

Introductory physics students' recognition of strong peers: Gender and racial or ethnic bias differ by course level and context

Meagan Sundstrom^{1,†}, Ashley B. Heim², Barum Park³, 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*

³*Department of Sociology, Cornell University, Ithaca, New York 14853, USA*



(Received 8 July 2022; accepted 28 November 2022; published 23 December 2022)

Researchers have pinpointed recognition from others as one of the most important dimensions of students' science and engineering identity. Studies, however, have found gender biases in students' recognition of their peers, with inconsistent patterns across introductory science and engineering courses. Toward finding the source of this variation, we examine whether a gender bias exists in students' nominations of strong peers across three different remote, introductory physics courses with varying student populations (varying demographics, majors, and course levels). We also uniquely evaluate possible racial or ethnic biases and probe the relationship between instructional context (whether lecture or laboratory) and recognition. Some of our results replicate previous findings (such as the association of course grade and small class section enrollment with nominations), while others offer contradictions. Comparing across our three courses and the prior work, results suggest that course level (whether first-year students or beyond-first-year students) might be more associated with a gender bias in peer recognition than other variables. Surprisingly, we also find instances of racial or ethnic biases in favor of students from backgrounds historically underrepresented in science. Finally, we find that the nomination patterns differ when students nominate individuals strong in the lecture material versus the laboratory material. This work serves as an important step in determining which courses and contexts exhibit biases in peer recognition, as well as how students' perceptions of one another form in remote teaching environments.

DOI: [10.1103/PhysRevPhysEducRes.18.020148](https://doi.org/10.1103/PhysRevPhysEducRes.18.020148)

I. INTRODUCTION

Introductory science, technology, engineering, and math (STEM) courses are particularly critical transition periods for students. These courses impact students' retention and persistence in their subsequent courses, their majors, and their future careers [1,2]. Retention rates, however, are lowest for students from marginalized groups, including women and underrepresented racial minorities [(URM); taken as any racial or ethnic group other than White and Asian/Asian American from here on] [3–11]. Research suggests that these retention issues are in part due to societal stereotypes that position women and URM students as less suitable to scientific fields than men and non-URM individuals, respectively [12–22]. This positioning leads underrepresented students to

experience low sense of belonging and low self-efficacy in their science and engineering courses [19,23–29]. These stereotypes manifest in many ways, including students' perceptions of their peers' abilities related to the course material [30–32].

Prior work has found conflicting results with regard to whether there is a gender bias in introductory STEM students' perceptions of their peers [30–32]. These studies observed that students disproportionately nominated men over women as strong in their biology and physics courses [30,32], but that men and women received comparable numbers of nominations in mechanical engineering courses [31]. Whether these discrepant results are due to varying student populations (e.g., students' majors, academic years, and demographics) in the observed courses, the scientific discipline of the course, or some other factor is still unresolved. We advance this body of work by collecting and analyzing students' nominations of strong peers in three different remote physics courses serving various student populations. We aim to determine how the student population of and context within a course (lecture or laboratory) are related to students' perceptions of their peers. We also expand previous work by considering possible racial or ethnic biases, in addition to gender biases, in students' nominations.

[†]mas899@cornell.edu

^{*}Corresponding author.
ngholmes@cornell.edu

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

A. Recognition in STEM

We situate our study in the theoretical framework of identity. An individual's identity refers to their being a "certain kind of person" in a given context [33]. *Science identity*, therefore, is the degree to which an individual believes they are a "science person." Researchers have conceptualized a model of science and engineering identity containing four dimensions: performance, competence, interest, and recognition [18,34]. Studies show that *recognition* is one of the most important of these dimensions in predicting students' participation, persistence, and career intentions in science and engineering [18,34–41]. Recognition is the degree to which meaningful others (e.g., peers, teachers, and family) perceive and acknowledge an individual as a science person. When a student receives ample recognition from others, they are likely to see themselves as a science person and develop a strong science identity [37,42].

Given the importance of recognition, it stands a reasonable desire for all students to feel recognized by their peers in a science classroom. Recognition, however, is 'culturally produced'—it is influenced by sociohistorical norms and stereotypes [18,39,43]. For physics in particular, stereotypes often position men and non-URM students as more suitable to the field than women and URM students [12–22,28,29]. In turn, empirical work shows that men and non-URM students report higher senses of recognition in their physics classes than women and URM students, respectively [36,39,44]. Close examinations of women and women of color in physics [18,43,45] also affirm that high-achieving or "exceptional" women in the field hinge on recognition from others to succeed. Thus, students' gender *and* race or ethnicity are closely related to their recognition and identity in physics. Previous work, however, has not analyzed racial or ethnic bias in students' nominations of strong peers (they analyze gender bias only) [30–32]. To contribute to this gap, we measure both gender and racial or ethnic bias in the current study.

Prior research also suggests that recognition may vary within different disciplines, instructional formats, and student populations. Grunspan and colleagues [30] examined three iterations of a large, introductory biology course (the second in the course sequence) with gender-balanced enrollment. They found in all semesters that men received significantly more nominations as strong in the course material than women. The extent to which this bias occurred, however, varied between instruction types: they observed that women nominated other women more frequently when the instructor employed "random call" during class. Salehi and colleagues [31] later performed a similar analysis of two offerings (one with traditional instruction and one with active learning instruction) of a medium-sized, gender-balanced mechanical engineering course taken by second and third-year students planning to major in engineering. They expected that the nature of

the engineering discipline would lead to more gender bias in peer recognition than that found in biology. Instead, they found no gender bias in their observed nominations in either course offering. Lastly, Bloodhart and colleagues [32] analyzed peer perceptions across many introductory life science and physics courses. They do not report the student populations of each course, but, in aggregate, students in the life science courses were mostly non-URM women in their first year of study and the physical science courses were mostly non-URM men across all four academic years. The researchers found in the two disciplines that both men and women under-nominated women as knowledgeable in the course material in comparison to women's actual final grades in the class. To examine whether the different results across these three studies are attributable to varying disciplines and/or student populations, we analyze three physics courses serving different populations of students.

The studies above also suggest that instruction type, such as a traditionally taught lecture versus an active learning course with frequent group work, may influence recognition. In physics, furthermore, recognition likely varies between the instructional contexts of lecture and laboratory (lab) work. Not only do these two contexts involve very different pedagogies (lectures contain many students who focus on the instructor and labs contain a small number of students who collaborate on tasks), but they also cover distinct content and aim to develop different sets of skills [46–51]. Correspondingly, research has shown that students relate knowledge of mathematics or theoretical physics with lectures, while they view "doing lab" as handling machinery and using technical skills [12,13]. Such differences in relevant skill sets suggest that different students may be recognized as strong in lecture and lab contexts. Thus, we probe and analyze students' recognition of peers in the two contexts separately in the current study.

B. Current study

We collected students' nominations of peers they believed were strong in the lecture and lab material in three different remote, introductory physics courses at Cornell University. These data allowed us to compare peer recognition across instructional contexts (within our study) as well as across disciplines (comparing our study of physics courses to prior research on courses in other disciplines). Each of the three physics courses also contained varying student populations in terms of students' major, academic year, gender, and race or ethnicity. This variation allowed us to examine whether gender and racial or ethnic biases in peer recognition depend on features of student population as suggested by prior studies [30–32]. The following two research questions guided our study:

- (1) To what extent do gender and racial or ethnic biases exist in students' recognition of strong peers across

three different introductory physics courses serving distinct student populations?

- (2) How does introductory physics students' recognition of strong peers differ in lecture and lab contexts?

Comparing the three courses in our study to previous studies [30–32], we find that whether students' perceptions of strong peers in lecture exhibit a gender bias might depend on course level over other variables (e.g., student populations or scientific discipline of the course). Courses serving first-year students exhibit a gender bias in lecture perceptions, while those serving the beyond-first-year level do not. Surprisingly, we also find in some cases that URM students tend to receive more nominations than their non-URM peers. With regard to instructional context, we observe that both the general patterns of nominations and whether a gender or racial or ethnic bias exists in nominations differ between the lecture and lab contexts.

II. METHODS

In this section, we first characterize the courses and students we analyzed. Then, we describe our data collection and analysis methods.

A. Courses and participants

Our data come from three introductory physics courses at Cornell University, which we call Courses A, B, and C. These courses were held during the Fall 2020 semester, when about three-quarters of students at the institution resided on campus but most courses were held online due to the COVID-19 pandemic. Course A was an introductory, calculus-based mechanics course aimed at first-year students intending to major in engineering or other STEM disciplines. Course B was also an introductory, calculus-based mechanics course, but primarily served first-year students intending to major in physics. Course C was an introductory, calculus-based electromagnetism course intended for first- and second-year students majoring in engineering or other STEM disciplines. Course C is typically taken after Course A in a course sequence, so it is reasonable to assume that students were familiar with one another to some extent before entering the course. Students in our data set for Courses A and C are different, however, because we examine the two courses during one semester. Only data for Course C provide a snapshot of this familiarity as many of those students took mechanics during the previous semester, but that offering of Course A is not analyzed here.

As summarized in Table I, each course had lecture, lab, and discussion sections. Most course components were held online (synchronously), with a few held in person. Lectures for all three courses were instructed by a male faculty member in the physics department. Courses A and C were “flipped,” such that students read relevant sections

of the textbook and took a reading quiz before coming to class. Lectures for Course A used conceptual iClicker questions and instructor demonstrations, while lecture time in Course C was spent on problem-solving questions through Learning Catalytics [52]. In both courses, students answered questions both individually and in groups. Course B was more traditional in that, during lectures, the instructor presented new content and asked iClicker questions that students answered individually. Courses B and C (but not Course A) used an online discussion forum where both students and teaching staff could post questions and answers related to course content at any time.

In all three courses, lab and discussion sections were led by graduate teaching assistants. Lab sections met once per week for 2 h and discussion sections met twice per week for 50 min. Each section contained approximately 20 students who worked together in small groups of two to four. The labs and most of the discussion sections took place online through Zoom and students worked in groups in virtual breakout rooms. In the few discussion sections held in person, students worked together at round tables. Labs were nontraditional [46–51], where students performed open-ended investigations using objects at home or in their dorm rooms. They submitted their lab work as a group, rather than individually. During discussion sections, students solved problems related to the lecture content. Despite working as a group, students submitted their discussion work individually in Course C. Students were not required to submit discussion work in Courses A and B.

We collected students' self-reported gender, race or ethnicity, intended major, and academic year via a survey at the beginning of the semester. We grouped race or ethnicity by URM status, where non-URM students are those solely identifying as White and/or Asian/Asian American and URM students are those identifying as at least one of any other race or ethnicity (including Black or African American, Hispanic/Latinx, and Native Hawaiian or other Pacific Islander). These student populations and average final course grades are summarized in Table I. Similar proportions of men (42%) and women (47%) were enrolled in Course A and roughly three-quarters of the students in this course identified as non-URM (71%). Men and women received comparable final course grades on average in this course, as did URM and non-URM students. Most students in Course B identified as men (70%) and more than three-quarters of the students identified as non-URM (81%). Men and non-URM students on average received higher final course grades in this course than women and URM students, respectively. Course C contained similar proportions of men (45%) and women (51%) and two-thirds of the students identified as non-URM (66%). Men and women on average received comparable final course grades in this course, while non-URM students on average received higher final course grades than URM students.

TABLE I. Summary of the modality, student demographics, and student final course grades for the three courses we analyzed. All online meetings were synchronous. Percentages are relative to the number of students included in the analysis. We denote students' gender or race or ethnicity as "unknown" if they preferred not to disclose this information on the survey or if they did not complete the survey (but did consent to other parts of the research). Grades are provided on a 4.0 scale, with standard deviations (SD) of the means given in parentheses.

	Course A	Course B	Course C
Modality			
Lecture	Online	Online	Online
Lab sections	Online	Online	Online
Discussion sections	12 Online, 2 In-person	3 Online, 2 In-person	Online
Total enrollment	208	89	433
Students in analysis	198	84	404
Gender			
Men	84 (42%)	59 (70%)	182 (45%)
Women	92 (47%)	23 (28%)	206 (51%)
Nonbinary	0	1 (1%)	3 (1%)
Unknown	22 (11%)	1 (1%)	13 (3%)
Race or ethnicity			
Non-URM	140 (71%)	68 (81%)	268 (66%)
URM	32 (16%)	12 (14%)	107 (27%)
Unknown	26 (13%)	4 (5%)	29 (7%)
Major			
Physics or Engineering Physics	11 (5%)	58 (69%)	14 (3%)
Engineering	128 (65%)	14 (17%)	331 (83%)
Other (STEM)	19 (10%)	7 (8%)	24 (6%)
Unknown	40 (20%)	5 (6%)	35 (8%)
Year			
First-year	166 (83%)	78 (93%)	10 (2%)
Second-year	24 (12%)	3 (4%)	374 (93%)
Third-year	6 (3%)	1 (1%)	12 (3%)
Other or unknown	2 (2%)	2 (2%)	8 (2%)
Grades			
Mean (SD) final course grade (men)	3.6 (0.6)	3.5 (0.6)	3.4 (0.5)
Mean (SD) final course grade (women)	3.6 (0.4)	3.2 (0.6)	3.4 (0.4)
Mean (SD) final course grade (non-URM)	3.5 (0.6)	3.5 (0.6)	3.4 (0.4)
Mean (SD) final course grade (URM)	3.6 (0.6)	3.2 (0.5)	3.2 (0.5)

B. Data collection

In all three courses, we administered an online survey during the eighth week of the 15-week semester as part of a lab homework assignment about students' group work experiences. On the survey, we asked students to nominate peers whom they believed were knowledgeable in the course with the following two prompts adapted from prior work [30–32]:

Please list any students in this physics class that you think are particularly strong in the lecture or discussion section material.

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

We refer to the first prompt as "lecture perceptions" and the second prompt as "lab perceptions."

The survey was in an open response format (one text box) and students could respond with an unlimited number

of names. This format avoids students feeling obligated to fill a quota and writing down extra names of peers they may not actually perceive as strong [53]. Students were also not given a class roster from which to choose or look up names. This resulted in some listings being hard to match to the class roster during analysis, as there were instances of students misspelling peers' names and reporting just a first or a last name. Thus, text processing of the responses was necessary. We compared each name reported on the survey to the class roster and matched up names for which the number of corrections needed to match the full name on the roster was fewer than 0.3 times the length of someone's full name. We chose the constant 0.3 via trial and error, finding that this worked best for capturing as many close matches as possible without producing false negatives. If a name (either first or last) appeared multiple times in the dataset, then we did not match on listings of just that name itself and only matched listings of the other half of the name or the full name.

TABLE II. Survey response rates and mean number of nominations made per student. The survey response rate is the percent of enrolled students who completed the survey. The mean number of nominations is the average number of peers' names that each student listed.

	Course A	Course B	Course C
Survey response rate	79%	92%	83%
Mean nominations (lecture)	1.2	1.9	1.0
Mean nominations (lab)	1.2	1.3	0.9

As summarized in Table II, survey response rates were reasonably high (all above 75%) and students in each course on average listed one or two peers for each prompt. Our analysis included all students who responded to the survey and/or were listed by at least one peer. We also only included the nominations made by students who consented to participate in research (more than 95% of survey responders). If a consenting student nominated a non-consenting student, we included the nomination, but removed all information (demographics, etc.) about the nonconsenting student. In all courses, at least 93% of the enrolled students are included in our analysis (see Table I).

We note that prior studies [30,31] used surveys late in the semester, which formed highly centralized networks (many nominations were concentrated to a few students) for course-level perceptions. In our study, two of the three courses exhibited highly centralized lecture perception networks at the mid-semester mark, so there is no reason to believe our results are impacted by the timing of survey administration.

At the end of the semester, we collected discussion and lab section enrollment for all courses. For Courses B and C, we also collected the number of student contributions to the course's online discussion forum (sum of their posted questions and posted answers to others' questions). Course A did not use a discussion forum. We used these discussion forum contributions as a measure of students' outspokenness because it quantifies students' communicative engagement during an online course. This is a similar, but distinct, measure to that of Grunspan and colleagues [30], who determined outspokenness by asking the instructor to name actively participating students after each class meeting. Some students had no discussion forum data, indicating that they likely did not ever register for or use the forum. For these students, we imputed their contributions to the discussion forum as zero, which was also the mode of each course's distribution of contributions. We imputed one student's and 55 students' discussion forum data in Courses B and C, respectively.

C. Analysis of nominations

We converted the survey responses into directed networks for each course (A, B, and C) and each context

(lecture and lab). *Nodes* represented students and *edges* (or *ties*) represented all nominations made between students. To first gain a sense of each network's overall structure, we calculated two network-level statistics: *density* and *indegree centralization*. Density measures the proportion of all possible edges in the network that we observed. Indegree centralization measures the extent to which the nominations are concentrated around a single student or a small subset of students (i.e., whether there are emergent *celebrities* who receive most of the nominations). Higher indegree centralization indicates higher concentration around one or a few students. 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 [54,55]. The bootstrapping was performed with 10 000 bootstrap trials for each network using the *snowboot* package in R [56].

We then used exponential random graph models (ERGMs) to statistically determine the salient structural characteristics of our networks. ERGMs assume that the observed network is a realization from a random graph that comes from a distribution belonging to the exponential family [57,58]. They allow us to perform many statistical tests at once, determining whether the frequency of certain patterns or configurations in our observed network is significantly different than if the ties were formed randomly. To formulate such a model, we first choose a principled set of predictor variables (i.e., configurations) that might explain the formation of the observed network. These variables may be structural (e.g., measuring the tendency for mutual nominations) or nodal (e.g., measuring the extent to which students of a certain gender are more likely to receive a nomination). The goal is to use these k network statistics $g_k(y)$ and their corresponding coefficients θ_k to predict the formation of the random network Y . The model takes the form

$$P_{\theta}[Y = y] = \frac{1}{\psi} \exp \left(\sum_k \theta_k g_k(y) \right),$$

where y is a realization of the random network Y and $\psi = \sum_y \exp [\sum_k \theta_k g_k(y)]$ is 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 [59], which we used to fit all models in our analysis.

There are two different ways to interpret the coefficients of ERGMs. In general, the coefficients weight the importance of each modeled configuration for the formation of the realized network, where positive (negative) coefficients indicate that the configuration is observed more (less)

frequently than by chance after accounting for all other configurations that are modeled. The second way to interpret the coefficients is to focus on specific ties of the network. In this “change statistics” interpretation, the coefficient θ_k of the k th configuration shows how much the log-odds of a tie being present changes if the formation of the tie increases the k th configuration by one unit, holding the rest of the network constant. For instance, if the predictor variable measures the number of mutual ties in the network, its coefficient represents how much the log-odds of a tie being present increases when the addition of this tie would reciprocate an existing tie.

We initially fit ERGMs with the same set of predictor variables used by Grunspan and colleagues [30] for each of our observed networks. Our model contained one additional variable for discussion section homophily (the tendency for students to nominate peers enrolled in their same discussion section) because discussion was an extra structural component in our courses. We also added three variables to measure the effects of race or ethnicity, which exactly mirrored the structure of the gender variables in the original model of Grunspan and colleagues [30]. Inspection of the goodness-of-fit diagnostics, however, revealed a significant inadequacy in this model: we were not appropriately capturing the presence of *triadic closure*. Triadic closure is the tendency for three nodes to be connected, given pairwise connections. That is, if ties exist between nodes A and B and between nodes B and C, then a tie between nodes A and C forms triadic closure. In some cases, we were also not adequately capturing the network’s *outdegree* distribution (the proportion of nodes *making* a certain number of nominations). Our model did sufficiently account for each network’s *indegree* distribution (the proportion of nodes *receiving* a certain number of nominations).

In response, we altered the original model from Grunspan and colleagues [30]. We added a geometrically weighted edgewise shared partner (GWESP) variable, which is typically used to account for triadic closure. The more ties two nodes have in common (i.e., the more shared partners they have), however, the higher the probability of an edge forming between them. Thus, a decay parameter for the GWESP variable determines the extent to which the probability of tie formation decreases for each additional partner already shared between two nodes [60]. This parameter can take on a value between 0 and 1, with lower values creating larger decreases in tie probability per subsequent shared partner. We used a fixed decay parameter of 0.25 as is commonly used in ERGM literature [61–63]. Because incorporating both a GWESP term and an isolates term (for students receiving zero nominations; used in the original model) produced degeneracy in the model, we removed the isolates term. We also changed the structure of the gender and race or ethnicity variables to allow for easier and more meaningful

interpretations. Specifically, we added a variable to the model for each possible directed tie for the gender and race or ethnicity attributes (e.g., man nominating a man, man nominating a woman, etc.) as in Ref. [64]. These network statistics allowed for a more direct comparison of ties by using common base terms for gender and race or ethnicity variables and thus an easier interpretation of gender and racial or ethnic biases. We note that in creating these variables, we fit models with each possible base term for gender and race or ethnicity to the observed networks. We ultimately chose to use nominations between majority demographic groups as the base terms, however the results are consistent with those of all the possible models.

These modifications resulted in an improved model fit for every observed network. For all six observed networks, the goodness-of-fit diagnostics showed that we were capturing the distributions of indegree, outdegree, and triadic closure well with the revised model. Coefficient estimates using the original model of Ref. [30] and an example of goodness-of-fit diagnostics for both models are provided in the Appendix. We report in the main text our results using the revised model, which contained the following predictor variables:

- (1) *Edges*: intercept term equal to the number of edges in the network
- (2) *Mutuality*: number of reciprocated or mutual nominations
- (3) *Geometrically weighted edgewise shared partners (GWESP)*: triadic closure
- (4) *Woman* $\rightarrow$ *woman*: number of edges for which a woman nominates another woman (base term is *man* $\rightarrow$ *man*)
- (5) *Woman* $\rightarrow$ *man*: number of edges for which a woman nominates a man (base term is *man* $\rightarrow$ *man*)
- (6) *Man* $\rightarrow$ *woman*: number of edges for which a man nominates a woman (base term is *man* $\rightarrow$ *man*)
- (7) *URM* $\rightarrow$ *URM*: number of edges for which a URM student nominates a URM student (base term is *non-URM* $\rightarrow$ *non-URM*)
- (8) *URM* $\rightarrow$ *non-URM*: number of edges for which a URM student nominates a non-URM student (base term is *non-URM* $\rightarrow$ *non-URM*)
- (9) *Non-URM* $\rightarrow$ *URM*: number of edges for which a non-URM student nominates a URM student (base term is *non-URM* $\rightarrow$ *non-URM*)
- (10) *Final course grade of nominee*: correlation between final course grade and indegree
- (11) *Discussion forum contributions of nominee* (only for Courses B and C): correlation between discussion forum contributions and indegree
- (12) *Homophily on lab section*: number of edges connecting students enrolled in the same lab section
- (13) *Homophily on discussion section*: number of edges connecting students enrolled in the same discussion section

We determined the coefficient estimates of these variables for each of our six observed networks using MCMC MLE and then compared the results across courses and contexts.

III. RESULTS

We first compare the structures and network-level statistics of each observed network. Then, we present the results of the exponential random graph models.

A. Structure of peer perception networks

Figures 1 and 2 show the network diagrams of lecture and lab perception networks, respectively, for each course. In each diagram, edges point from the nominator to the nominee and larger nodes represent students who received more nominations (i.e., higher indegree). Nodes are colored by gender and nodes with bold outlines indicate celebrities. These same network diagrams with nodes colored by race or ethnicity can be found in the Supplemental Material [65].

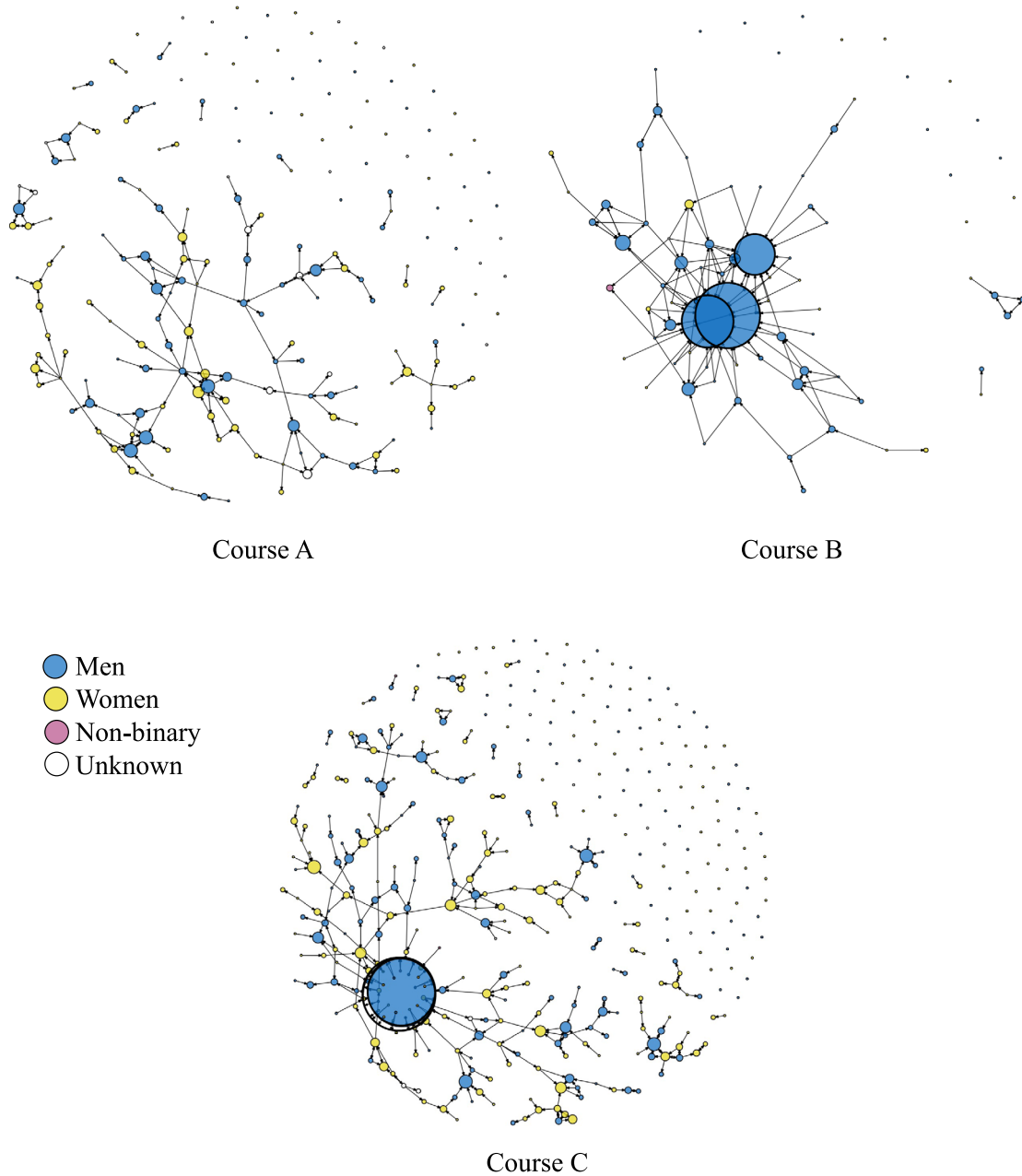


FIG. 1. Lecture perception networks. Nodes are colored by self-reported gender and sized proportional to indegree (number of received nominations). Nodes with bold outlines indicate celebrities (three in Course B and two in Course C). Edges point from the nominator to the nominee.

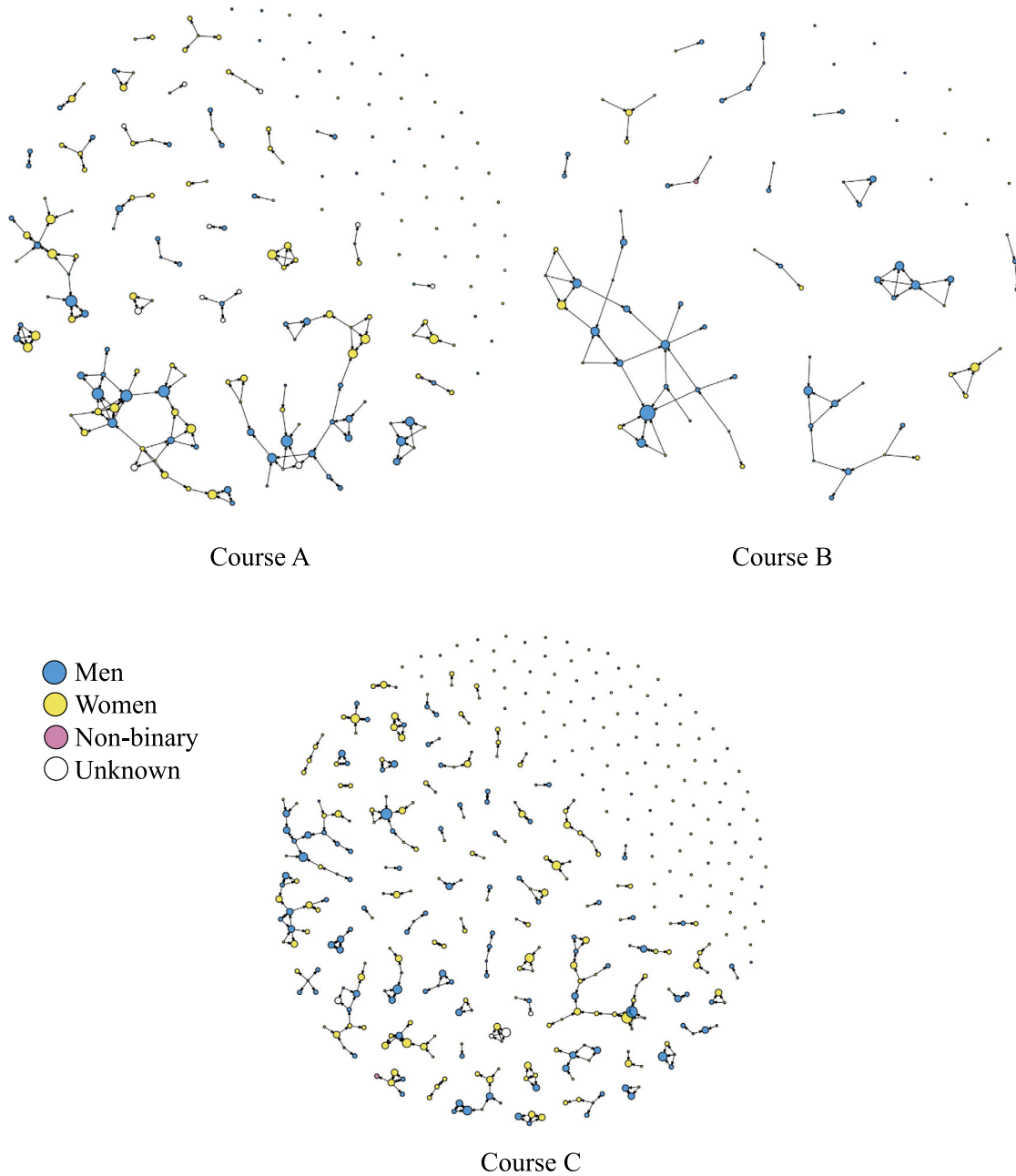


FIG. 2. Lab perception networks. Nodes are colored by self-reported gender and sized proportional to indegree (number of received nominations). Edges point from the nominator to the nominee.

Table III summarizes the density and indegree centralization of each observed network. Within each course, the network densities of the lecture and lab perception networks are similar to one another. We see this as similar levels of connectedness (proportion of possible edges present) in Figs. 1 and 2 for lecture and lab perception networks within each course. This similarity in network density, however, is despite the very different structures of these connections across contexts. That is, students nominate similar numbers of peers in each context, but the

distribution of who receives the nominations is different between lecture and lab contexts.

In Course A, the indegree centralization (the extent to which the network is concentrated around just a few students) is similarly low for both lecture and lab perceptions. Correspondingly, we see in Figs. 1 and 2 that there are no emergent celebrities in either network for this course (no nodes are drastically larger than the others). In Courses B and C, however, the indegree centralization of the lecture perception network is larger than that of the lab perception

TABLE III. Network-level statistics for all observed networks. Density is the proportion of observed to possible edges. Indegree centralization is the extent to which the nominations are concentrated around one or a few students. Standard errors of the last digit are shown in parentheses.

	Course A		Course B		Course C	
	Lecture	Lab	Lecture	Lab	Lecture	Lab
Density	0.005(7)	0.005(6)	0.02(6)	0.01(2)	0.002(3)	0.002(2)
Indegree centralization	0.02(9)	0.02(8)	0.3(8)	0.06(2)	0.08(2)	0.008(3)

network by an order of magnitude. This suggests that the lecture perception networks of these two courses are much more concentrated around a few prominent students (celebrities) than the lab perception networks. We observe in Fig. 1 that the lecture perception networks of Courses B and C contain three and two celebrities (nodes with bold outlines that are much larger than the rest, having received many more nominations), respectively. On the other hand, we see in Fig. 2 that there are no central nodes receiving many nominations in either of the two lab perception networks (all nodes are similar in size). Thus, for Courses B and C, despite the similar density measures, the lecture perception networks are much more concentrated around a few celebrities, with no outstanding celebrities in the lab perception networks.

B. Evaluating gender and racial or ethnic bias in peer recognition

Table IV shows the coefficient estimates for our revised exponential random graph model fit to all observed networks. We interpret the coefficient estimates as the log-odds of tie formation. For example, the coefficient estimate for the *homophily on discussion section* variable for Course A's lecture perception network is 1.59. This means that the log-odds of a tie forming in the network increases by 1.59 for each additional tie connecting students in the same discussion section, holding the rest of the network the same. In other words, ties connecting students in the same discussion section are more probable than ties connecting students in different discussion sections, even after accounting for the other configurations included in the model.

For five out of six analyzed networks, the coefficient estimates for the *woman → woman* and *woman → man* variables, shown in light and medium green dots in Fig. 3, are not statistically significant. This means the frequency with which women nominate either a woman or man is not significantly different than the frequency with which men nominate other men (the base term) after adjusting for the other variables in the model. In other words, women proportionately nominate their female and male peers in these five networks. In the lecture perception network of Course C, however, women nominate other women significantly more than men nominate other men.

The coefficient estimates for the *man → woman* variable, shown in dark green dots in Fig. 3, indicate that men significantly undernominate women in the lecture perception network of Course A and both networks of Course B. The lecture and lab perception networks of Course B, moreover, have the largest and second-largest coefficient magnitudes for this variable, respectively. This suggests that the strongest gender bias occurs in Course B's lecture perception network. Making direct comparisons of ERGM coefficients across different networks, however, has limitations [66], so we consider this claim preliminary.

We note that the coefficient estimate for the *man → woman* variable is not statistically significant in the lecture perception network of Course C, however this might be due to one of the two celebrities having unknown gender. If we impute this celebrity's gender as a man, the *man → woman* variable becomes negative and statistically significant (implying a gender bias against women), though the other gender variables are not statistically significant. If we impute this celebrity's gender as a woman, the results related to gender are the same as when this celebrity's gender is unknown. The dependency of the statistical results on this one celebrity's gender offers an important caveat to our interpretations discussed in the next section.

The gender patterns suggested by our model fits are illustrated in Figs. 1 and 2. Despite there being no clear celebrities in the lecture perception network of Course A (the top-nominated man and woman received five and four nominations, respectively), our statistical model indicates a gender bias in this network. Therefore, men, on average, receive more nominations than women (the average size of the blue nodes is greater than the average size of the yellow nodes). For the lab perception network of Course A, men and women have an even distribution of nominations (the average size of the blue nodes is similar to the average size of the yellow nodes) as indicated by our statistical analysis. On the other hand, all three emergent celebrities in Course B's lecture perception network are men (the three largest nodes, outlined in bold, are blue). In this network, the top-nominated man received ten times as many nominations (30) as the top-nominated woman (three). Course B's lab perception network is less centralized around a few celebrities, but we see that, on average, men receive more nominations than women (the average size of the blue nodes is greater than the average size of

TABLE IV. Exponential random graph model results. We fit models with all possible permutations of the gender and race or ethnicity variables serving as the base terms, but here we show results using nominations between majority demographic groups as the base terms (*man* $\rightarrow$ *man* for gender and *non-URM* $\rightarrow$ *non-URM* for race or ethnicity). Standard errors of the coefficient estimates are in parentheses. Asterisks indicate statistical significance (* $p < 0.05$; ** $p < 0.01$).

	Course A		Course B		Course C	
	Lecture	Lab	Lecture	Lab	Lecture	Lab
<i>Edges</i>	-9.09 ** (0.72)	-7.85 ** (0.54)	-12.04 ** (1.11)	-7.88 ** (0.86)	-8.85 ** (0.31)	-8.67 ** (0.53)
<i>Mutuality</i>	2.92 ** (0.40)	1.82 ** (0.35)	0.96 * (0.45)	1.21 * (0.51)	3.30 ** (0.19)	1.90 ** (0.31)
<i>GWESP (triadic closure; decay parameter = 0.25)</i>	1.21 ** (0.17)	1.06 ** (0.14)	1.05 ** (0.15)	0.96 ** (0.19)	1.91 ** (0.07)	0.98 ** (0.11)
<i>Woman $\rightarrow$ woman</i>	0.06 (0.17)	0.08 (0.15)	-0.70 (0.73)	0.47 (0.28)	0.46 ** (0.09)	0.01 (0.18)
<i>Woman $\rightarrow$ man</i>	-0.03 (0.21)	-0.04 (0.21)	-0.13 (0.19)	-0.16 (0.28)	0.22 (0.14)	-0.16 (0.18)
<i>Man $\rightarrow$ woman</i>	-0.52 * (0.24)	-0.23 (0.24)	-1.45 ** (0.55)	-0.77 ** (0.37)	0.07 (0.14)	-0.31 (0.19)
<i>URM $\rightarrow$ URM</i>	-0.17 (0.55)	0.11 (0.43)	1.73 ** (0.43)	1.08 * (0.54)	0.23 (0.13)	-0.19 (0.29)
<i>URM $\rightarrow$ non-URM</i>	0.34 (0.24)	0.24 (0.28)	0.31 (0.22)	-0.04 (0.33)	-0.40 ** (0.14)	-0.24 (0.19)
<i>Non-URM $\rightarrow$ URM</i>	-0.02 (0.27)	-0.52 (0.34)	-0.07 (0.40)	-0.22 (0.38)	0.53 ** (0.11)	-0.16 (0.17)
<i>Final course grade of nominee</i>	0.86 ** (0.19)	0.29 * (0.14)	1.91 ** (0.28)	0.46 * (0.22)	0.54 ** (0.08)	0.29 * (0.15)
<i>Discussion forum contributions of nominee</i>	N/A	N/A	0.02 ** (0.003)	0.008 (0.006)	0.02 ** (0.005)	-0.003 (0.01)
<i>Homophily on lab section</i>	1.01 ** (0.17)	3.37 ** (0.21)	0.59 ** (0.17)	2.97 ** (0.32)	1.01 ** (0.12)	4.19 ** (0.16)
<i>Homophily on discussion section</i>	1.59 ** (0.17)	0.20 (0.25)	0.66 ** (0.18)	0.09 (0.24)	1.76 ** (0.09)	0.27 (0.20)

the yellow nodes). For Course C, the lecture perception network has one male celebrity and a second celebrity with unknown gender (these two nodes are outlined in bold and overlap in the diagram). As mentioned above, the latter might explain why we did not resolve any gender bias in our model fit. Finally, Course C's lab perception network is similar to that of Course A in that there is a relatively even distribution of nominations across men and women, in line with our quantitative findings above.

We examine the coefficient estimates for the race or ethnicity variables in a similar manner. These results are summarized in Fig. 4. In both networks of Course A and the lab perception network of Course C, none of the coefficient estimates for the race or ethnicity variables are statistically significant after adjusting for the remaining variables in the model. This means that no particular nomination type (e.g., URM student nominating URM student) occurs more frequently than another—both URM and non-URM

students proportionately nominate their URM and non-URM peers (no racial or ethnic bias). In both networks of Course B, however, URM students disproportionately over-nominate URM peers, even after controlling for the other network configurations in the model (yellow dots in Fig. 4). Similarly, in Course C's lecture perception network, URM students significantly undernominate non-URM peers (orange dots in Fig. 4) and non-URM students disproportionately overnominate URM peers (brown dots in Fig. 4). Accordingly, one of the two prominent celebrities in this particular network (shown in Fig. 1) is a URM student.

Similar to the results related to gender, we note that the results vary for the lecture perception network of Course C if we impute the race or ethnicity of the second celebrity whose race or ethnicity is unknown. If we impute this student as non-URM, we find that URM students disproportionately overnominate URM peers in this network, with no change to the other two race or ethnicity variables.

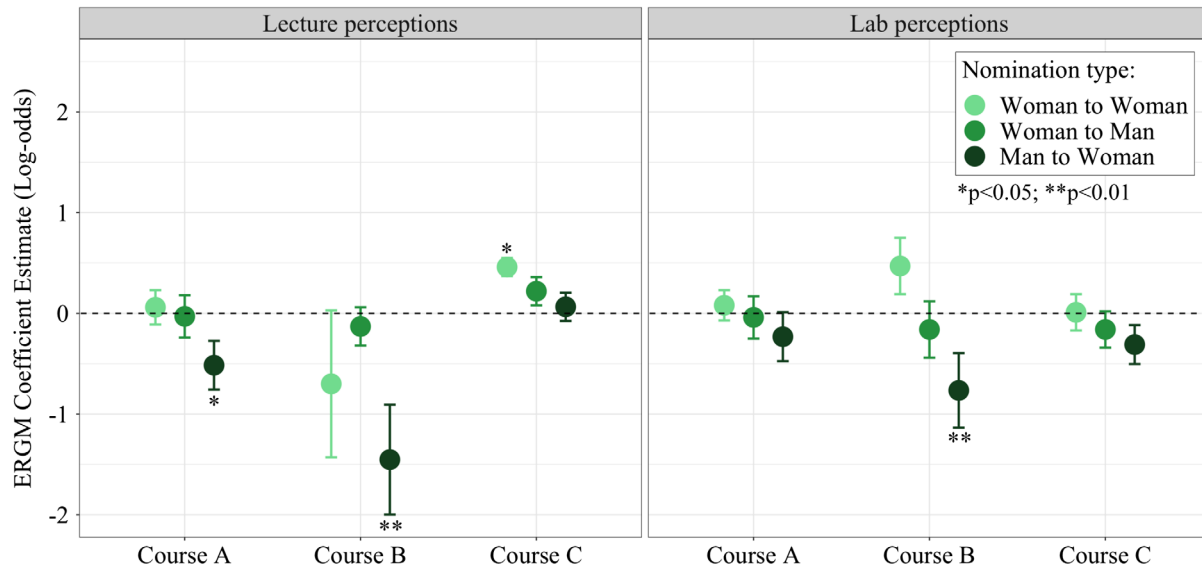


FIG. 3. ERGM coefficient estimates for the gender nomination variables. The base term (i.e., coefficient estimate of zero) is nominations from man to man. Error bars are the standard errors of the coefficient estimates (values shown in Table IV) and asterisks indicate statistical significance.

If we impute this student as URM, we find that both URM and non-URM students disproportionately overnominate URM peers. In both cases, therefore, we still find a tendency for URM students to receive more nominations than their non-URM peers.

C. Roles of final course grade, outspokenness, and section enrollment in shaping peer recognition

The remaining predictor variables in the model lend insight into the association between final course grade,

outspokenness on the online discussion forum, and section enrollment and the structure of our observed perception networks. All coefficient estimates for the *final course grade of nominee* variable, summarized in Table IV, are positive and statistically significant. That is, in all three courses and in both contexts, students with higher final course grades receive significantly more nominations than students with lower final course grades. The magnitudes of the coefficients also suggest that final course grade is a stronger predictor of receiving nominations in the lecture context than the lab context in every course, though again

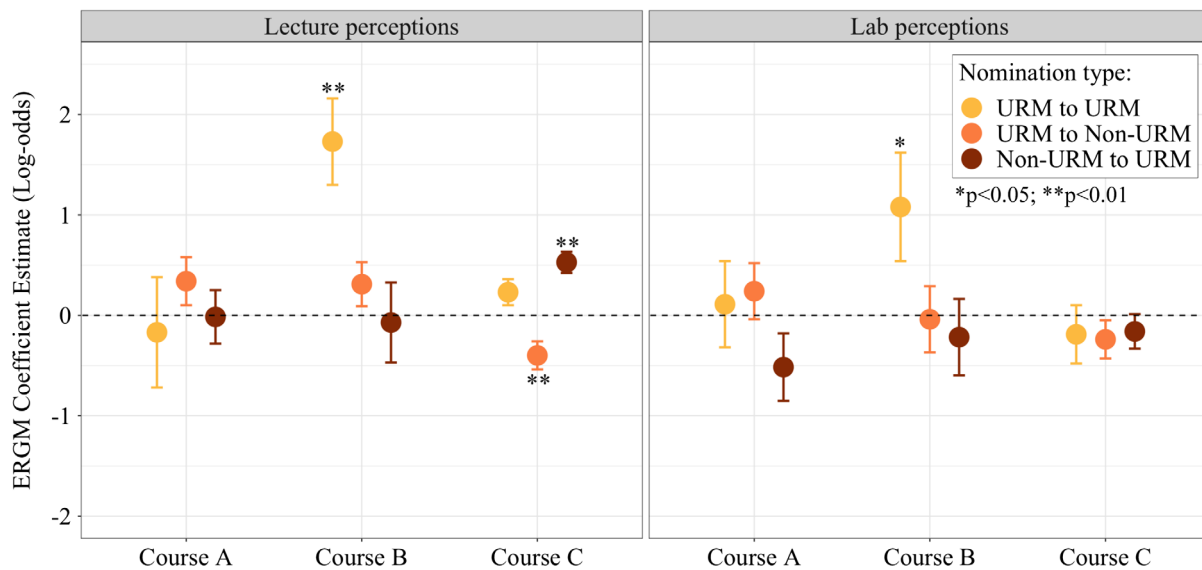


FIG. 4. ERGM coefficient estimates for the race or ethnicity nomination variables. The base term (i.e., coefficient estimate of zero) is nominations from non-URM to non-URM. Error bars are the standard errors of the coefficient estimates (values shown in Table IV) and asterisks indicate statistical significance.

such comparisons should be considered tentative [66]. We provide plots comparing the indegree distributions to final course grade in the Supplemental Material [65].

With regard to outspokenness (the *discussion forum contributions of nominee* variable, which was only measurable for Courses B and C), we find that students who contribute more to the discussion forum are significantly more likely to receive nominations as strong in the lecture context. Contributions to the discussion forum, however, are not a significant predictor of receiving nominations as strong in the lab context. We provide plots comparing the indegree distributions to number of discussion forum contributions in the Supplemental Material [65].

We observe similar patterns across courses regarding the relationship between lab and discussion section enrollment and recognition among peers. Coefficient estimates for the *homophily on lab section* variable are positive and statistically significant in every observed network, meaning that students are more likely to nominate peers in their lab section than peers outside of their lab section as strong in both lecture and lab content. Viewing the magnitude of the coefficients [66], this effect is, unsurprisingly, more pronounced in lab perception networks than lecture perception networks. On the other hand, coefficient estimates for the *homophily on discussion section* variable are positive and statistically significant in all three lecture perception networks, but they are not statistically significant in any of the three lab perception networks. This suggests that students tend to nominate peers in their discussion section as strong in the lecture material, but they do not systematically nominate discussion peers as strong in the lab material.

IV. DISCUSSION

In this study, we collected students' nominations of strong peers in three different remote, introductory physics courses with varying student populations. We advance previous work by measuring both gender and racial or ethnic biases and differentiating perceptions related to lecture and lab contexts. The remainder of this section synthesizes our findings for each research question and concludes by noting recommendations for instruction and limitations to the study.

A. Mixed evidence of gender and racial or ethnic biases in recognition

Our analyses found mixed results regarding the presence or absence of a gender bias in students' recognition of their peers. After adjustments for various measures reflecting structural tendencies of tie formation, women proportionately nominated their male and female peers in all courses and contexts (lecture and lab) except Course C's lecture perception network. In this network, women disproportionately nominated other women over men as strong in the lecture material. In contrast, men proportionately

nominated their male and female peers in three out of six observed networks (lab perception network of Course A and both networks of Course C) after controlling for other network configurations. Men significantly undernominated their female peers in Course A's lecture perception network and in both perception networks of Course B. Recall that if we impute the second celebrity in Course C's lecture perception network as a man, men also significantly undernominated their female peers in this network.

The results related to gender bias in lecture peer perceptions add insight to those found in prior work [30–32]. Across these studies, the courses vary by student population (majors, course level, and gender), instructional type (traditional and interactive lectures or nontraditional labs), and institution. This variability, understandably, leads to different conclusions in each study (including across the courses examined in our study). Contrary to expectations, a course's gender composition (whether gender-balanced, majority men, or majority women) and discipline (whether physical sciences and engineering or biology) do not seem to predict the presence or absence of a gender bias in students' recognition of their peers. Neither does the instructional style—whether traditional lecture, interactive lecture, or lab instruction—or class size (whether the course contains 90 or 400 enrolled students).

One common factor that is consistently associated with gender bias across different studies, however, is the course level. Across the four studies, courses at the first-year level (those in Ref. [30], Courses A and B in our study, and those in Ref. [32], assuming the student populations in the lower level courses are primarily first-year based on typical course enrollments) all exhibit a gender bias in peer recognition, whereas those at the beyond-first-year level (those in Ref. [31] and Course C in our study) do not. We posit that developing familiarity and friendship with peers in previous semesters allows for a more diverse set of students to gain recognition in subsequent courses. Students in beyond-first-year courses, for example, likely have had more opportunities to showcase their knowledge or skills in front of their peers during prior courses they take together. In students' first introductory courses, in contrast, a gender bias in peer recognition aligned with socio-historical stereotypes [12–17] endures before the students get to know each other. Alternatively, this pattern could be due to selection effects where only those women who received substantial recognition in their first course enrolled in subsequent courses. Thus, all of the gender bias may have occurred in the first year courses, creating unequal representation of students in the subsequent courses where we no longer find a bias. We note that this relationship between course level and gender bias in peer recognition is a tentative claim given the celebrity of unknown gender in Course C's lecture perception network.

The modified analyses used in our study also add to the understanding of the nature of the gender bias in

introductory STEM courses. As in the perceptions study in biology [30], we found that, when a gender bias in peer recognition existed, men undernominated women, but women proportionately nominated both men and women. This result differs from that of Bloodhart and colleagues [32], however, who found that *both* men and women undernominated women in introductory life sciences and physics courses. More details about their analyzed courses are necessary to determine which, if any, course features may have led to these different results.

Our study also uniquely evaluated whether a racial or ethnic bias exists in students' recognition of peers. Race or ethnicity was not a significant predictor of nominations in either network of Course A or the lab perception network of Course C. In both networks of Course B and the lecture perception network of Course C (even when imputing the second celebrity of the latter network as a URM or non-URM student), however, URM students were more likely to receive nominations than their non-URM peers. This suggests that, when the nomination probabilities are adjusted for other variables in the model, URM students received more recognition than their non-URM counterparts, despite the documented stereotypes against URM students in science [18–22,67] and indications that URM students report significantly lower senses of recognition than non-URM peers in their physics courses [36,39,44].

We discuss several possible explanations that may have influenced these surprising findings. First, for the networks where we found no racial or ethnic bias (both networks of Course A and the lab perception network of Course C), one might expect low statistical power would explain the lack of a measurable effect: URM students made up less than 30% of each analyzed course. This explanation is unlikely, however, given that we were able to statistically discern an effect in the other networks with comparable proportions of URM and non-URM students. Alternatively, we note that this study was fielded in the aftermath of the Black Lives Matter protests following the murder of George Floyd. Students (especially at Cornell University) were aware of the political climate [68,69], which might have created more awareness of racial or ethnic bias (compared to gender bias) and thus social desirability biases in the responses. We note the plausibility of this explanation given that the effect on URM students' nominations is either unbiased or in the *opposite* direction to what research would have predicted. Another explanation, particularly for the tendency for URM students to nominate URM peers in Course B, is friendship tie homophily. A host of research suggests that friendship serves as a mechanism for recognition [70,71] and that students tend to form friendships with peers of their same race or ethnicity [72,73]. URM students in Course B, therefore, might have formed friendships with one another and in turn recognized one another as strong in the course. Finally, in this study we measured actual recognition, whereas some prior work

measures perceived recognition [36,39,44]. Students' actual recognition may differ from their perceptions of recognition, resulting in these different outcomes. For example, students from underrepresented groups may perceive lower recognition than they are actually given based on their awareness of existing stereotypes. We recommend for future research to directly compare perceived and actual recognition across demographic groups to examine this viable phenomenon.

B. Recognition differs between lecture and lab contexts

Different from previous studies, we probed peer perceptions related to lecture and lab contexts separately. We observed very different network structures across these contexts, with celebrities emerging in two out of three lecture perception networks but in none of the lab perception networks. We suspect that the structure of coursework in each context impacted the distribution of nominations. In the courses we analyzed, lectures contained half or all of the enrolled students (depending on whether there were one or two lecture sections). The few highly motivated students (i.e., the celebrities) who frequently participated in lecture by answering or asking questions in front of the rest of the class likely gained considerable recognition from peers as strong in the lecture context. Because lectures were held on Zoom, students could also readily see the names of these celebrities and recall them on the survey. A similar phenomenon may have also occurred during online office hours, which were (anecdotally) very busy. By contrast, labs were held on Zoom and used breakout rooms, so students were restricted to interacting with and seeing the names of just a few peers. Lab groups were also held stable throughout the semester, allowing for meaningful yet limited recognition [37,74]. We note that while discussion sections also used online breakout rooms, they did not necessitate interaction between students (students submitted individual work, if at all, and some groups would leave their cameras and microphones off and work independently). Labs, on the other hand, required students to set up their experiments, collect and analyze data, and coordinate lab notes for a weekly group submission, all of which were negotiated through conversations.

Our findings pertaining to gender, moreover, suggest that students perceive male and female peers' expertise with lecture and lab material differently. The presence or absence of a gender bias in peer recognition varied between contexts in Course A, with a gender bias in lecture perceptions but not in lab perceptions. In Course B, there was a stronger gender bias in lecture perceptions than lab perceptions. The lecture, but not lab, perception results mostly agree with prior work, which has found a gender bias in course-level perceptions [30,32]. We speculate that, as was the case in the study of Grunspan and colleagues [30], men may have been more verbally outspoken than women in the lecture sections (though we did

not measure this). Indeed, it has been shown in introductory science courses that women are less comfortable participating in whole-class discussions than men [75] and that women respond less frequently than men to instructor-posed questions to the class [28,76]. In remote courses in particular, research has found that men participate more than women both verbally and in the chat window and that students acknowledge chat messages from male peers more than female peers [77]. Recognition in labs, on the other hand, likely occurs among students working together in their lab groups. In our observed courses, lab groups were created based on a group-forming survey where students could indicate their preferences related to group work, for instance if they wished to work with (or not work with) certain peers. Instructors intentionally created lab groups based on the survey and also avoided groups with isolated women. One possible explanation for observing less gender bias in labs, therefore, is that women had sufficient opportunities for recognition within their majority-women or all-women lab groups. Alternatively, students may hold different conceptions of what it means to be “strong in the lecture/discussion section material” and “strong in the lab material.” This seems plausible given that we found more gender bias in the lecture context than the lab context, yet stereotypes typically associate physics and masculinity with technical skills (lab) *and* natural brilliance (lecture) [12–17]. To explore this possibility, we have modified the perceptions survey to also ask students to briefly explain their nominations. We will use these data to unpack the traits or behaviors students attribute to being strong in each context and determine whether this explains the difference in gender bias.

We also found that higher final course grades predict more received nominations from peers across all courses and contexts, in agreement with prior work [30,31]. Results related to discussion forum contributions and section enrollment, however, varied by context. We found that students’ contributions to the course discussion forum was a significant predictor of receiving nominations in lecture, but not in lab, perception networks. We suspect that this difference occurred because most content posted on the discussion forum was related to lecture material rather than lab material. Participation in the discussion forum likely served as a means of becoming more visible to and recognized by peers, but only in regard to the content being discussed. The lecture perception networks, moreover, contained emergent celebrities in the two courses using a discussion forum, Courses B and C, but not in Course A. Frequent visibility in the discussion forum, where students’ names are explicitly tied to their contributions, might be a mechanism for students receiving many nominations as strong in their course. Similar to previous studies [30,31], we also observed that lab section enrollment is important for shaping peer perceptions in both contexts, while discussion section enrollment is only a strong predictor of lecture perceptions. In other words,

students learn about one another’s strengths and acquire recognition related to lecture material in both discussion and lab sections, while they learn about each other’s strengths related to lab material only in lab section.

C. Recommendations for instruction

Together, our results related to gender and racial or ethnic bias in peer recognition point to courses (first-year level) and contexts (both lecture and lab) in which students from underrepresented backgrounds (mostly women) may receive less recognition than their peers and, therefore, may be at a disadvantage for developing their physics identity [37,42]. Because recognition is one of the most important dimensions of physics identity [18,34–41], instructors may support all students’ identity development by facilitating more equitable peer recognition.

Our findings suggest that instructors, particularly of STEM courses at the first-year level, should aim to create opportunities for meaningful recognition in all aspects of a course. For example, research suggests that friendship and collaboration with peers is one mechanism for recognition: interacting with others allows for students to showcase their knowledge and skills and acquire validation from others [70,78]. Opportunities for recognition, therefore, may be achieved through more student-centered instructional styles, where students work closely with one another in small groups [37]. Though small group work is already common in labs, lectures often place emphasis on individuals answering questions in front of the class. Poll questions and other active learning activities implemented in lectures may be turned into group discussions and group submissions rather than individual work. Further, if students are presenting group work in front of the whole class, instructors can create opportunities for positive recognition by allowing groups to discuss the ideas before asking them to share, increasing the likelihood of groups landing on the correct answer [79].

In terms of forming groups, prior research suggests that historically underrepresented students (women and URM students) benefit from working in groups where they are not isolated [80,81]. One study, for example, found that gender homogeneous and majority-women groups performed better when solving physics problems than majority-men groups [80]. The researchers observed in these majority-men groups that the men dominated the conversation, ignoring suggestions from their female peer. Our results agree with this work from a different perspective, namely recognition. We found that the gender bias in peer recognition was weaker in the lab context (where groups intentionally avoided isolated women) than the lecture context (where any group work was completed in unintentionally formed groups). Our results, therefore, support the prior work recommending that underrepresented students (based on both gender and race or ethnicity) be placed in groups where they are not the minority. Previous research

also suggests that students tend to become friends with, and therefore may be more prone to recognize, peers of similar academic achievement [71]. Intentionally forming heterogeneous groups of students based on performance, therefore, might enable students of different levels of academic achievement to recognize each other.

As to whether groups should be held the same or changed throughout the semester, our study cannot make a strong recommendation one way or another. Close and consistent collaboration with group members seems to allow students to overcome any implicit biases with which they enter the course [74]. Meanwhile, altering groups may be beneficial in allowing students to gain recognition from many of their peers.

Outside of group work, instructor-posed questions to the whole class can still provide many students with opportunities to gain recognition. For a given question, an instructor may ask for multiple volunteers to share their thinking before hearing from any individual student (“many hands”) or randomly select individuals to share (“warm” or cold call) [82]. The instructor can also explicitly ask for different volunteers each time. Instructors should also be cautious when allotting praise to students’ answers to the whole class. If a student’s answer is followed by “Perfect!” there is little room for other students to contribute or ask questions, limiting exposure to peers to just the one student who volunteered.

D. Limitations

We end this section by acknowledging multiple limitations of our study. First and foremost, we collected our data during a global pandemic. Students’ learning experiences were certainly impacted by this event [83–87] and, as a result, our findings may not be generalizable to physics instruction during normal circumstances. In addition, the courses analyzed here were almost entirely held online. While our results align with some of the previous work from in-person courses, future work should perform a more systematic comparison of peer perceptions in face-to-face and remote courses.

With regard to our methods, our perceptions survey may not have captured all nominations. We did not provide students with a list of names to look at when filling out the perceptions survey, so students may not have remembered the names of individuals they perceived as strong in the material. Additionally, we only collected survey responses at the midpoint of the semester. Other work administered surveys either both at the middle and end of the course [30] or just at the end point of the course [31]. Future work comparing physics students’ perceptions at multiple points in the course or just at the end of the course may add nuance to our findings.

We also performed text processing on the surveys to match the reported names to the class roster. This process dropped fewer than 5% of the reported nominations

(for instance, due to students misspelling a peer’s name). We note the possibility for bias in the text processing because certain kinds of names might be less likely to be matched and thus more likely to be dropped from analysis. On one hand, students with common first or last names in the class might be less likely to be matched, particularly if only their first or last name is listed. Complicated names might be more prone to misspelling and, therefore, may also have a low probability of being matched. Rare first or last names, on the other hand, are more likely to be matched because they are unique. Whether certain kinds of students, for instance non-URM and URM students, have common or rare names might influence the representation of demographic groups in the data. Because we were able to match a high percentage of the data (more than 95%), however, we do not believe this potential bias impacted our study’s results.

In addition, we categorized race or ethnicity in terms of URM status because the number of students in each racial or ethnic group was too small for our quantitative analysis to produce useful and interpretable results. However, this inevitably masks differences in recognition between students of individual racial or ethnic groups. Future research should seek to study more diverse student populations with statistically sufficient representation from all racial or ethnic groups. We also treated gender and race or ethnicity separately. It seems valuable for future work to determine whether gender and race or ethnicity are separately important for peer perceptions or whether it is the *intersection* between gender and race or ethnicity that significantly explains recognition patterns.

Finally, peer recognition might depend on other variables that we did not measure or analyze in this study. For example, research suggests that students view their friends as strong in the course [70,71]. To determine whether this is the case, students’ friendship ties could be collected and added as a predictor variable in the statistical model. Students’ majors might also be related to recognition: perhaps students view peers in particular majors (e.g., STEM majors or physics majors in particular) as stronger in physics than peers in other majors (e.g., life sciences or non-STEM majors). We did not examine the relationship between student majors and recognition in this study because the courses were quite homogeneous on major (either most students were studying engineering or most students were studying physics), many students did not report their subfield within the engineering school, and some students did not yet declare a major. Future work should investigate whether and how friendship ties, students’ majors, and other variables relate to students’ perceptions of strong peers.

V. CONCLUSION

Examining students’ nominations of strong peers, we found variation in gender and racial or ethnic biases across three different remote, introductory physics courses. We observed that courses primarily serving first-year

students exhibited a gender bias in lecture perceptions, while those serving beyond-first-year students did not. Additionally, URM students were either more or equally likely to receive nominations than their non-URM peers, contrary to what prior research would predict. Recognition also varied between lecture and lab contexts. Lecture perception networks contained a few central students receiving many nominations, however no outstanding celebrities emerged in the lab perception networks. These results suggest that recognition varies within different student populations and instructional contexts. Findings also point to advantages of instruction that emphasizes small group work and allows for many different students to speak up in front of the class. These instructional efforts are hopefully the first steps toward creating more widespread, rather than skewed, distributions of peer recognition, such that all students can develop their physics identities—a critical predictor of participation, persistence, and career intentions in physics and other science disciplines.

ACKNOWLEDGMENTS

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program Grant No. DGE-2139899 and National Science Foundation Grant No. DUE-1836617. The authors thank Cole Walsh, David Wu, Andy Schang, and David Esparza for providing meaningful feedback on this work.

APPENDIX: ORIGINAL MODEL RESULTS AND GOODNESS-OF-FIT COMPARISON

Table V shows the coefficient estimates using the original model from Ref. [30] with our data. We include four additional variables to measure discussion section homophily and racial or ethnic bias. Figure 5 compares the goodness-of-fit diagnostics of the original model and our revised model (presented in the main text) for Course A's lecture perception network. The horizontal axis represents a network measure—outdegree (number of nominations

TABLE V. ERGM coefficient estimates using the original model of Grunspan and colleagues [30]. Standard errors of the coefficient estimates are in parentheses. Asterisks indicate statistical significance (* $p < 0.05$; ** $p < 0.01$). For every network, the goodness-of-fit diagnostics for the revised model are better than those for the original model (shown in Fig. 5).

	Course A		Course B		Course C	
	Lecture	Lab	Lecture	Lab	Lecture	Lab
<i>Edges</i>	−9.00 ** (0.76)	−8.12 ** (0.63)	−13.90 ** (1.57)	−9.07 ** (1.04)	−7.69 ** (0.45)	−9.79 ** (0.56)
<i>Mutuality</i>	3.38 ** (0.32)	2.50 ** (0.29)	2.23 ** (0.36)	1.76 ** (0.43)	4.26 ** (0.29)	2.54 ** (0.24)
<i>Isolates (0-indegree)</i>	0.53 * (0.25)	0.23 (0.24)	0.68 (0.49)	0.03 (0.38)	1.02 ** (0.18)	0.02 (0.19)
<i>Female nominator</i>	0.20 (0.24)	0.14 (0.25)	0.65 (0.42)	0.41 (0.45)	0.29 (0.18)	0.16 (0.21)
<i>Woman-woman bias</i>	0.23 (0.21)	0.30 (0.21)	−0.56 (0.71)	0.67 (0.44)	0.09 (0.14)	0.10 (0.18)
<i>Man-man bias</i>	0.52 * (0.21)	0.48 * (0.20)	1.01 ** (0.37)	0.70 * (0.35)	0.307 (0.16)	0.52 ** (0.178)
<i>URM nominator</i>	0.06 (0.25)	0.14 (0.25)	−0.37 (0.31)	−0.01 (0.46)	−0.61 ** (0.21)	0.11 (0.21)
<i>URM-URM bias</i>	−0.26 (0.55)	−0.07 (0.47)	1.52 ** (0.56)	1.19 (0.64)	0.82 ** (0.23)	−0.13 (0.26)
<i>Non-URM—non-URM bias</i>	0.04 (0.15)	−0.03 (0.15)	−0.73 ** (0.23)	0.08 (0.31)	−0.26 * (0.12)	0.35 * (0.15)
<i>Final course grade of nominee</i>	0.79 ** (0.18)	0.29 (0.15)	2.41 ** (0.37)	0.58 * (0.26)	0.41 ** (0.12)	0.43 ** (0.14)
<i>Discussion forum contributions of nominee</i>	N/A	N/A	0.03 ** (0.004)	0.01 * (0.006)	0.01 (0.006)	0.007 (0.009)
<i>Homophily on lab section</i>	1.20 ** (0.15)	3.68 ** (0.20)	0.64 ** (0.18)	3.20 ** (0.34)	1.13 ** (0.15)	4.37 ** (0.16)
<i>Homophily on discussion section</i>	1.84 ** (0.15)	0.14 (0.21)	0.61 ** (0.19)	0.13 (0.24)	1.77 ** (0.14)	0.46 * (0.18)

Course A: Lecture Perceptions

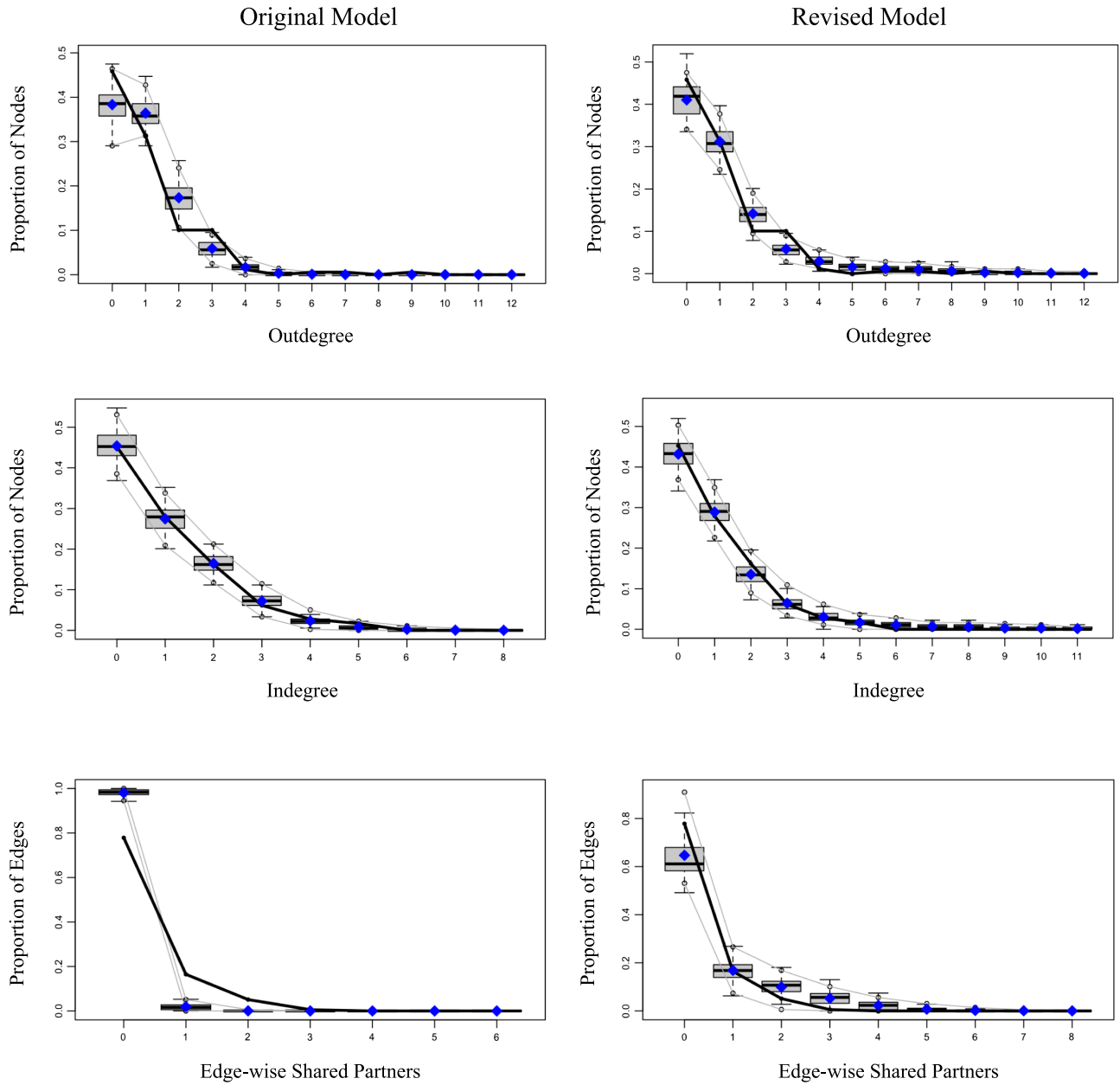


FIG. 5. Goodness-of-fit diagnostics for both the original ERGM model [30] and the revised ERGM model for the lecture perception network of Course A.

made), indegree (number of nominations received), and edge-wise shared partners (measure of triadic closure)—and the vertical axis represents the proportion of students or edges in the network. Each plot shows the distribution of the network measure in the observed data (thick black line)

and in 10 networks simulated using the estimated model coefficients (boxplots). While the original model captures the outdegree and indegree distributions sufficiently, the revised model significantly improves how well the model captures the distribution of edge-wise shared partners.

- [1] Radhika Suresh, The relationship between barrier courses and persistence in engineering, *J. College Student Retention: Res. Theory Pract.* **8**, 215 (2006).
- [2] Jessica Watkins and Eric Mazur, Retaining students in science, technology, engineering, and mathematics (STEM) majors, *J. Coll. Sci. Teach.* **42**, 36 (2013), https://mazur.harvard.edu/files/mazur/files/rep_685.pdf.
- [3] Maya A. Beasley and Mary J. Fischer, Why they leave: The impact of stereotype threat on the attrition of women and minorities from science, math and engineering majors, *Social Psychol. Educ.* **15**, 427 (2012).
- [4] Catherine Riegler-Crumb, Barbara King, and Yasmiyn Irizarry, Does STEM stand out? Examining racial/ethnic gaps in persistence across postsecondary fields, *Educ. Res.* **48**, 133 (2019).
- [5] Jacob Clark Blickenstaff, Women and science careers: Leaky pipeline or gender filter?, *Gender Educ.* **17**, 369 (2005).
- [6] 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).
- [7] Sapna Cheryan, Sianna A. Ziegler, Amanda K. Montoya, and Lily Jiang, Why are some STEM fields more gender balanced than others?, *Psychol. Bull.* **143**, 1 (2017).
- [8] L. Merner and J. Tyler, African-American participation among bachelors in the physical sciences and engineering: Results from the 2005 to 2015 data of the National Center for Education Statistics (2019), <https://www.aip.org/statistics/reports/african-american-participation-among-bachelors-physical-sciences>.
- [9] Susan C. White, Race and ethnicity of physics faculty members, *Phys. Teach.* **56**, 539 (2018).
- [10] Angela Johnson, Rose Young, and Elizabeth Mulvey, Where do women of color complete physics degrees?, *Phys. Teach.* **58**, 620 (2020).
- [11] Anne Marie Porter and Rachel Ivie, *Women in physics and astronomy*, 2019 (AIP Statistical Research Center, 2019), <https://www.aip.org/statistics/reports/women-physics-and-astronomy-2019>.
- [12] Anna T. Danielsson, Exploring woman university physics students 'doing gender' and 'doing physics,' *Gender Educ.* **24**, 25 (2012).
- [13] 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).
- [14] 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).
- [15] Toni Schmader, Michael Johns, and Marchelle Barquissau, The costs of accepting gender differences: The role of stereotype endorsement in women's experience in the math domain, *Sex Roles* **50**, 835 (2004).
- [16] Elena Makarova and Walter Herzog, Trapped in the gender stereotype? The image of science among secondary school students and teachers, *Equality, Diversity and Inclusion* **34**, 106 (2015).
- [17] 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).
- [18] 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).
- [19] 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).
- [20] Robert Ceglie, Underrepresentation of women of color in the science pipeline: The construction of science identities, *J. Women Minorities Sci. Engin.* **17**, 271 (2011).
- [21] Jennifer M. Grossman and Michelle V. Porche, Perceived gender and racial/ethnic barriers to STEM success, *Urban Educ.* **49**, 698 (2014).
- [22] B. E. Blaine, *Understanding Race, Racial Stereotypes, and Racism* (Sage Publications, Inc., Thousand Oaks, CA, 2013).
- [23] Roberta M. Hall and Bernice R. Sandler, *The Classroom Climate: A Chilly One for Women?* (Association of American Colleges, Washington, DC, 1982).
- [24] Ramona Gunter and Amy Stambach, Differences in men and women scientists' perceptions of workplace climate, *J. Women Minorities Sci. Engin.* **11**, 97 (2005).
- [25] Elaine Seymour and Nancy M. Hewitt, *Talking about Leaving: Why Undergraduates Leave the Sciences* (Westview Press, Boulder, CO, 1997).
- [26] Lisa Rosenthal, Bonita London, Sheri Robin Levy, and Marci Lobel, The roles of perceived identity compatibility and social support for women in a single-sex STEM program at a co-educational university, *Sex Roles* **65**, 725 (2011).
- [27] Sonja Cwik and Chandralekha Singh, Damage caused by societal stereotypes: Women have lower physics self-efficacy controlling for grade even in courses in which they outnumber men, *Phys. Rev. Phys. Educ. Res.* **17**, 020138 (2021).
- [28] 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).
- [29] 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).
- [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] 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).

- [33] James Paul Gee, Chapter 3: Identity as an analytic lens for research in education, *Rev. Res. Educ.* **25**, 99 (2000).
- [34] 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).
- [35] 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).
- [36] 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).
- [37] Zahra Hazari and Cheryl Cass, Towards meaningful physics recognition: What does this recognition actually look like?, *Phys. Teach.* **56**, 442 (2018).
- [38] Zahra Hazari, Eric Brewae, 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).
- [39] 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).
- [40] Allison Godwin, Geoff Potvin, Zahra Hazari, and Robynne Lock, Identity, critical agency, and engineering: An affective model for predicting engineering as a career choice, *J. Engin. Educ.* **105**, 312 (2016).
- [41] Allison Godwin and Geoff Potvin, Pushing and pulling Sara: A case study of the contrasting influences of high school and university experiences on engineering agency, identity, and participation, *J. Res. Sci. Teach.* **54**, 439 (2017).
- [42] 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).
- [43] Lucy Avraamidou, Identities in/out of physics and the politics of recognition, *J. Res. Sci. Teach.* **59**, 58 (2022).
- [44] 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), <https://www.jstor.org/stable/43631586>.
- [45] Louise Archer, Julie Moote, Becky Francis, Jennifer DeWitt, and Lucy Yeomans, The "exceptional" physics girl: A sociological analysis of multimethod data from young women aged 10–16 to explore gendered patterns of post-16 participation, *Am. Educ. Res. J.* **54**, 88 (2017).
- [46] 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).
- [47] Joseph Kozminski, Heather Lewandowski, Nancy Beverly, Steve Lindaas, Duane Deardorff, Ann Reagan, Richard Dietz, Randy Tagg, M. Eblen Zayas, J. Williams *et al.*, *AAPT Recommendations for the Undergraduate Physics Laboratory Curriculum* (American Association of Physics Teachers, College Park, MD, 2014).
- [48] N. G. Holmes and Carl E. Wieman, Introductory physics labs: We can do better, *Phys. Today* **71**, 1, 38 (2018).
- [49] Emily M. Smith and N. G. Holmes, Best practice for instructional labs, *Nat. Phys.* **17**, 662 (2021).
- [50] 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).
- [51] N. G. Holmes, Carl E. Wieman, and D. A. Bonn, Teaching critical thinking, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 11199 (2015).
- [52] Sara A. Newland, Review of learning catalytics, an interactive student response tool, *J. Political Sci. Educ.* **17**, 159 (2021).
- [53] 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).
- [54] Adrienne L. Traxler, Tyme Suda, Eric Brewae, and Kelley Commeford, Network positions in active learning environments in physics, *Phys. Rev. Phys. Educ. Res.* **16**, 020129 (2020).
- [55] Tom A. B. Snijders and Stephen P. Borgatti, Non-parametric standard errors and tests for network statistics, *Connections* **22**, 161 (1999), http://www.analytictech.com/borgatti/papers/snijders_borgatti-density_significance.pdf.
- [56] Yuzhou Chen, Yulia R. Gel, Vyacheslav Lyubchich, and Kusha Nezafati, Snowboot: Bootstrap methods for network inference, *arXiv:1902.09029*.
- [57] Carolyn J. Anderson, Stanley Wasserman, and Bradley Crouch, A p* primer: Logit models for social networks, *Soc. Networks* **21**, 37 (1999).
- [58] Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher, An introduction to exponential random graph (p*) models for social networks, *Soc. Networks* **29**, 173 (2007).
- [59] 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).
- [60] David R. Hunter, Curved exponential family models for social networks, *Soc. Networks* **29**, 216 (2007).
- [61] Ruth M. Hummel, David R. Hunter, and Mark S. Handcock, Improving simulation-based algorithms for fitting ERGMs, *J. Comput. Graph. Stat.* **21**, 920 (2012).
- [62] 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*.
- [63] Carter T. Butts, Martina Morris, Pavel N. Krivitsky, Zack Almquist, Mark S. Handcock, David R. Hunter, Steven M. Goodreau, and Skye Bender de Moll, Introduction to exponential-family random graph (erg or p*) modeling with ERGM. European University Institute, Florence,

- <https://cran.r-project.org/web/packages/ergm/vignettes/ergm.pdf>, (2014).
- [64] James E. Wells. Modeling student collaborations using ERGMs, *Proceedings of the PER Conference 2019, Provo, UT*, 10.1119/perc.2019.pr.Wells.
- [65] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevPhysEducRes.18.020148> for the same network diagrams shown in the main text with nodes colored by students' race or ethnicity and plots comparing the indegree distributions to final course grades and discussion forum contributions.
- [66] Scott W. Duxbury, The problem of scaling in exponential random graph models, *Sociol. Methods Res.* **0049124120986178** (2021).
- [67] Asia A. Eaton, Jessica F. Saunders, Ryan K. Jacobson, and Keon West, How gender and race stereotypes impact the advancement of scholars in STEM: Professors' biased evaluations of physics and biology post-doctoral candidates, *Sex Roles* **82**, 127 (2020).
- [68] Matt Steecker, Cornell community members join George Floyd protests against police brutality, racism, Ithaca Journal (2020), <https://www.ithacajournal.com/story/news/local/2020/06/03/cornell-university-community-members-protest-ithaca-after-floyds-death/3133045001/>.
- [69] Maya Rader, A year later, Cornell event reflects on George Floyd protests, The Cornell Daily Sun (2021), <https://cornellsun.com/2021/04/07/a-year-later-cornell-event-reflects-on-george-floyd-protests/>.
- [70] Dina Zohrabi Alaei, Micah K. Campbell, and Benjamin M. Zwickl, Impact of virtual research experience for undergraduates experiences on students' psychosocial gains during the COVID-19 pandemic, *Phys. Rev. Phys. Educ. Res.* **18**, 010101 (2022).
- [71] Lydia Laninga-Wijnen, Allison M. Ryan, Zeena Harakeh, Huiyoung Shin, and Wilma A. M. Vollebergh, The moderating role of popular peers' achievement goals in 5th- and 6th-graders' achievement-related friendships: A social network analysis, *J. Educ. Psychol.* **110**, 289 (2018).
- [72] Miller McPherson, Lynn Smith-Lovin, and James M. Cook, Birds of a feather: Homophily in social networks, *Annu. Rev. Sociol.* **27**, 415 (2001).
- [73] Sergio Currarini, Matthew O. Jackson, and Paolo Pin, Identifying the roles of race-based choice and chance in high school friendship network formation, *Proc. Natl. Acad. Sci. U.S.A.* **107**, 4857 (2010).
- [74] Danny Doucette and Chandralekha Singh, Why are there so few women in physics? Reflections on the experiences of two women, *Phys. Teach.* **58**, 297 (2020).
- [75] Sarah L. Eddy, Sara E. Brownell, Phonraphee Thummaphan, Ming-Chih Lan, and Mary Pat Wenderoth, Caution, student experience may vary: Social identities impact a student's experience in peer discussions, *CBE Life Sci. Educ.* **14**, ar45 (2015).
- [76] Stephanie M. Aguillon, Gregor-Fausto Siegmund, Renee H. Petipas, Abby Grace Drake, Sehoya Cotner, and Cissy J. Ballen, Gender differences in student participation in an active-learning classroom, *CBE Life Sci. Educ.* **19**, ar12 (2020).
- [77] Sierra C. Nichols, Yongyong Y. Xia, Mikaylie Parco, and Elizabeth G. Bailey, Participation and performance by gender in synchronous online lectures: Three unique case studies during emergency remote teaching, *J. Microbiol. Biol. Educ.* **23**, e00281 (2022).
- [78] Gina M. Quan, Chandra Turpen, and Andrew Elby, Analyzing identity trajectories within the physics community, *Phys. Rev. Phys. Educ. Res.* **18**, 020125 (2022).
- [79] Michelle K. Smith, William B. Wood, Wendy K. Adams, Carl Wieman, Jennifer K. Knight, Nancy Guild, and Tin Tin Su, Why peer discussion improves student performance on in-class concept questions, *Science* **323**, 122 (2009).
- [80] Patricia Heller and Mark Hollabaugh, Teaching problem solving through cooperative grouping. Part 2: Designing problems and structuring groups, *Am. J. Phys.* **60**, 637 (1992).
- [81] Lauren L. Sullivan, Cissy J. Ballen, and Sehoya Cotner, Small group gender ratios impact biology class performance and peer evaluations, *PLoS One* **13**, e0195129 (2018).
- [82] Kimberly D. Tanner, Structure matters: Twenty-one teaching strategies to promote student engagement and cultivate classroom equity, *CBE Life Sci. Educ.* **12**, 322 (2013).
- [83] Bethany Wilcox and Michael Vignal, Recommendations for emergency remote teaching based on the student experience, *The Physics Teacher* **58**, 374 (2020).
- [84] Matthew Dew, Lewis Ford, Dawson T. Nodurft, Tatiana Erukhimova, and Jonathan Perry, Student responses to changes in introductory physics learning due to the COVID-19 pandemic, *Phys. Teach.* **59**, 162 (2021).
- [85] Pascal Klein, Lana Ivanjek, Merten Nikolay Dahlkemper, Katarina Jeličić, M. A. Geyer, Stefan Küchemann, and Ana Susac, Studying physics during the COVID-19 pandemic: Student assessments of learning achievement, perceived effectiveness of online recitations, and online laboratories, *Phys. Rev. Phys. Educ. Res.* **17**, 010117 (2021).
- [86] Irene Marzoli, Arturo Colantonio, Claudio Fazio, Marco Giliberti, Umberto Scotti di Uccio, and Italo Testa, Effects of emergency remote instruction during the COVID-19 pandemic on university physics students in Italy, *Phys. Rev. Phys. Educ. Res.* **17**, 020130 (2021).
- [87] Danny Doucette, Sonja Cwik, and Chandralekha Singh, Everyone is new to this: Student reflections on different aspects of online learning, *Am. J. Phys.* **89**, 1042 (2021).