Higher indegree centralization (i.e., closer to 1) indicates higher concentration of nominations around one or a few students (i.e., "celebrities" [30] who receive many more nominations than their peers). Finally, transitivity measures the tendency of nodes to cluster together and is calculated as the proportion of two paths (two edges connecting three nodes) that have a third edge closing the triangle, not considering edge direction. If node A is connected to node B and node C, for example, an edge between nodes B and C would form a triangle. A higher proportion of such triangles would lead to higher transitivity values (i.e., closer to 1).

We determined the standard errors of each of these statistics via bootstrapping: resampling the observed network many times, calculating the statistic of each sampled network, and then determining the standard deviation of the statistic among all of the sampled networks [54,61]. The bootstrapping was performed with 5000 bootstrap trials for each network using the *snowboot* package in R [62].

2. Exponential random graph models

We determined the extent to which a gender bias exists in each observed recognition network using exponential random graph models (ERGMs). Such models assume that an observed network is a realization from a random graph that comes from a distribution belonging to the exponential family [63,64]. ERGMs allow us to perform many statistical tests at once, determining whether the frequencies of certain configurations (e.g., ties between students of the same gender) in our observed network are significantly different than if the ties were formed randomly. The goal is to use these k configurations $g_k(y)$ and their corresponding coefficients θ_k to predict the formation of the random network Y. The model takes the form

$$P_\theta[Y = y] = \frac{\exp\left(\sum_k \theta_k g_k(y)\right)}{\sum_y \exp\left(\sum_k \theta_k g_k(y)\right)},$$

where y is a realization of the random network Y and the denominator serves as a normalization constant that ensures that the probability sums to one. Given an observed network y, the coefficients of the model are estimated using maximum likelihood estimation (MLE). Because of

the dependence between the network ties, the MLE is commonly approximated with Markov chain Monte Carlo (MCMC) techniques [65]. The coefficients θ_k represent log-odds of tie formation and can be interpreted as a weighting of the importance of each modeled configuration for the realized network, where positive (negative) coefficients show that the configuration is observed more (less) frequently than by chance after accounting for all other configurations that are modeled.

In our study, we fit an ERGM to each observed network using a similar set of configurations, or predictor variables, as our prior work [14]. For the lab course, we added two new variables. The first variable measured the tendency for students to nominate peers in their immediate lab group given prior work that suggests students often report connections to their group members on network surveys [66]. The second variable measured the tendency for students to nominate peers enrolled in their same lecture course (the separate course structure is different than in Ref. [14]). We also only measured discussion section homophily in the lecture course because our prior research found no significant tendency for students to nominate peers in their discussion section as strong in the lab material [14] and the discussion sections are now even further removed from lab material given the distinct courses. Different from our previous work, students received separate final course grades in the lab and lecture courses rather than one overall course grade that encompassed lab and lecture content. Therefore, we used the lab course final grades in the ERGMs for the lab course recognition networks and the lecture course final grades in the ERGMs for the lecture course recognition networks. Finally, we did not include a variable measuring transitivity in the models as we did in previous work [14] because the MCMC MLE did not converge with this variable added. The goodness-of-fit diagnostics, however, showed that our model sufficiently captured the distributions of indegree, outdegree (number of nominations reported by each student), and transitivity for all four observed networks (see Fig. 8 in Appendix A). The following predictor variables were included in our model:

- *Edges*: intercept term equal to the number of edges in the network.
- *Reciprocity*: number of mutual nominations (i.e., student A nominates student B and student B nominates student A).
- *Woman → woman*: number of edges for which a woman nominates another woman (base term is *man → man*).
- *Woman → man*: number of edges for which a woman nominates a man (base term is *man → man*).
- *Man → woman*: number of edges for which a man nominates a woman (base term is *man → man*).
- *URM → URM*: number of edges for which a URM student nominates a URM student (base term is *non-URM → non-URM*).

- *URM → non-URM*: number of edges for which a URM student nominates a non-URM student (base term is *non-URM → non-URM*).
- *Non-URM → URM*: number of edges for which a non-URM student nominates a URM student (base term is *non-URM → non-URM*).
- *Physics majors → physics majors* (lab course only): number of edges for which a student in the physics majors lecture course nominates another student in the physics majors lecture course (base term is *nonmajors → nonmajors*).
- *Physics majors → nonmajors* (lab course only): number of edges for which a student in the physics majors lecture course nominates a student in the nonmajors lecture course (base term is *nonmajors → nonmajors*).
- *Nonmajors → physics majors* (lab course only): number of edges for which a student in the nonmajors lecture course nominates a student in the physics majors lecture course (base term is *nonmajors → nonmajors*).
- *Lab group homophily*: number of edges connecting students in the same lab group.
- *Lab section homophily*: number of edges connecting students enrolled in the same lab section.
- *Discussion section homophily* (lecture course only): number of edges connecting students enrolled in the same discussion section.
- *Grade of nominee*: correlation between final course grade and number of received nominations.

We used the coefficient estimates of the *woman → woman*, *woman → man*, and *man → woman* variables for the four observed recognition networks to determine whether a gender bias exists in student nominations of strong peers after adjusting for the other network configurations included in the model. While we only focus on these gender variables in this paper, we keep the other variables in the model to account for as many different aspects of students' identity and participation in the course as possible and because an exploratory analysis indicated that removing these other variables can change the results for the gender variables [67]. In particular, a few of the predictor variables (e.g., *lab group homophily*, *lab section homophily*, and *discussion section homophily*) explicitly control for patterns of student interactions, allowing us to identify whether a gender bias in students' recognition of strong peers exists even after accounting for any strong interaction trends (e.g., gender homophily) [41].

Additionally, we note that the final four predictor variables listed above (related to lab group, lab section, discussion section, and grade) cannot handle unknown data. Therefore, only students with known data for these four variables were included in the ERGM analysis. While this predominantly restricted analysis to students who completed the course (i.e., students who did not have a final course grade likely dropped or withdrew from the course after we administered the survey), the ERGM analysis still included more than 90% of students that are part of our overall analysis. Thus, the statistical models provide an accurate description of most students in the class. We recommend for future work to investigate the network positionality of students who do not complete their physics course.

We also note that some sample sizes, particularly for URM students, seem too small to make statistical comparisons with our models (Table I). ERGMs, however, consider edges rather than nodes as the unit of analysis. Though the number of URM students (i.e., nodes) may be small, the networks we study include many of the possible edges between students (Table II and Fig. 4). Smaller sample sizes, furthermore, do not prevent valid estimation of the coefficient values. Instead, they are reflected in the standard errors and *p* values of the coefficients [68]. Quantitative modifications to ERGMs are only necessary for very small networks (less than six nodes) [69].

We finally note that students' final grades in the lab course were fairly skewed, with many students earning an A or A-. This may introduce issues of range restriction for the *grade of nominee* term, where low variability in students' final course grades limits the possibility of finding a significant correlation between grades and received nominations. We find in both offerings of the lab course, however, that the model is able to distinguish a significant effect of course grade on peer recognition (Table V).

3. Peer interactions

To measure the extent to which peer interactions are related to peer recognition, we converted students' self-reported interactions (last two questions on the survey, see Fig. 1) into directed networks. Similar to our analysis of the recognition network structure, we first calculated the density, the proportion of possible edges in the network that we observe, of each interaction network. For each offering of each course, we compared the recognition and interaction network densities to determine whether there were comparable numbers of edges in both networks or if one of the two networks contained many more edges than the other. This comparison was observational and not statistical.

For a more interpretable metric, we calculated the *percent overlap*, the percent of directed edges in the recognition network that also appear in the corresponding interaction network (see Fig. 3, for example), for each course. This measure allowed us to determine the extent to which peer recognition was *interaction based* or *observation based*. Higher percent overlap values indicate that most of students' nominations of strong peers were interaction-based recognition: many students nominated peers with whom they also reported interacting. For interaction-based recognition, we assume that the nominator came to understand the nominee's skill set through their direct interactions (e.g., talking to immediate group mates during lab). Lower percent overlap
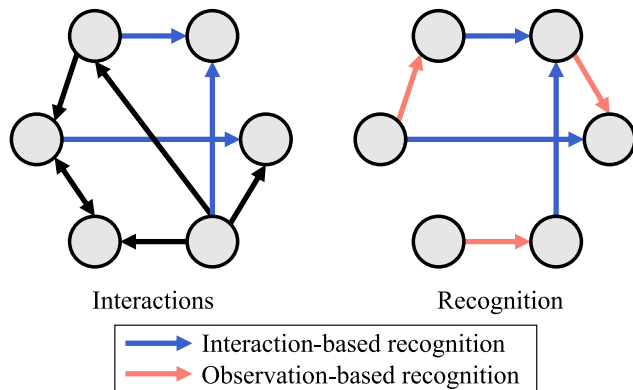
FIG. 3. Toy interaction and recognition networks to exemplify how we calculated percent overlap and fraction of interaction network edges kept in the recognition network. Black edges indicate edges that appear in the interaction network but not the recognition network, blue edges indicate edges that appear in both the interaction and recognition networks (edges kept), and orange edges indicate edges that appear in the recognition network but not the interaction network. In this case, the percent overlap is $\frac{1}{2}$ because three out of six recognition network edges also appear in the interaction network. The fraction of interaction network edges kept in the recognition network is $\frac{1}{3}$ because three out of nine interaction network edges also appear in the recognition network.

values, on the other hand, indicate that most of students' nominations of strong peers were observation-based recognition: many students nominated peers with whom they did not report interacting. For observation-based recognition, we assume that the nominator came to understand the nominee's skill set through their indirect observations of them (e.g., seeing someone frequently participate in lecture) rather than direct interactions.

We also calculated *gender homophily* as the percent of edges in the interaction network where both the nominator and the nominee are of the same gender (i.e., edges from men to men and edges from women to women). Research has shown that gender homophily is prevalent in student interaction networks [40,41,70], thus this measure helped us understand patterns between the interaction and recognition networks related to gender, discussed next.

Finally, we compared the extent to which any gender bias in peer recognition was specifically related to a gender bias in interaction-based recognition (i.e., if there was a gender bias in which of students' interaction ties they also nominated as strong in the course). This analysis only included edges for which both the nominator and the nominee self-reported their gender as either man or woman (859 out of 1000 total nominations of strong peers and 1590 out of 1789 total self-reported interactions across all four courses). For every possible combination of men and women nominating each other (i.e., man nominating a man, man nominating a woman, woman nominating a man, and woman nominating a woman), we calculated the *fraction of*

interaction network edges kept in the recognition network as the number of directed edges that appear in both the interaction and recognition networks divided by the number of directed edges in the interaction network (see Fig. 3, for example). We compared this measure by student gender in each network. We did not perform statistical tests because the goal of this analysis was to determine large-scale trends in the measure and relying on p values can be problematic [71–73]. Statistical tests of distinguishability would involve many comparisons that increase the risk of finding apparent statistical significance due only to chance. Instead, we use overlap in error bars (given by standard errors) to make qualitative interpretations about differences in the measure between men and women and do not comment on the possible distinguishability of small effects. While this approach is more appropriate than statistical testing, we acknowledge that using error bars may come with its own set of limitations [74].

We note that we could not measure the extent to which any gender bias in peer recognition was related to a gender bias in observation-based recognition. An appropriate measure of such a bias would be to calculate the fraction of peers that students indirectly observed, but did not interact with, that they nominated as strong in the course material. However, we did not collect data about which peers students indirectly observed—all we know is who they do not interact with. In large classes, such as the ones we analyze here, it is not reasonable to assume students indirectly observed all of the peers with whom they did not interact. We recommend for future work to investigate gender bias in observation-based recognition by examining peer recognition in small courses (where it is reasonable to assume students have the chance to indirectly observe all of their peers) or by also collecting data about peers with whom students are familiar but did not directly interact.

4. Explanations

We conducted a thematic coding analysis of student responses to the survey prompt, "Please briefly explain why you chose this student as strong in the course material," to identify what gets recognized in peer recognition—that is, the skill sets for which students recognized their strong physics peers. The first author initially read all responses to gain a sense of the data as a whole [75]. Upon recognizing similar themes to those identified in prior research [18,19,42–51], the first author drafted an *a priori* codebook informed by these themes. The research team then iteratively developed this codebook by individually coding a subset of the data and then meeting to modify the code definitions based on coding disagreements [76]. We also grouped together similar codes into four overarching categories: knowledge, processes, interactions, and other. These categories were inspired by those in Ref. [77], in which the authors categorized students' problem solving skills as related to knowledge, processes, and beliefs.

After the coding scheme was agreed upon, three members of the research team coded a stratified random sample of 10% of the 1000 total explanations in our dataset. We stratified the random sample by course (lab and lecture) because the two courses had different learning objectives and course structures, which might have led students to associate different skill sets with being strong in each course. Therefore, half of the random sample contained explanations from the lab course and the other half contained explanations from the lecture course. We determined interrater reliability by calculating Fuzzy Kappa [78] between each of the three pairs of coders because each explanation could receive multiple codes. All three pairwise Fuzzy Kappa values were greater than 0.8 indicating sufficient interrater reliability [78]. After reaching this level of agreement, the first author coded the remaining explanations.

We then compared the fractions of nominations between every possible combination of men and women nominating each other (i.e., man nominating a man, man nominating a woman, woman nominating a man, and woman nominating a woman) containing each code. This comparison only included explanations for which both the nominator and the nominee self-reported their gender as either man or woman (859 out of the 1000 total explanations across all four courses). We also aggregated the data from the fall and spring offerings because the results for each individual offering were not substantially different from the aggregated results and the larger dataset reduces possible statistical noise. Similar to our explanations analysis, we did not perform any statistical tests because the goal of this analysis was to identify any large-scale differences in these fractions. Instead, we used overlap in error bars (given by standard errors) to make qualitative interpretations about differences in code frequencies between men and women. This comparison allowed us to determine whether and how any gender bias in peer recognition in each course is related to what gets recognized (i.e., students nominating one another for different skill sets).

III. RESULTS

We present the results from each of the four stages of analysis (Fig. 2): recognition network structure, ERGM analysis, comparison to interaction networks, and explanations of nominations of strong peers.

A. Recognition network structure

The structural features of the observed recognition networks, summarized in Table II and shown in Fig. 4, provide information about broad patterns of recognition. We observe that the densities of the lab and lecture recognition networks are similar within each semester (fall and spring). Because there are many more nodes in the lab recognition network than the lecture recognition network in each semester (Table II), the densities suggest that there is a higher level of connectedness in the lab networks

than the lecture networks (i.e., a higher proportion of possible edges exist in the networks in the left column of Fig. 4 than in the corresponding network in the right column). We also observe relatively low indegree centralization values in each network. This observation indicates that the nominations are fairly spread out among the students in a course rather than concentrated around only one or a few students. Correspondingly, there are no outstanding "celebrities" in any network; no nodes are much larger than the rest, which would be associated with receiving many nominations (Fig. 4). While there is one man in the fall lecture network who receives more nominations than anyone else (seven), outstanding celebrities in the large science courses analyzed in prior work receive more than 30 nominations [14,30].

We also observe that all four networks contain one or two relatively large components that connect many nodes along chainlike formations and many smaller components of two to four nodes that are only connected to each other. The prevalence of these smaller components, however, is stronger in the lab course than the lecture course as indicated by the higher levels of transitivity. This pattern is likely due to the lab course placing more emphasis on small group work during the lab sessions (e.g., coordinating experimental investigations and submitting the lab notes for a group grade) than the lecture course does during the discussion sessions (e.g., collaborating on problems but not submitting work for a grade). We also see a large fraction of isolated nodes (at least 30% of the total nodes in each network), representing individuals who responded to the survey but did not nominate any peers as being strong in the course material and who were also not nominated by any other students. The demographics of the isolated nodes (e.g., gender and race or ethnicity) in each network are proportional to the demographics of the course population.

B. Exponential random graph models

As per our research questions, we focus on the coefficient estimates of the ERGM terms that speak to the extent to which a gender bias exists in the recognition networks (see Fig. 5 and Table V in Appendix B). We find that in both offerings of the lab course, women proportionately nominate men and women as strong in the course material as compared to men nominating men (red and orange dots on the left panel of Fig. 5). Women also proportionately nominate men and women as strong in the course material as compared to men nominating men in the fall offering of the lecture course (orange dots on the right panel of Fig. 5). Women nominate men less frequently than men nominate men, however, in the spring offering of the lecture course.

Additionally, in both offerings of both courses men disproportionately undernominate women as strong in the course material as compared to men nominating men (pink dots on both panels of Fig. 5). Although comparing ERGM

TABLE II. Network-level statistics for the observed recognition networks. Standard errors of the last digit are shown in parentheses.

	Lab course		Lecture course	
	Fall	Spring	Fall	Spring
Nodes	387	646	237	513
Density	0.002(2)	0.0009(8)	0.002(3)	0.0008(9)
Indegree centralization	0.011(4)	0.005(2)	0.03(1)	0.009(3)
Transitivity	0.24(3)	0.26(2)	0.13(3)	0.11(2)

coefficient values across different-sized networks is ill defined [79], we tentatively observe that this bias from men occurs to a similar extent in every course, with the possible exception of the spring offering of the lab course which has a slightly smaller coefficient estimate for the *man → woman* variable.

In all cases, the comparisons are made after adjusting for the other variables in our model. We particularly note that these results related to gender bias hold even after controlling for the gender composition of lab groups (*lab group homophily* variable), which were intentionally made to
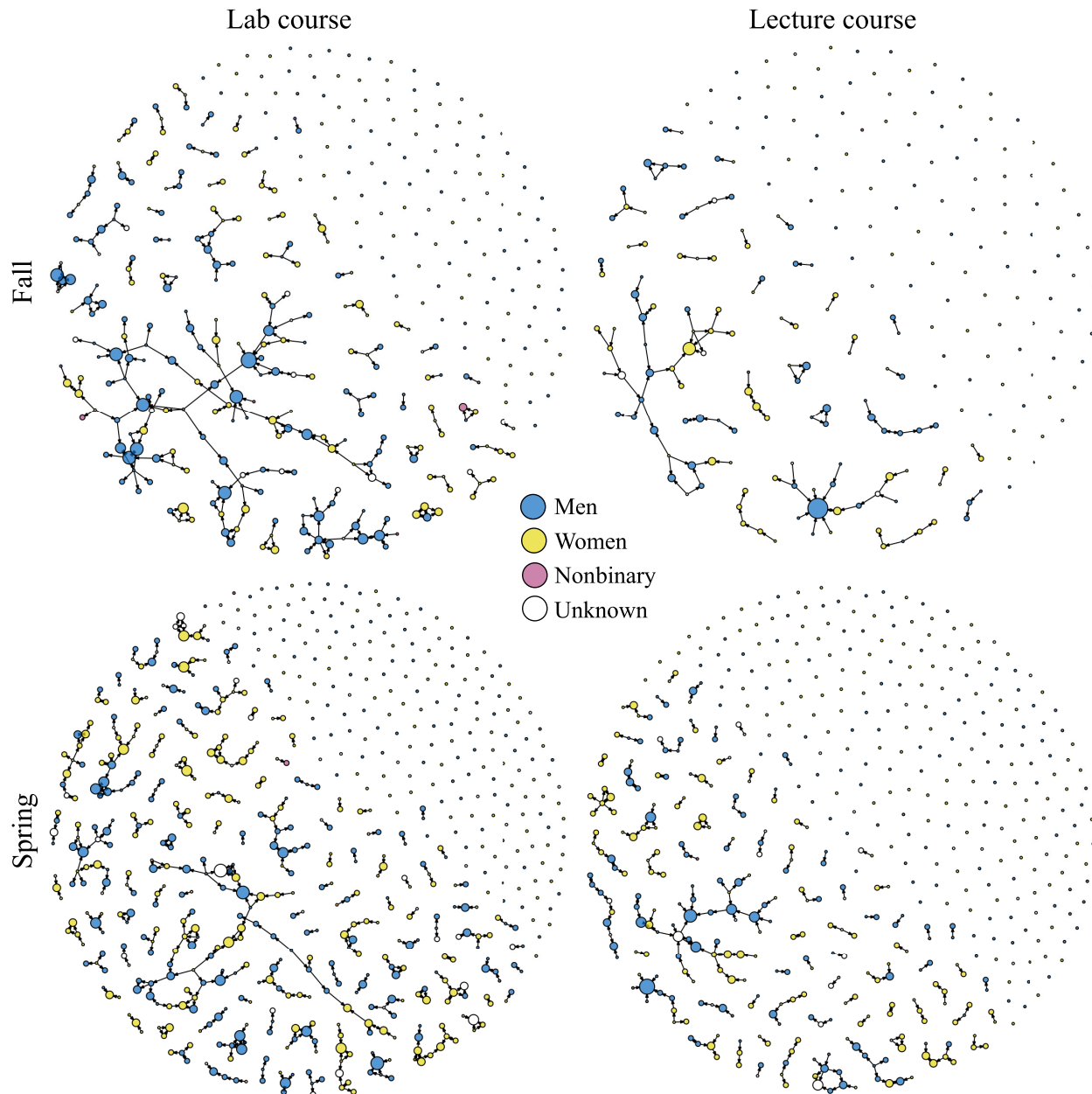

FIG. 4. Recognition networks for all analyzed courses. Nodes are colored by gender and sized proportional to indegree (number of received nominations as strong in the course). Edges point from the nominator to the nominee.
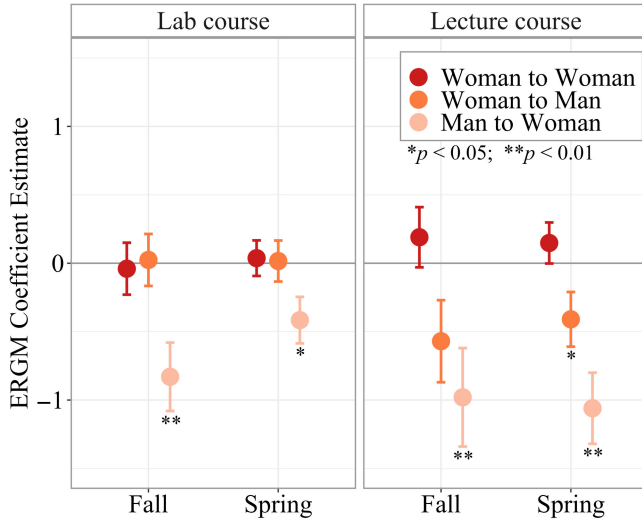
FIG. 5. Coefficient estimates, represented as log-odds, for the gender variables of the exponential random graph models (values shown in Table V in Appendix B). The base term (i.e., coefficient estimate of zero) is nominations from man to man. Error bars represent standard errors and asterisks indicate statistical significance.

avoid isolated women, and other patterns of student interactions (e.g., *lab section homophily* and *discussion section homophily* variables). In the lab course, the results also hold after controlling for any bias based on lecture course enrollment (*physics majors → physics majors*, *physics majors → nonmajors*, and *nonmajors → physics majors* variables). While the nonmajors lecture course is fairly gender balanced (Table I), the physics majors course contains a majority of men (70%–80%, depending on the semester). A bias in which nominations favor students in the physics majors course, therefore, would make men more likely to be nominated than women. The gender bias we observe, however, is present even after controlling for the lecture enrollment variables included in our model (and thus the different student populations in the two lecture courses).

C. Peer interactions

To further understand how students identify who to recognize as strong in their physics course, we also analyzed students' self-reported interactions. The interaction network diagrams are shown in Appendix C (see Fig. 9).

Comparing the densities of the interaction networks (Table III) to the densities of the corresponding recognition networks (Table II), we find in all four courses that the interaction network is more dense (i.e., students are more connected by edges) than the recognition network. This comparison may suggest that when making their nominations of strong peers, students select a subset of the peers with whom they interact to nominate as strong in the course material.

The percent overlap values, the percent of edges in the recognition network that are also in the corresponding

TABLE III. Densities and gender homophily of the observed interaction networks and percent overlap of the recognition and interaction networks. Standard errors of the last digit of the densities are shown in parentheses.

	Lab course		Lecture course	
	Fall	Spring	Fall	Spring
Density	0.003(3)	0.002(1)	0.005(5)	0.0015(9)
Percent overlap	56%	54%	57%	55%
Gender homophily	67%	67%	69%	72%

interaction network, however, indicate that interactions only account for a little more than half of the nominations of strong peers in each recognition network (Table III). That is, students also nominate peers with whom they did not report interacting. These results indicate that interaction-based recognition (e.g., learning about peers' skills by working together on a problem set or discussing concepts together in lecture) and observation-based recognition (e.g., learning about peers' skills by seeing a student ask a question in class or watching a student in a nearby lab group collect data for an experiment) occur with similar frequencies in the observed courses.

We examined the extent to which the gender biases in peer recognition identified in the ERGMs (in which men undernominate women in all courses and women undernominate men in the spring lecture course) are related to gender biases in the interaction-based nominations. We identified the fraction of students' interaction ties that they also nominate as strong in the course, split by gender of the nominee (Fig. 6). In three out of the four courses (all but the spring offering of the lab course), we observe that men disproportionately "keep" more of their interaction ties with men than with women (see top panel of Fig. 6). This pattern indicates that the gender bias coming from men that we observed in the ERGM analysis for these courses is (at least partly) related to a gender bias in interaction-based recognition. Men exhibit strong gender homophily in their peer interactions (Table III) and, after adjusting for this homophily (i.e., the metric shown in Fig. 6 is normalized by the number of interactions made to peers of a given gender), men exhibit a gender bias in which of their interaction ties they select as strong in the course material.

In the spring offering of the lab course, in contrast, men proportionately nominate their interaction ties to men and women as strong in the course (see top panel of Fig. 6). The gender bias in peer recognition coming from men in this course, therefore, is likely related to a gender bias in observation-based recognition, though we could not measure such a bias in our study (see Sec. II C 3).

Finally, the gender bias in the spring lecture course, in which women undernominate men as strong in the course, is (at least partly) related to interaction-based recognition: women disproportionately nominate more of their interaction ties with women than with men (see bottom panel of
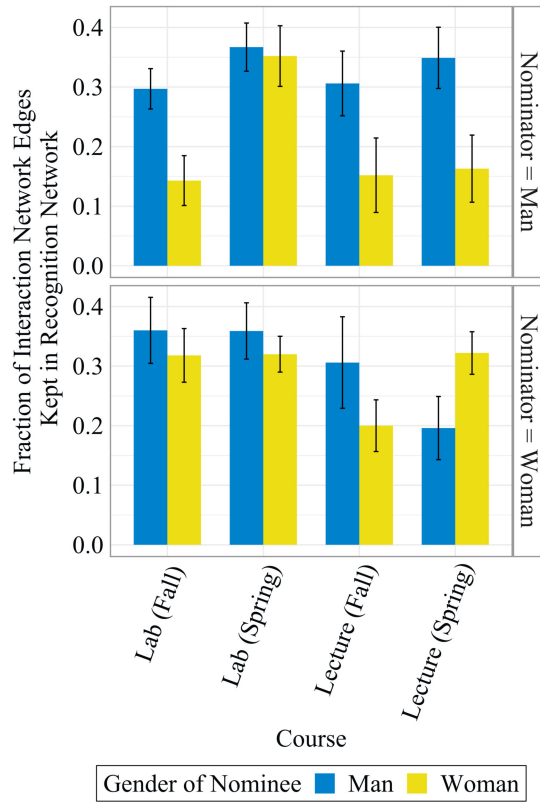
FIG. 6.   Fraction of edges in each interaction network that also appear in the corresponding recognition network for each combination of nominator and nominee gender. Error bars represent standard errors of the proportions.

Fig. 6). In all other semesters, where we do not observe a gender bias coming from women in the ERGMs, we correspondingly find that women nominate proportional numbers of men and women from their interaction ties as strong in the course material.

D. Explanations

We devised a coding scheme characterizing students' written explanations of their nominations of strong peers to determine what gets recognized in peer recognition—that is, what are the specific skill sets that students associate with being strong in their physics course (see Table IV). The coding scheme illuminates the features of peer interactions and indirect observations of peers that students consider when selecting which of their peers to nominate.

Related to knowledge, students describe strong peers as those who have *understanding* of the course material and have high *performance* (e.g., earn high grades). Students also mention that the peers they nominate have *experience* or background knowledge relevant to the course and have a *natural ability* for learning physics. Nominees were also described as hard-working or having *motivation* to learn the course material.

Students identify multiple processes associated with being strong in their physics courses. These codes are course specific. In the lab course, students describe strong peers as those who carried out the data *analysis* for their experiment, contributed to the *planning* or experimental design, participated in *data collection*, and engaged in *writing* up the lab

TABLE IV.   Definitions and examples of our coding scheme for students' explanations of why they nominated their peers as strong in the course material. Some codes were only present in either lab or lecture course nominations, indicated in parentheses. N indicates the total number of occurrences of each code in each course, lab or lecture.

Category or code	Definition	Example	N_{lab}	$N_{lecture}$
Knowledge				
Understanding	Knowledgeable about the course material	"Seems to have a strong sense of the topics in class."	228	112
Performance	Receiving good grades; answering questions correctly	"Did well on the quiz."	16	38
Experience	Having relevant background knowledge or experiences outside of the course	"Previous understanding of the class from AP Physics."	36	16
Natural ability	Having an innate aptitude for understanding the course material	"Has an innate ability to view a problem in its simplest terms."	6	12
Motivation	Putting a lot of time or effort into the course; determined	"Works harder than anyone I know to improve their physics lab knowledge."	88	44
Processes				
Analysis (lab only)	Analyzing and interpreting experimental data	"Has great understanding of software to help analyze data."	148	
Planning (lab only)	Designing and evaluating an experimental procedure	"Knows a lot about experimental design and what to do to answer the experimental question."	90	

(Table continued)

TABLE IV. *(Continued)*

Category or code	Definition	Example	N_{lab}	$N_{lecture}$
Data collection (lab only)	Carrying out an experimental procedure	"Good at experimental setup/ conducting trials with relatively small errors when possible."	26	
Writing (lab only)	Writing lab notes about an experiment	"Good at making detailed write-ups."	21	
Problem solving (lecture only)	Visualizing or reasoning through problems; applying the right equations to problems	"Very good at identifying the topics that a problem contains and quickly connecting it with a formula."		54
Interactions				
Helping	Providing support with the course material to others (nominator mentions benefits from this support)	"Helped me understand the homework."	77	58
Explaining	Describing or clarifying the course material to others (nominator does not mention benefits from this explaining)	"Explained concepts to me very well."	23	31
Participation	Active contributor to in-class discussions and activities; asking questions	"Participates effectively during the lecture."	54	7
Leading	Taking charge during in-class group work	"Organizes the group quite well and gets the group rolling, keeping us on track throughout the lab."	17	2
Other				
Other	Part or all of the explanation is vague or does not fit with above codes[†]	"Thinks outside the box."	52	22
None	No explanation provided	[Blank]	57	26

[†]A more detailed description of explanations coded as other, with examples, is provided in Appendix D.

notes. In the lecture course, students acknowledge the *problem solving* abilities of their strong peers.

Students also consider features of their interactions with others in the course when forming perceptions of their peers. Some students describe strong peers as *helping* them learn the course material and *explaining* the course material to others. Students also describe nominees as having high levels of verbal *participation* during lectures or group work and *leading* group work during in-class activities.

Finally, some students note *other* reasons that they viewed their peers as strong in the course material. These explanations are either too vague to associate with one of the above codes or too infrequent to create a separate code. Explanations left completely blank are coded as *none*.

Students proportionately recognize their interaction-based and observation-based nominees for each of these skills, except for process-related codes. In the lab course, students are more likely to describe their observation-based nominees than their interaction-based nominees as being strong in processes (specifically, data analysis). In the lecture course, students are more likely to describe their interaction-based nominees than their observation-based nominees as being strong in processes (specifically, problem solving). More detail about this comparison can be found in Appendix E (see Fig. 10).

Comparing fractions of nominations containing each code by the nominator and nominee gender, we find that in both the lab and lecture courses women proportionately nominate men and women for each category of the coding scheme [bottom row of Fig. 7(a)], with small differences for most individual codes [bottom row of Fig. 7(b)]. One interesting exception is that women disproportionately nominate men more than women for *understanding* in the lab course, though women do not exhibit a gender bias in their nominations in either offering of this course (Fig. 5).

In the lecture course, men also proportionately nominate men and women for each category despite the gender bias identified in the ERGMs [top right box of Fig. 7(a)]. In the lab course, however, men nominate men more than women for features of their interactions [top left box of Fig. 7(a)]. Examining the individual codes, we see that men disproportionately overnominate men as strong in the lab course for two of the individual interactions codes [top row of Fig. 7(b)]—*helping* and *explaining*—with the largest difference occurring for *helping*. It is important to note that these comparisons for individual codes are limited by small sample sizes (Table IV).

These results add nuance to our previous stages of analysis. The gender bias in the lecture course is seemingly related to who gets recognized (i.e., students disproportionately overnominating interaction ties to peers of their
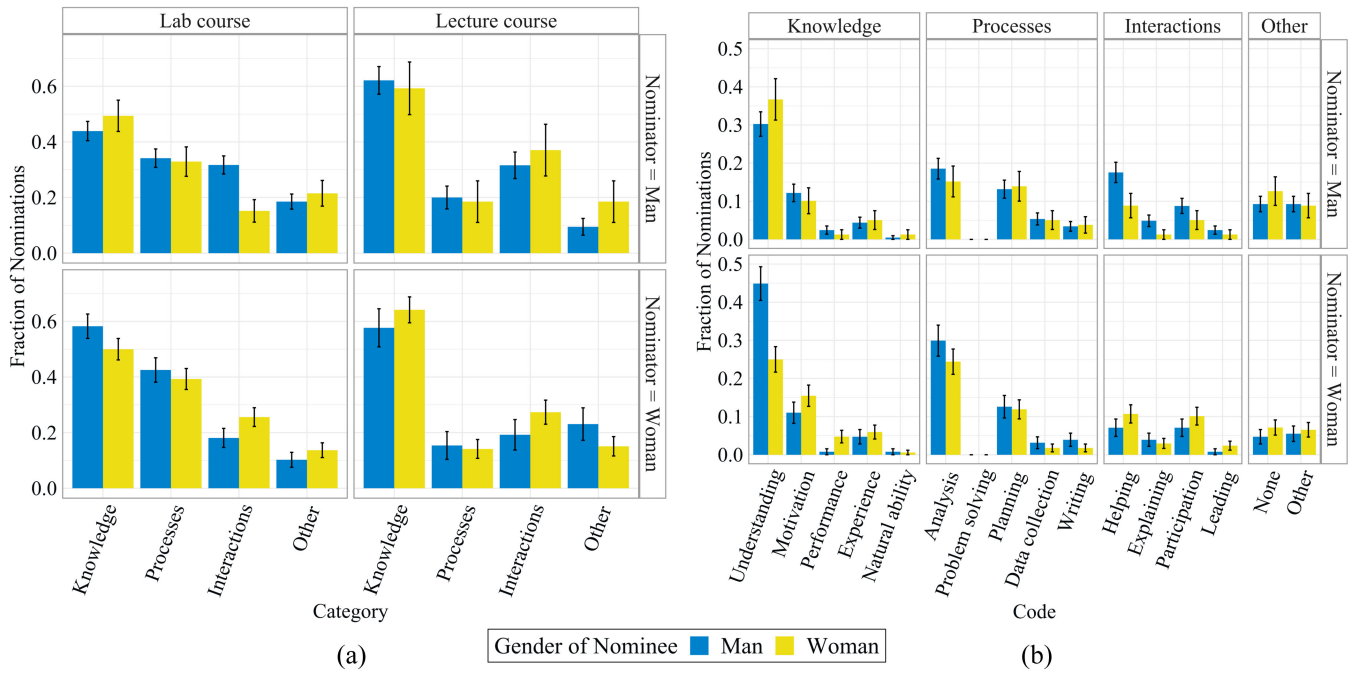
FIG. 7.   Fraction of nominations falling under each (a) category of our coding scheme within each course and (b) code of our coding scheme for the *lab course only*, split by gender of the nominator and nominee. Results are aggregated across the fall and spring offerings of each course. We coded 300 (205 in lab and 95 in lecture), 106 (79 in lab and 27 in lecture), 179 (127 in lab and 52 in lecture), and 274 (168 in lab and 106 in lecture) explanations from man to man, man to woman, woman to man, and woman to woman, respectively. Fractions do not necessarily sum to one because each explanation could receive multiple codes. Nominations made by men and women received comparable numbers of codes: men's nominations received an average of 1.4 codes and women's nominations received an average of 1.3 codes. Error bars represent standard errors of the proportions.

same gender) and not what gets recognized. In the lab course, however, the observed gender bias is also due to what gets recognized: men nominate more men than women because of the ways they interacted.

IV. DISCUSSION

In this study, we aimed to disentangle who and what gets recognized in peer recognition to better understand the nature of gender bias in such recognition (identified in, e.g., Refs. [13,14,30]). Across two offerings of distinct lab and lecture physics courses, we found that students determine who gets recognized in peer recognition in two ways, each with a similar frequency: interacting with peers and indirectly observing peers with whom they do not interact. We also identified what gets recognized in peer recognition: students mention skill sets related to knowledge, processes, and interactions in their written explanations of nominations.

In the following sections, we synthesize these results related to gender bias in peer recognition (in which men disproportionately undernominated women as strong in both courses and women disproportionately undernominated men as strong in one offering of the lecture course) and relate our findings to prior work. We also discuss other implications for research suggested by our analysis.

A. The nature of gender bias in peer recognition

In both offerings of the lecture course, we found a similar gender bias in peer recognition (in which men undernominated women compared to men) to previous work examining science courses aimed at first-year students during in-person and remote physics courses [13,14] and in-person biology courses [30]. Such a bias was not previously observed in in-person mechanical engineering courses [31] and remote physics courses [14] aimed at beyond-first-year students, which is likely due to students in those courses being more familiar with each other, as noted in Ref. [14]. Different from prior work [13,14,30], we also observed a bias from women in the spring offering of the lecture course: women disproportionately undernominated men as strong in this course. Surprisingly, we also found a gender bias in peer recognition (in which men disproportionately undernominated women) in both offerings of the lab course even though we did not previously observe a gender bias within a comparable lab context of a remote physics course (Course A in Ref. [14]). We build on the body of previous work by also probing the nature of these gender biases in peer recognition, finding that whether the gender bias is related to who and/or what gets recognized varies by instructional context, whether lecture or lab, described next.

1. Gender bias in the lecture course is related to who gets recognized, not what gets recognized

Our analysis suggests that in both offerings of the lecture course, the gender bias in peer recognition coming from men's nominations is related to men disproportionately overnominating men with whom they interact as compared to women with whom they interact as strong in the course material (i.e., there was a gender bias in interaction-based recognition). We observed that this bias is likely *not* attributable to men recognizing men and women for different skill sets because, of the people they recognize, men nominated men and women for similar skill sets.

The latter result is somewhat surprising given the research literature showing that students associate men and women with having different skill sets in lecture [44,46,49]. One study, for example, found that students associate men more than women with having a natural ability for learning physics and associate women more than men with asking questions [44]. If our observed gender bias in the lecture course is instead due to who gets recognized, rather than what gets recognized, then the observed bias may be due to gender stereotypes more broadly (i.e., students generally associating men more than women with being strong in physics) [3,7,9,17–26]. Alternatively, this gender bias may be related to prior literature's suggestion that students' social networks, and subsequently peer perceptions, exhibit strong gender homophily [40,41,70]. This interpretation is supported by our analysis of women's nominations, described next.

We found that women undernominated men in the spring offering of the lecture course, *opposite* of what prior work would predict (i.e., a gender bias in peer recognition in favor of men, rather than against men) [13,14,30]. In this offering, the observed bias was related to women nominating more of their interaction edges to women than men as strong in the course. Similar to nominations from men in this course, women proportionately nominated men and women for the skills identified in our analysis. The gender bias in the lecture course coming from both men and women, therefore, is related to who gets recognized (particularly, students nominating their same-gender interaction ties) and not what gets recognized.

2. Gender bias in the lab course is related to both who and what gets recognized

We found that the gender bias in the lab course (in which men disproportionately undernominated women as strong in the course) is related to a gender bias in interaction-based recognition in the fall offering (similar to the patterns observed in the lecture courses described above), but is likely related to a gender bias in observation-based recognition in the spring offering (though we could not measure this). Different from the lecture course, the nature of peer interactions *are also* a possible source of the gender bias in the lab course: men nominated men more than women for skills related to their interactions, such as helping them with the course material and explaining course material to others more generally. This result is consistent with prior work proposing that the "chilly climate" for women in physics may be due to the nature of their peer interactions rather than their number of peer interactions [66]. It is surprising, however, that the men and women in our study proportionally nominated one another for specific experimental skills in lab, such as handling the equipment to collect data or leading the group, despite evidence of students' gendered engagement in these roles [18,19,80].

One possible explanation for why we observed a gender bias (from men) in the in-person lab course but not the remote labs (Course A in Ref. [14]) is that these two labs had very different course structures, which ultimately shaped opportunities for peer interactions and observations. In the remote course, the lab was attached to the lecture course and almost all work for lab was done during the lab sessions within lab groups in isolated breakout rooms on Zoom. The lab had no whole-class lecture session or summative assessments (e.g., exams or quizzes) and students' lab grades were combined with their lecture course grade. The in-person lab course, in contrast, included a whole-class lecture session that included peer instruction activities and three summative quizzes. Students also received a separate course grade for the lab material. We suspect that these changes in course structure facilitated more out-of-lab-group interactions, including both in-lecture interactions and out-of-class interactions. Out-of-class interactions may also have been more common during the in-person course simply due to the evolving nature of the COVID-19 pandemic.

3. Summary

The results from both courses add nuance to our previous work which found that the presence or absence of a gender bias in peer recognition varies between the instructional contexts of lab and lecture [14]. In this study, we observed a gender bias in peer recognition (either from both men and women or just men) in *both* the lab and lecture course but found that the sources of this bias varied—whether only due to a bias in who gets recognized (lecture) or also due to a bias in what gets recognized (lab). These results may suggest that pedagogical style (e.g., having both lecture and small-group sessions related to instructional material) may impact gender bias in peer recognition more than the instructional material itself. Future investigations of peer recognition, however, should continue to analyze peer recognition in these instructional contexts separately and should examine the nature of gender bias in peer recognition in other lab and lecture contexts, such as those at other institutions, with students from other types of majors, or with students in studio physics courses. Researchers should also more closely investigate how students develop perceptions of strong physics peers, such as through interviews, that may better illuminate some of

the differences we observed between courses and semesters and between men's and women's nominations.

B. Other implications for research

Here we synthesize our findings that are not directly tied to the research questions of this study with those of previous research and suggest directions for future work.

1. Structure of peer recognition networks

The structures of the peer recognition networks in this study differ from those observed in prior work. Specifically, a relatively small fraction of nodes in the recognition networks of the in-person lecture courses in this study comprised the *giant component* (the largest cluster of connected nodes), with connections instead forming short chainlike structures and some small, disconnected components. In previous studies [14,30,31], in contrast, the lecture recognition networks for both in-person and remote science courses had relatively large giant components (i.e., a large proportion of the nodes connected together in a main cluster). We propose that this difference in network structure is due to differences in course structure and populations. Refs. [30,31], for example, examined second-semester (or later) in-person courses within a multicourse sequence, when students have likely developed familiarity with one another. Thus, students could likely identify many of their peers, including outspoken students in lecture who then become celebrities in the network. Similarly, in the remote physics courses studied in Ref. [14], Zoom likely expedited name familiarity and connections across many different peers (not just those in close physical proximity) even for students in the first course of the sequence. The in-person lecture courses analyzed in this study, however, are the first in the physics sequence when students are likely still getting to know one another and each other's names (e.g., even if some students are really outspoken during in-person lectures, other peers likely do not know their name). Future work should examine this effect of peer familiarity on peer recognition further through a longitudinal study, for example, modeling how recognition networks change (or not) throughout a course sequence (e.g., similar to Ref. [81] in which the authors analyze changes in peer interaction networks over time).

We also observed lower transitivity (small group clustering) in the peer recognition networks of in-person lab courses (this study) than the peer recognition networks of remote physics labs [14]. This difference is likely due to the different course structures mentioned above and the different instructional modalities. In the remote physics labs [14], students only attended low-enrollment lab sessions where they worked in Zoom breakout rooms with the same peers every week. Therefore, students had limited access to peers (and these peers' names) outside of their immediate lab group. Students in the in-person lab course (this study), however, could see and interact with other students in their lab section, including those outside of their immediate lab

group. These students also attended a large lecture each week where they could interact with peers outside of their lab group and lab section entirely. Increased visibility of out-of-lab-group peers within individual lab sections and the addition of a lecture session to the lab, therefore, likely increased the number of peers students had access to, which in turn reduced the amount of small, isolated clusters in the recognition network. Future work should investigate peer recognition in other course structures containing structured small group activities, such as studio physics and modeling instruction, where patterns of recognition might also vary.

2. How does peer recognition form?

We found that both peer interactions and indirect observations of peers play an important role in shaping peer recognition, in line with what prior work suggests [15,30,38,39]. Our analysis, therefore, illuminates the complexities behind how students form perceptions of their peers (i.e., interaction-based versus observation-based recognition). Future work should further investigate the relationship between interactions and peer recognition, for example by conducting egocentric network analyses to probe the more fine-grained social processes underlying the development of peer recognition. Our results also suggest that future analyses of peer recognition should continue to probe and analyze interaction networks alongside student nominations of strong peers. Examining both networks together will likely provide a more nuanced understanding of how patterns of recognition form.

Finally, we devised a coding scheme to identify the different skill sets for which students nominate their peers as strong in their physics course. These codes encompassed a variety of skills students identified in their explanations of nominations: knowledge (e.g., content knowledge, getting high grades, and having a natural ability to learn physics), engagement in processes (e.g., designing an experiment, analyzing data, and solving problems), and interactions (e.g., helping, explaining, and leading). Such skills have been identified across qualitative research studies investigating how students define being a good physics student or a good physicist [18,19,42–51], however our study is the first to evaluate these skills at the scale of a large introductory physics course. Future work should examine whether this coding scheme is applicable to explanations from students in other institutions or experiencing different instructional styles and pedagogies.

C. Limitations

We end this section by acknowledging the limitations of our study. First, the online network survey may not have captured all nominations of strong peers and all peer interactions. Students may not have remembered the names of individuals they perceived as strong in the material and/ or with whom they interacted, for example due to recall bias. We also only collected survey responses in the middle

of the semester. Previous work administered surveys either both at the middle and end of the course [30] or only at the end of the course [31] and so this methodological choice allowed us to compare our results to that previous work. Future research examining in-person physics students' recognition of strong peers at multiple points in their physics course, or just at the end of their physics course, however, may add nuance to our results.

Additionally, we categorized students' race or ethnicity by URM status because the number of students in some of the individual racial or ethnic groups was too small for our statistical analysis to produce useful and interpretable results. This treatment of race or ethnicity, however, inevitably masks differences in peer recognition between students of individual racial or ethnic groups. Because our research questions were focused on the role of gender in peer recognition, this categorization allowed us to account for race and ethnicity in a cursory way in our analysis. Future work should seek to study the role of race and ethnicity in peer recognition explicitly by using more diverse student populations with statistically sufficient representation across racial or ethnic groups. Researchers should also aim to differentiate peer recognition among White and Asian students and between Asian and Asian American students [82,83].

We also observed more sparse recognition networks than those in prior work [14,30,31], particularly with regard to the proportion of students who were isolates (nodes with zero adjacent edges in the network). It is impossible for us as researchers to know if these isolates are "true isolates," students who truly have zero nominations to make or receive, or if they are essentially non-respondents who fill out the survey quickly and do not nominate anyone, even though they may actually have nominations to make. Therefore, our analysis may have missed some nominations and the network structures may not represent all connections between students. Fortunately, social network data is robust to up to 30% missing data [57] and our recognition networks contain a range of 32%–48% isolates. Assuming some students are true isolates, our analysis likely captures a fairly accurate picture of peer recognition. Furthermore, all of our ERGMs converged with appropriate goodness-of-fit diagnostics, indicating that the statistical analysis is reliable; any uncertainties due to small sample sizes are reflected in the standard errors and p values of the coefficient estimates.

Finally, the majority of our analysis assumes that any overlap in the recognition and interaction networks implies that peer recognition was formed through peer interactions. Alternatively, interactions may plausibly be formed through recognition, such as by a student deciding to interact with peers they perceive as strong in the course material. We believe the former is more likely than the latter in this study because the courses we analyze are the first in the course sequence, when students have likely not met each other before, and because the survey was administered in the middle of the semester, when students may not have a strong sense of each other's grades. Nonetheless, future work should seek to disentangle this relationship between interactions and recognition, if possible.

V. CONCLUSION

In this study, we observed a gender bias in student nominations of strong physics peers in distinct lab and lecture courses. In both courses, men undernominated women as strong in the course. Women also undernominated men as strong in one offering of the lecture course. To understand the nature of this gender bias in peer recognition, we built upon two existing threads of research. First, we investigated the role of peer interactions in students' determination of who gets recognized in peer recognition. Second, we examined what gets recognized—whether the skills students associated with being a strong physics student or strong physicist are related to the gender bias in peer recognition. We found that roughly half of nominations of strong peers formed through peer interactions, with the remaining nominations likely coming from indirect observations of peers. We also observed that the nature of the gender bias varied between the lab and lecture courses. In the lecture course, the bias was related to who gets recognized: both men and women disproportionately overnominated their interaction ties with students of their same gender as strong in the course material. At the same time, men and women nominated men and women for similar skill sets in this course. In the lab course, in contrast, men also disproportionately undernominated women for certain skill sets, particularly those related to their interactions, such as being helpful. These results add nuance to our understanding of how students form perceptions of strong peers in their physics courses and prompt future work to examine the nature of gender bias in peer recognition in other course structures, instructional styles, and student populations.

ACKNOWLEDGMENTS

APPENDIX A: ERGMs: GOODNESS-OF-FIT DIAGNOSTICS

Figure 8 shows one example (for the fall offering of the lecture course) of the diagnostics used to assess the goodness-of-fit of ERGMs. When an ERGM fits an observed network well, the thick black line always falls within the boxplots. Because this is the case in Fig. 8 and for the other three analyzed courses, our model sufficiently captures the characteristics of the observed networks.
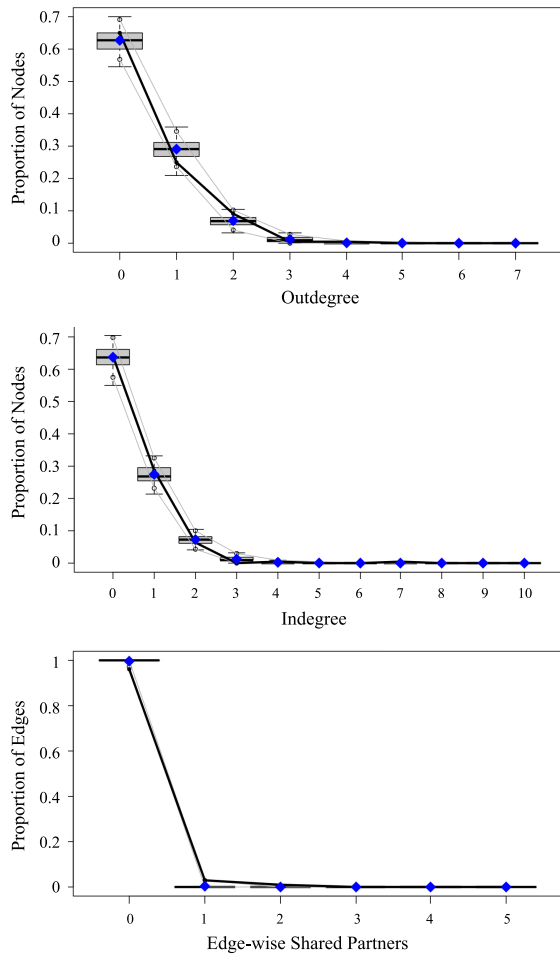
FIG. 8. ERGM goodness-of-fit diagnostics for the recognition network of the fall offering of the lecture course. The horizontal axis represents a network measure (outdegree, indegree, or edgewise shared partners—a measure of transitivity) and the vertical axis represents either the proportion of nodes or proportion of edges in the network. Plots compare the distribution of each network measure for the observed data (thick black line) to that for 10 network simulations generated using the estimated model coefficients (box plots).

APPENDIX B: ERGMs: FULL MODEL RESULTS

Table V shows the coefficient estimates for all predictor variables in our exponential random graph models, fit to each of the four analyzed networks.

APPENDIX C: INTERACTION NETWORK DIAGRAMS

Figure 9 shows the network diagrams of students' self-reported peer interactions for all four analyzed courses.

APPENDIX D: EXPLANATIONS CODING SCHEME: *Other* CODE EXAMPLES

We coded an explanation as *other* if it was not related to one of the other codes (see Table IV) or if the reason(s) provided in the explanations did not appear often enough in the full dataset to form a new code. Examples of vague explanations coded as *other* are "Good at working in the lab" and "She's amazing." Examples of ideas that were too infrequent to form their own code are:

- drawing (e.g., "She draws good.")
- open-minded (e.g., "[He] is open-minded and explains concepts materials when I don't understand.")
- patient (e.g., "Patient with other members.")
- listening (e.g., "He listens to new ideas and knows the course material very well.")
- revising their own thinking (e.g., "She understands and is confident in the course material and can iteratively revise her thinking.")
- creative (e.g., "He quickly comes up with good ideas for experiments and is very creative with collecting data.")
- hearing from someone else that the nominee is strong in the course (e.g., "I've heard great things of this man.").

TABLE V. Coefficient estimates for the exponential random graph models, represented as log-odds, fit to each observed recognition network. Standard errors are given in parentheses. Asterisks indicate statistical significance ($^*p < 0.05$; $^{**}p < 0.01$).

	Lab course		Lecture course	
	Fall	Spring	Fall	Spring
Edges	−10.37** (0.73)	−11.11** (0.56)	−9.49** (0.87)	−10.70** (0.48)
Reciprocity	−0.14 (0.31)	0.46* (0.23)	2.95** (0.52)	4.84** (0.32)
Woman → woman	−0.04 (0.19)	0.04 (0.13)	0.19 (0.22)	0.15 (0.15)
Woman → man	0.02 (0.19)	0.02 (0.15)	−0.57 (0.30)	−0.41* (0.20)
Man → woman	−0.83** (0.25)	−0.42* (0.17)	−0.98** (0.36)	−1.06** (0.26)
URM → URM	0.19 (0.44)	−0.23 (0.24)	0.92* (0.41)	0.47 (0.25)
URM → non-URM	0.01 (0.25)	0.10 (0.15)	0.19 (0.29)	−0.17 (0.20)
Non-URM → URM	−0.25 (0.26)	0.08 (0.16)	0.30 (0.31)	−0.11 (0.24)
Physics majors → physics majors	1.36** (0.41)	1.59** (0.43)		

(Table continued)

TABLE V. *(Continued)*

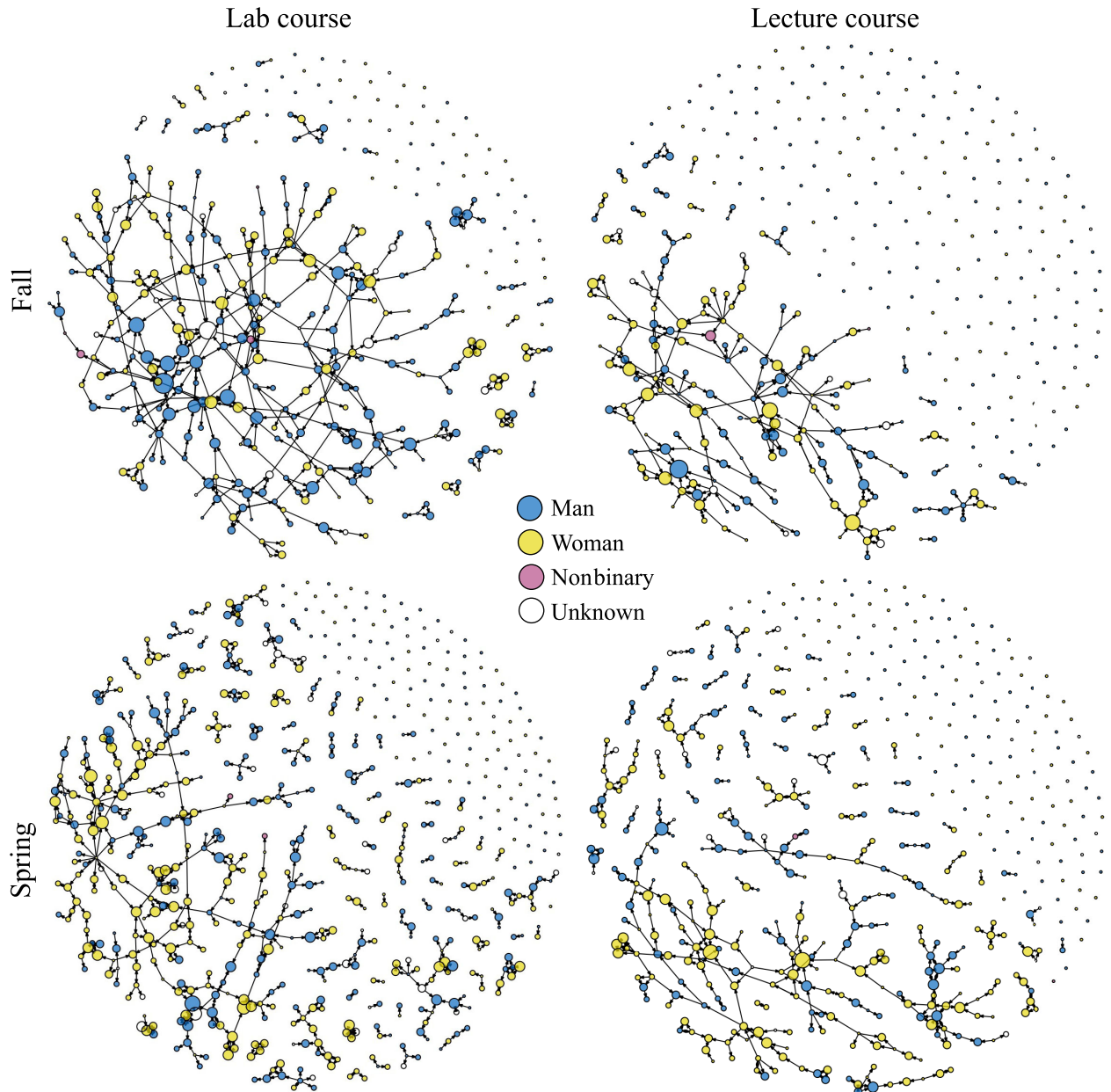|	Lab course		Lecture course	
	Fall	Spring	Fall	Spring
Physics majors → nonmajors	−0.96* (0.40)	−0.07 (0.28)		
Nonmajors → physics majors	0.20 (0.25)	0.69** (0.21)		
Lab group homophily	4.64** (0.26)	4.20** (0.18)	1.70** (0.56)	2.07** (0.38)
Lab section homophily	1.96** (0.27)	2.50** (0.20)	−0.04 (0.47)	0.17 (0.33)
Discussion section homophily			2.14** (0.21)	1.62** (0.32)
Grade of nominee	0.71** (0.19)	0.75** (0.15)	0.79** (0.23)	1.01** (0.13)

FIG. 9. Interaction networks for all analyzed courses. Nodes are colored by gender and sized proportional to indegree (number of received nominations). Edges point from the nominator to the nominee.

APPENDIX E: EXPLANATIONS SPLIT BY INTERACTION-BASED AND OBSERVATION-BASED NOMINATIONS

Figure 10 shows the fraction of interaction-based nominations (i.e., nominations of strong peers with whom the nominee also reported interacting) and observation-based nominations (i.e., nominations of strong peers with whom the nominee did not report interacting) falling under each category or code from our coding scheme.
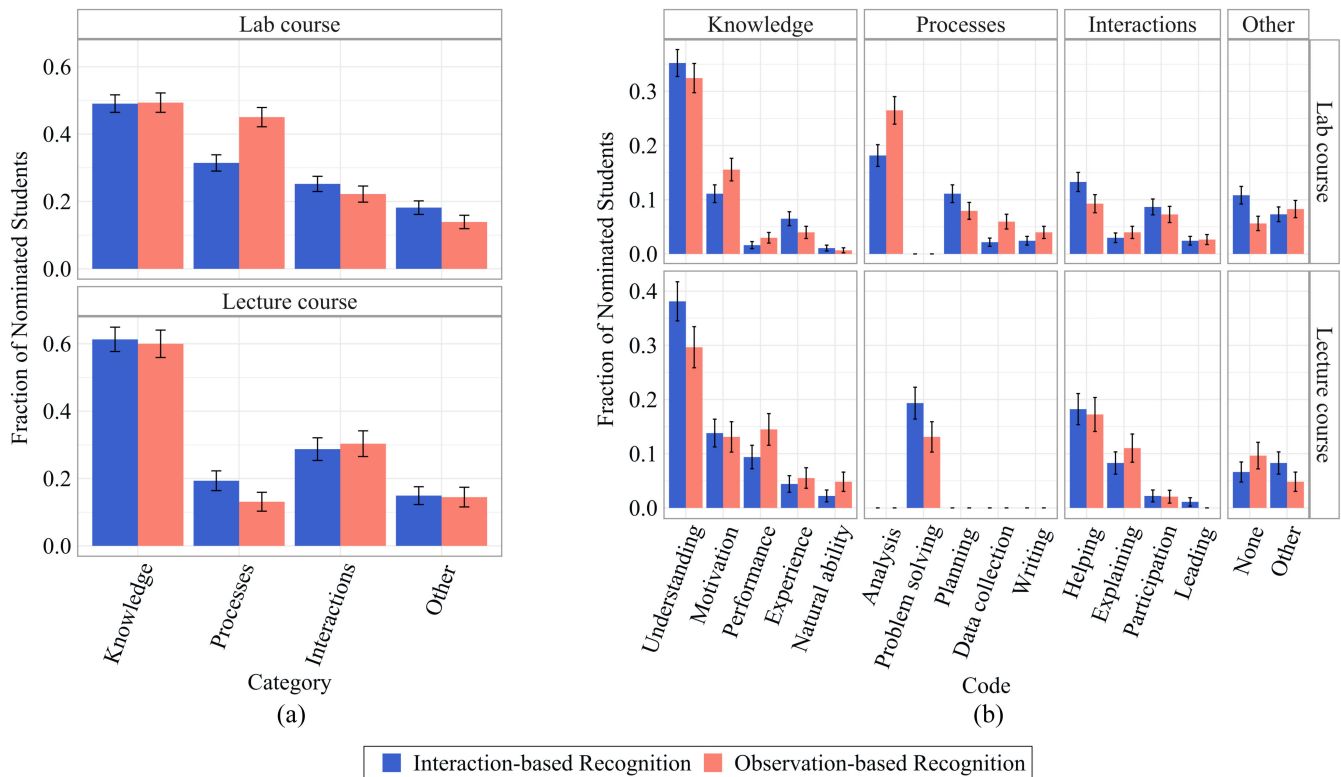

FIG. 10. Fraction of nominations that are interaction-based and observation-based falling under each (a) category and (b) code of our coding scheme, split by course. Fractions do not necessarily sum to one because each explanation could receive multiple codes. Error bars represent standard errors of the proportions.

[1] Simone Hyater-Adams, Claudia Fracchiolla, Noah Finkelstein, and Kathleen Hinko, Critical look at physics identity: An operationalized framework for examining race and physics identity, Phys. Rev. Phys. Educ. Res. **14,** 010132 (2018).

[2] Robynne M. Lock, Zahra Hazari, and Geoff Potvin, Physics career intentions: The effect of physics identity, math identity, and gender, AIP Conf. Proc. **1513,** 262 (2013).

[3] Heidi B. Carlone and Angela Johnson, Understanding the science experiences of successful women of color: Science identity as an analytic lens, J. Res. Sci. Teach. **44,** 1187 (2007).

[4] Zahra Hazari, Gerhard Sonnert, Philip M. Sadler, and Marie-Claire Shanahan, Connecting high school physics experiences, outcome expectations, physics identity, and physics career choice: A gender study, J. Res. Sci. Teach. **47,** 978 (2010).

[5] Zahra Hazari and Cheryl Cass, Towards meaningful physics recognition: What does this recognition actually look like?, Phys. Teach. **56,** 442 (2018).

[6] Zahra Hazari, Eric Brewe, Renee Michelle Goertzen, and Theodore Hodapp, The importance of high school physics teachers for female students' physics identity and persistence, Phys. Teach. **55,** 96 (2017).

[7] Z. Yasemin Kalender, Emily Marshman, Christian D. Schunn, Timothy J. Nokes-Malach, and Chandralekha Singh, Gendered patterns in the construction of physics identity from motivational factors, Phys. Rev. Phys. Educ. Res. **15,** 020119 (2019).

[8] Maria Vetleseter Bøe, Staying recognised as clever: High-achieving physics students' identity performances, Phys. Educ. **58**, 035012 (2023).

[9] Lucy Avraamidou, Identities in/out of physics and the politics of recognition, J. Res. Sci. Teach. **59**, 58 (2022).

[10] Yangqiuting Li and Chandralekha Singh, The impact of perceived recognition by physics instructors on women's self-efficacy and interest, Phys. Rev. Phys. Educ. Res. **19**, 020125 (2023).

[11] Zahra Hazari, Philip M. Sadler, and Gerhard Sonnert, The science identity of college students: Exploring the intersection of gender race, and ethnicity, J. Coll. Sci. Teach. **42**, 82 (2013), http://www.jstor.org/stable/43631586.

[12] Ewan Bottomley, Antje Kohnle, Ken Mavor, Paula Jean Miles, and Vivienne Wild, The relationship between gender and academic performance in undergraduate physics students: The role of physics identity, perceived recognition, and self-efficacy, Eur. J. Phys. **44**, 025701 (2022).

[13] Brittany Bloodhart, Meena M. Balgopal, Anne Marie A. Casper, Laura B. Sample McMeeking, and Emily V. Fischer, Outperforming yet undervalued: Undergraduate women in STEM, PLoS One **15**, e0234685 (2020).

[14] Meagan Sundstrom, Ashley B. Heim, Barum Park, and N. G. Holmes, Introductory physics students' recognition of strong peers: Gender and racial or ethnic bias differ by course level and context, Phys. Rev. Phys. Educ. Res. **18**, 020148 (2022).

[15] James Paul Gee, Chapter 3: Identity as an analytic lens for research in education, Rev. Res. Educ. **25**, 99 (2000).

[16] Z. Yasemin Kalender, Emily Marshman, Christian D. Schunn, Timothy J. Nokes-Malach, and Chandralekha Singh, Why female science, technology, engineering, and mathematics majors do not identify with physics: They do not think others see them that way, Phys. Rev. Phys. Educ. Res. **15**, 020148 (2019).

[17] Cathrine Hasse, Gender diversity in play with physics: The problem of premises for participation in activities, Mind Cult. Act., **9**, 250 (2002).

[18] Anna T. Danielsson, Exploring woman university physics students 'doing gender' and 'doing physics', Gender Educ. **24**, 25 (2012).

[19] Allison J. Gonsalves, Anna Danielsson, and Helena Pettersson, Masculinities and experimental practices in physics: The view from three case studies, Phys. Rev. Phys. Educ. Res. **12**, 020120 (2016).

[20] Ursula Kessels, Melanie Rau, and Bettina Hannover, What goes well with physics? Measuring and altering the image of science, Br. J. Educ. Psychol. **76**, 761 (2006).

[21] Elena Makarova and Walter Herzog, Trapped in the gender stereotype? The image of science among secondary school students and teachers, Equality, Diversity Inclusion **34**, 106 (2015).

[22] Elena Makarova, Belinda Aeschlimann, and Walter Herzog, The gender gap in STEM fields: The impact of the gender stereotype of math and science on secondary students' career aspirations, Front. Educ. **4**, 60 (2019).

[23] Erika D. Tate and Marcia C. Linn, How does identity shape the experiences of women of color engineering students?, J. Sci. Educ. Technol. **14**, 483 (2005).

[24] Robert Ceglie, Underrepresentation of women of color in the science pipeline: The construction of science identities, J. Women Minorities Sci. Engin. **17**, 271 (2011).

[25] Corinne A. Moss-Racusin, John F. Dovidio, Victoria L. Brescoll, Mark J. Graham, and Jo Handelsman, Science faculty's subtle gender biases favor male students, Proc. Natl. Acad. Sci. U.S.A. **109**, 16474 (2012).

[26] Sarah L. Eddy, Sara E. Brownell, and Mary Pat Wenderoth, Gender gaps in achievement and participation in multiple introductory biology classrooms, CBE Life Sci. Educ. **13**, 478 (2014).

[27] Linda J. Sax, Kathleen J. Lehman, Ramón S. Barthelemy, and Gloria Lim, Women in physics: A comparison to science, technology, engineering, and math education over four decades, Phys. Rev. Phys. Educ. Res. **12**, 020108 (2016).

[28] Angela Johnson, Rose Young, and Elizabeth Mulvey, Where do women of color complete physics degrees?, Phys. Teach. **58**, 620 (2020).

[29] Anne Marie Porter and Rachel Ivie, *Women in Physics and Astronomy, 2019* (AIP Statistical Research Center, College Park, MD, 2019).

[30] Daniel Z. Grunspan, Sarah L. Eddy, Sara E. Brownell, Benjamin L. Wiggins, Alison J. Crowe, and Steven M. Goodreau, Males under-estimate academic performance of their female peers in undergraduate biology classrooms, PLoS One **11**, e0148405 (2016).

[31] Shima Salehi, N. G. Holmes, and Carl Wieman, Exploring bias in mechanical engineering students' perceptions of classmates, PLoS One **14**, e0212477 (2019).

[32] Paula Heron and Laurie E. McNeil, *Phys21: Preparing Physics Students for 21st Century Careers, technical report, Joint Task Force on Undergraduate Physics Programs* (American Physical Society and American Association of Physics Teachers, College Park, MD, 2016).

[33] Joseph Kozminski, Heather Lewandowski, Nancy Beverly, Steve Lindaas, Duane Deardorff, Ann Reagan, Richard Dietz, Randy Tagg, M EblenZayas, J. Williams *et al.*, *AAPT Recommendations for the Undergraduate Physics Laboratory Curriculum* (American Association of Physics Teachers, College Park, MD, 2014), p. 29.

[34] N. G. Holmes and Carl E. Wieman, Introductory physics labs: We can do better, Phys. Today **71**, No. 1, 38 (2018).

[35] Emily M. Smith and N. G. Holmes, Best practice for instructional labs, Nat. Phys. **17**, 662 (2021).

[36] Emily M. Smith, Martin M. Stein, Cole Walsh, and N. G. Holmes, Direct measurement of the impact of teaching experimentation in physics labs, Phys. Rev. X **10**, 011029 (2020).

[37] N. G. Holmes, Carl E. Wieman, and D. A. Bonn, Teaching critical thinking, Proc. Natl. Acad. Sci. U.S.A. **112**, 11199 (2015).

[38] Gina M. Quan, Chandra Turpen, and Andrew Elby, Analyzing identity trajectories within the physics community, Phys. Rev. Phys. Educ. Res. **18**, 020125 (2022).

[39] Dina Zohrabi Alaee, Micah K. Campbell, and Benjamin M. Zwickl, Impact of virtual research experience for under-

graduates experiences on students' psychosocial gains during the COVID-19 pandemic, Phys. Rev. Phys. Educ. Res. **18,** 010101 (2022).

[40] Sofia Dokuka, Diliara Valeeva, and Maria Yudkevich, How academic achievement spreads: The role of distinct social networks in academic performance diffusion, PLoS One **15,** e0236737 (2020).

[41] Meagan Sundstrom, Andy Schang, Ashley B. Heim, and N. G. Holmes, Understanding interaction network formation across instructional contexts in remote physics courses, Phys. Rev. Phys. Educ. Res. **18,** 020141 (2022).

[42] Karen L. Tonso, Student engineers and engineer identity: Campus engineer identities as figured world, Cultural Studies Sci. Educ. **1,** 273 (2006).

[43] Deborah Fields and Noel Enyedy, Picking up the mantle of "expert": Assigned roles assertion of identity, and peer recognition within a programming class, Mind Cult. Act. **20,** 113 (2013).

[44] Karin Due, Who is the competent physics student? A study of students' positions and social interaction in small-group discussions, Cultural Studies Sci. Educ. **9,** 441 (2014).

[45] Allison J. Gonsalves, Physics and the girly girl—there is a contradiction somewhere: Doctoral students' positioning around discourses of gender and competence in physics, Cultural Studies Sci. Educ. **9,** 503 (2014).

[46] Allison Gonsalves, Persistent discourses in physics education: Gender neutrality and the gendering of competence, Cultural Studies Sci. Educ., **9,** 461 (2014).

[47] Paul W. Irving and Eleanor C. Sayre, Becoming a physicist: The roles of research, mindsets, and milestones in upper-division student perceptions, Phys. Rev. ST Phys. Educ. Res. **11,** 020120 (2015).

[48] Danny Doucette, Russell Clark, and Chandralekha Singh, Hermione and the secretary: How gendered task division in introductory physics labs can disrupt equitable learning, Eur. J. Phys. **41,** 035702 (2020).

[49] Danny Doucette, Russell Clark, and Chandralekha Singh, What makes a good physics lab partner?, presented at PER Conf. 2020, virtual conference, 10.1119/perc .2020.pr.Doucette.

[50] Katelyn M. Cooper, Anna Krieg, and Sara E. Brownell, Who perceives they are smarter? Exploring the influence of student characteristics on student academic self-concept in physiology, Adv. Physiol. Educ. **42,** 200 (2018).

[51] Emily M. Stump and Natasha G. Holmes, Student views of what counts as doing physics in the lab, presented at PER Conf. 2022, Grand Rapids, MI, 10.1119/perc.2022.pr .Stump.

[52] Rebeckah Fussell, Megan Renz, Philip Krasicky, Robert Fulbright, Jr., and Natasha Holmes, Introductory data analysis with Jupyter Notebooks, PICUP Collection (2022), https://doi.org/10.1119/PICUP.Exercise .DataAnalysis.

[53] Justyna P. Zwolak, Michael Zwolak, and Eric Brewe, Educational commitment and social networking: The power of informal networks, Phys. Rev. Phys. Educ. Res. **14,** 010131 (2018).

[54] Adrienne L. Traxler, Tyme Suda, Eric Brewe, and Kelley Commeford, Network positions in active learning environments in physics, Phys. Rev. Phys. Educ. Res. **16,** 020129 (2020).

[55] Remy Dou and Justyna P. Zwolak, Practitioner's guide to social network analysis: Examining physics anxiety in an active-learning setting, Phys. Rev. Phys. Educ. Res. **15,** 020105 (2019).

[56] Kelley Commeford, Eric Brewe, and Adrienne Traxler, Characterizing active learning environments in physics using network analysis and classroom observations, Phys. Rev. Phys. Educ. Res. **17,** 020136 (2021).

[57] Jeffrey A. Smith and James Moody, Structural effects of network sampling coverage i: Nodes missing at random, Soc. Netw. **35,** 652 (2013).

[58] Devyn Shafer, Maggie S. Mahmood, and Tim Stelzer, Impact of broad categorization on statistical results: How underrepresented minority designation can mask the struggles of both Asian American and African American students, Phys. Rev. Phys. Educ. Res. **17,** 010113 (2021).

[59] Daniel Z. Grunspan, Benjamin L. Wiggins, and Steven M. Goodreau, Understanding classrooms through social network analysis: A primer for social network analysis in education research., CBE Life Sci. Educ. **13,** 167 (2014).

[60] Eric Brewe, The roles of engagement: Network analysis in physics education research, in *Getting Started in PER* (American Association of Physics Teachers, College Park, MD, 2018), Vol. 2.

[61] Tom A. B. Snijders and Stephen P. Borgatti, Non-parametric standard errors and tests for network statistics, Connections **22,** 61 (1999), https://www.stats.ox.ac.uk/ ~snijders/Snijders_Borgatti.pdf.

[62] Yuzhou Chen, Yulia R. Gel, Vyacheslav Lyubchich, and Kusha Nezafati, Snowboot: Bootstrap methods for network inference, arXiv:1902.09029.

[63] Carolyn J. Anderson, Stanley Wasserman, and Bradley Crouch, A p* primer: Logit models for social networks, Soc. Netw. **21,** 37 (1999).

[64] Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher, An introduction to exponential random graph (p*) models for social networks, Soc. Netw. **29,** 173 (2007).

[65] David R. Hunter, Mark S. Handcock, Carter T. Butts, Steven M. Goodreau, and Martina Morris, ergm: A package to fit, simulate and diagnose exponential-family models for networks, J. Stat. Softw. **24,** 54860 (2008).

[66] N. Pearson, Allison Godwin, and Adam Kirn, Developing an instrument to understand the social-structural integration of diverse students, in *Proceedings of the American Society for Engineering Education Annual Conference and Exposition* (American Society for Engineering Education, Columbus, OH, 2017), https://peer.asee.org/28140.

[67] Cole Walsh, Martin M. Stein, Ryan Tapping, Emily M. Smith, and N. G. Holmes, Exploring the effects of omitted variable bias in physics education research, Phys. Rev. Phys. Educ. Res. **17,** 010119 (2021).

[68] Eric D. Kolaczyk and Pavel N. Krivitsky, On the question of effective sample size in network modeling: An asymptotic inquiry, Stat. Sci. **30,** 184 (2015).

[69] Fan Yin and Carter T. Butts, Highly scalable maximum likelihood and conjugate Bayesian inference for ERGMs on graph sets with equivalent vertices, arXiv:2110.13527.

[70] Miller McPherson, Lynn Smith-Lovin, and James M. Cook, Birds of a feather: Homophily in social networks, Annu. Rev. Sociol. **27**, 415 (2001).

[71] Jacob Cohen, The earth is round ($p < .05$), Am. Psychol. **49**, 997 (1994).

[72] Geoff Cumming, *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-analysis* (Routledge, London, 2013).

[73] Brian A. Nosek, Charles R. Ebersole, Alexander C. DeHaven, and David T. Mellor, The preregistration revolution, Proc. Natl. Acad. Sci. U.S.A. **115**, 2600 (2018).

[74] Michael Correll and Michael Gleicher, Error bars considered harmful: Exploring alternate encodings for mean and error, IEEE Trans. Vis. Comput. Graph., **20**, 2142 (2014).

[75] R. Tesch, *Qualitative research—analysis types and software protocols* (The Falmer Press, Hampshire, UK, 1990).

[76] John L. Campbell, Charles Quincy, Jordan Osserman, and Ove K. Pedersen, Coding in-depth semistructured interviews: Problems of unitization and intercoder reliability and agreement, Sociol. Methods Res. **42**, 294 (2013).

[77] Wendy K. Adams and Carl E. Wieman, Analyzing the many skills involved in solving complex physics problems, Am. J. Phys. **83**, 459 (2015).

[78] Andrei P. Kirilenko and Svetlana Stepchenkova, Intercoder agreement in one-to-many classification: Fuzzy kappa, PLoS One **11**, e0149787 (2016).

[79] Scott W. Duxbury, The problem of scaling in exponential random graph models, Sociol. Methods Res. **52**, 764 (2021).

[80] Katherine N. Quinn, Michelle M. Kelley, Kathryn L. McGill, Emily M. Smith, Zachary Whipps, and N. G. Holmes, Group roles in unstructured labs show inequitable gender divide, Phys. Rev. Phys. Educ. Res. **16**, 010129 (2020).

[81] Jesper Bruun and Ian G. Bearden, Time development in the early history of social networks: Link stabilization group dynamics, and segregation, PLoS One **9**, e112775 (2014).

[82] Jean Yonemura Wing, Beyond black and white: The model minority myth and the invisibility of Asian American students, Urban Rev. **39**, 455 (2007).

[83] Nolan L. Cabrera, Beyond black and white: How white, male, college students see their Asian American peers, Equity Excell. Educ. **47**, 133 (2014).