

Computing Educational Activities Involving PEOPLE Rather Than THINGS Appeal More to Women (Recruitment Perspective)

Ingrid Maria Christensen*
CCER: Center for
Computing Education Research
IT University of Copenhagen
Copenhagen, Denmark

Melissa Høegh Marcher*
CCER: Center for
Computing Education Research
IT University of Copenhagen
Copenhagen, Denmark

Paweł Grabarczyk
CCER: Center for
Computing Education Research
IT University of Copenhagen
Copenhagen, Denmark

Therese Graverson
CCER: Center for
Computing Education Research
IT University of Copenhagen
Copenhagen, Denmark

Claus Brabrand
CCER: Center for
Computing Education Research
IT University of Copenhagen
Copenhagen, Denmark

ABSTRACT

There is a strong need for a more equal gender balance within the computing field. In 1998, Richard A. Lippa [29] uncovered a relationship between gender and preference within the PEOPLE–THINGS spectrum, with women preferring PEOPLE-oriented activities to a higher degree than men. The aim of this paper is twofold. First of all, we wish to determine if a similar relation can be established in the particular context of computing educational activities. Second of all, we wish to see if Lippa’s findings can be extrapolated to contemporary high-school students. To do that, we designed and conducted an experiment involving around 500 Danish high-school students who have been asked to choose between a PEOPLE-themed version vs an isomorphic THINGS-themed version of four activities representative for computing education. The results show that the odds of a woman preferring a task involving PEOPLE is 2.7 times higher than those of a man. The odds of a student *without* prior programming experience preferring a task involving PEOPLE is 1.4 times higher than those of a student *with* programming experience. If we compare women without programming experience to men with programming experience the effect is even more pronounced; indeed, the combined effect is 3.8 (2.7×1.4). Our study implies a recommendation for computing educators to, whenever possible, favor educational activities involving PEOPLE over THINGS. This makes educational activities appeal more to female students (and to students without programming experience), while not making a difference for male students (or students with programming experience). Since the experiment measured only the appeal of tasks (the users were not expected to perform them) the results we obtained

can be useful for recruitment processes where the overall image and appeal of material makes a difference.

CCS CONCEPTS

• **Social and professional topics** → **Computing education programs**; *Computer science education*; *Information technology education*; *Software engineering education*.

KEYWORDS

recruitment, computing, computer science, software engineering, STEM, education, student experiments, diversity, gender

ACM Reference Format:

Ingrid Maria Christensen, Melissa Høegh Marcher, Paweł Grabarczyk, Therese Graverson, and Claus Brabrand. 2021. Computing Educational Activities Involving PEOPLE Rather Than THINGS Appeal More to Women (Recruitment Perspective). In *Proceedings of the 17th ACM Conference on International Computing Education Research (ICER 2021), August 16–19, 2021, Virtual Event, USA*. ACM, New York, NY, USA, 18 pages. <https://doi.org/10.1145/3446871.3469758>

1 INTRODUCTION

Women are widely underrepresented in Information and Communication Technology (ICT) jobs. In the EU, only 17.9% of people working in ICT jobs in 2019 were women [23]. In the US, women held just 25% of computing occupations in 2015 [5]. There are, at least, three reasons why this state of affairs demands action.

The first reason is that, when it comes to gender equality, computing is still, sadly, lagging behind some of the other university disciplines. As pointed out by Janet Abbate [1], the early days of computing were much more progressive in this respect. The decrease in the number of women who major in computer science [3] suggests that the methods used to recruit new students are biased and ineffective at appealing to and attracting the female population. There is no doubt that gender inequality remains to be one of the most important human aspects of computing.

The second reason is that ICT skills are currently needed across all sectors as there is a large demand for ICT specialists, which many companies struggle to meet. In 2017, 53% of EU companies that tried to recruit ICT specialists reported difficulties in filling vacancies [22] and the current supply of ICT and STEM graduates

*Both authors contributed equally to this research.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
ICER 2021, August 16–19, 2021, Virtual Event, USA

© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8326-4/21/08...\$15.00
<https://doi.org/10.1145/3446871.3469758>

is insufficient to meet the demand [21]. Educating and hiring more women in the ICT industry could contribute to catching up with the high demand. Recruiting from only “half of the talent pool” is a big waste of talent.

Thirdly, more balanced gender distribution in the ICT industry could also contribute to a less biased discourse and production of software, as well as un-stereotyping the computing environment. Software engineering is a key force in the digital transformation of society and plays an increasing role in shaping our future. Since the artifacts we design always reflect the values and biases of their creators (sometimes in a tacit way) [10] they need to be created by a more representative part of the population. What it means is that we badly need more women in computing.

From now on, we will deliberately use the term *computing* which is intended as a broader term, including ICT, Software Engineering, and Computer Science.

As easy as it is to recognize the need for action, identifying the main causes why computing education fails at attracting women is a difficult task because it may be the result of many social and systemic factors. The initial hypothesis of our experiment described in Section 3 came from a study conducted by Richard A. Lippa in 1998 [29] that shows a correlation between vocational interests and the location of a given task on the scale between PEOPLE and THINGS oriented topics.

2 BACKGROUND

In 1998, Richard A. Lippa [29] put this claim to the test and established empirical evidence that:

Gender is strongly correlated with the preference for working with PEOPLE vs THINGS: Women are, on average, more interested in working with PEOPLE than men; whereas men are, on average, more interested in working with THINGS, than women.

– Richard A. Lippa, 1998 [paraphrased]

Lippa’s 1998 paper presents three studies that measure the vocational interests of two groups of psychology students and a group of twin pairs. The participants were asked to rank vocations (in the case of the second experiment also some every-day tasks) on an interest scale. The results of Lippa’s experiments were as follows: In the first study, only 20 out of 103 male respondents leaned towards PEOPLE-oriented vocations. In contrast to this, female participants displayed the exact opposite inclinations: 130 out of 186 chose PEOPLE-oriented vocations as preferred. The second study revealed a similar pattern: THINGS-oriented vocations and activities have been chosen by 124 out of 148 male participants and only by 65 out of 246 female participants. The third study showed similar results to the other two: The majority of female participants preferred PEOPLE-oriented vocations and the majority of male participants preferred THINGS-oriented vocations. A similar conclusion has been reported by Lippa in a paper from 2010 [30], where he stated that “Men tend to be much more THINGS-oriented and much less PEOPLE-oriented than women” (with a mean standard deviation of 1.18).

We believe that these results suggest a pattern that may be important for the transformation of computing education, but that their results cannot be trivially extrapolated to this new, specific context. Firstly, it is important to check if a similar pattern can

be found in a different cultural context of Western-European students. Second of all, the sub-optimal aspect of Lippa’s experiment was that it tested vocational preferences on respondents who had already made a certain vocational choice (choosing psychology as their study subject). Third of all, the original experiments test vocational preferences on a very general scale as the respondents were choosing amongst many different vocations. We believe that it may be useful to narrow down the experiment so it can focus on a single discipline and tasks that are representative of it. The experiment described below is an attempt to accommodate all of these differences in scope. To address the first and the second issue, the experiment replaces the context of an American university with a Danish high-school. To address the third issue, it presents the respondents with specific representative computing educational tasks (instead of more generic vocation descriptions and everyday activities used in Lippa’s experiment).

3 EXPERIMENT

We now describe the design and execution of our experiment.

3.1 Objectives

Lippa’s original studies [29] involving American twins (from 1976) and Californian college students (from the 1990es) demonstrated a *strong correlation* of gender with a preference for tasks involving PEOPLE vs THINGS.

Our objective is to investigate to what extent this correlation applies to our context of contemporary (2020) Danish high-school students (as a proxy for future university students). The idea is that if the correlation indeed extends to our context, it ought to be possible to “take advantage of the correlation” when designing computing educational activities. In particular, an educator could, e.g., consciously favor presenting educational tasks involving PEOPLE OVER THINGS to make them appeal more to women and thereby influence the perception of the discipline.

To this end, we designed an experiment to quantify the effect of orienting educational activities towards PEOPLE versus THINGS on women and men.

This paper will address three research questions:

RQ1: To what extent does the gender of high-school students impact preferences for PEOPLE vs THINGS in computing educational activities?

RQ2: To what extent does prior programming experience impact preferences?

RQ3: To what extent do preferences vary between tasks?

3.2 Hypothetical Tasks

We selected four qualitatively different types of educational tasks (aka, teaching/learning activities), representative of what computing students are often exposed to: a *project*, an *article*, a *presentation*, and an *exercise*. All educational activities were hypothetical in that the students did not actually have to perform the tasks, but merely choose which of two versions of the task they would prefer to complete. (Recall that our subjects were high-school students and thus

Table 1: Characterization of the four educational tasks (T1–T4).

TASK #	ACTIVITY	COMPETENCE	VISUALS	TOPIC	LENGTH
T1	project	develop	website	parts-to-whole compatibility	88 words, 1 image
T2	article	summarize	document	technology news article about AI	68 words, 1 image
T3	presentation	evaluate	mobile app	organizational planning	90 words, 1 image
T4	exercise	implement	assignment	programming constructs	156 words, 2 images

not necessarily schooled in computing.) Table 1 gives a characterization of the tasks in terms of their activity, competence, visuals, topic, and length. Task T1 was a software development *project* about the creation of a web service involving parts-to-whole compatibility (see Figure 1). Task T2 involved summarizing a news *article* about the use of artificial intelligence. Task T3 was about evaluating a *presentation* of an organizational planning application developed by another group of students. Task T4 was a simple programming *exercise* involving variables and conditional statements. All tasks came with a short introduction (see, for example, Figure 1).

3.3 Treatment

As treatment, we created two *isomorphic* versions of all tasks: a PEOPLE-themed version and a THINGS-themed version. Figure 1 depicts task T1 with the PEOPLE-version to the left and the THINGS-version to the right. We went through several iterations of the tasks to make sure that the final versions differed only with respect to the theme (PEOPLE vs THINGS) and not regarding other aspects such as, for instance, new vs old or useful vs useless.

We made sure the versions were *syntactically isomorphic*; i.e., that the sentences followed the same structure. Consider, for instance, the text from Task T1 where the differences have been highlighted in boldface:

“Compatible”	personality	types”	[PEOPLE version]
“Compatible”	spare	parts”	[THINGS version]
<<adjective>>	<<noun>>	<<noun (plural)>>	

We also applied this principle to full sentences, as illustrated by the following two grammatically equivalent sentences taken from task T3 (again, with the PEOPLE-version above the THINGS-version and with differences marked by bold font):

“[...] important for the **manager** to have an overview of **employees** at all times.” [PEOPLE version]
 “[...] important for the **factory** to have an overview of **machines** at all times.” [THINGS version]

The only difference is that the PEOPLE-version involves people (*manager* and *employees*) whereas the THINGS-version involves things (*factory* and *machines*).

We also made sure the overarching scenarios were *semantically isomorphic*. In task T1 (see Figure 1), e.g., both the PEOPLE-version and THINGS-version deal with the overall scenario of parts-to-whole compatibility; i.e., whether or not personality types/spare parts are compatible. The other three tasks (T2, T3, and T4) can be found in appendix A.

Finally, for images, we made sure the two versions were *compositionally isomorphic*.¹ For instance, in the images in task T2 (see Figure 2), the semantic instigator of the action portrayed (the person for PEOPLE vs the industrial robot for THINGS) is positioned on the right-hand side of the image; the action (talking into a phone vs picking up a box) is just to the left, roughly occurring in the center of the image. The rest of the image, mostly on the left-hand side, is plain, irrelevant contextual background. The color schemes of the two images are also similar with a lot of warm, beige, and yellow/orange colors.

3.4 Participant Selection

Our recruitment strategy was to recruit high-school students from Danish high-schools in the Copenhagen area. We chose high-school students for our experiment because they represent potential future university students. After all, many promotional campaigns by universities target precisely this cohort. The Communication Department at our university facilitated contact with teachers and administrators from six high-schools.

We first explained the nature and intention of our experiment to the teachers responsible for distributing our questionnaire to their students, so that the teachers could judge whether or not they would be willing to have their students potentially participate in the experiment. However, we asked the teachers not to pass on this information to their students, so as to not interfere with the outcomes (see also Section 5.2).

3.5 Design

Each high-school student received an invitation to participate from their teacher. The students, of course, each voluntarily elected whether or not they were willing to participate in the experiment. Each student had to explicitly consent; i.e., they could opt in or opt out. The invitation contained a link to a SURVEYMONKEY online questionnaire service [40] tasked with organizing and presenting questions, collecting timestamped responses, and offering cookie-based protection against participants attempting to participate more than once. The questionnaire contained a welcome greeting, including a photo of our university. It also explained how to fill in the answers for the subsequent tasks. The subjects were asked for some relevant background information. Highly relevant to the study, the form contained a question about their self-reported gender identification, comprising three options: “*female*” and “*male*” as well

¹Due to potential copyright issues, the PEOPLE-version of the image is not the one originally used in the questionnaire, but one that is virtually equivalent.

Task 1: You get a description of a project you need to do

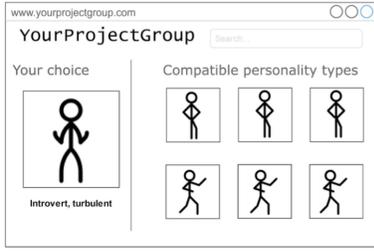
Imagine that you have been given sufficient instruction on programming websites. Your class will be divided into groups and your group will do a project.

You can choose between two projects:

Project A

You have to **program a website** called `yourprojectgroup.com`. The website should be used to find **personality types** that fit together when making teams. It should be possible to select a specific personality type and then see which other types fit with the selected type.

The mockup below is an example of what the website could look like:



Project B

You have to **program a website** called `yourspareparts.com`. The website should be used to find **spare parts** that fit together when repairing machines. It should be possible to select a specific spare part and then see which other parts fit with the selected part.

The mockup below is an example of what the website could look like:

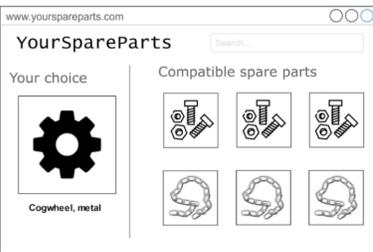
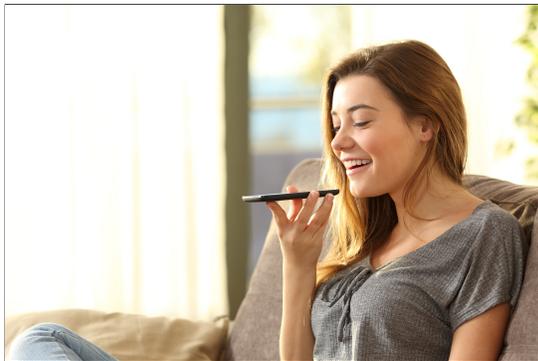


Figure 1: Task T1 (website project).



(a) PEOPLE-version.



(b) THINGS-version.

Figure 2: Pictures from Task T2 (news article).

as “other” which came with a free-text field. The subjects were asked to self-assess to what extent they had prior programming experience (“yes”, “little”, “none”, and “don’t know”). They were also asked which type of high-school they were enrolled in, so that we were able to filter out students not enrolled in a general education high-school². Finally, on the last page of the survey, participants were optionally given a chance to win a gift card for which they then had to voluntarily provide an email address.

²The Higher General Examination Programme (STX).

Table 2: The meaning of the scores: 1–5.

SCORE	EXPLANATION FOR SCORE
1	I would much rather do A
2	I would rather do A
3	I have no preference for neither A nor B
4	I would rather do B
5	I would much rather do B

For each of the four tasks (T1–T4), participants were asked to select their preference as to which of the two isomorphic versions of the task they would prefer to do (imagining that they hypothetically had to perform the task). Table 2 shows the possible answers (translated from Danish), categorized as numbers on a scale from one to five (1–5). A score of 1, for instance, means that the participant in question would give their preference as: “*I would much rather do A*” (where A would correspond to a particular version of the task; either PEOPLE or THINGS). The participants were asked to provide their score (preference) for each of the four tasks (T1–T4).

To mitigate the threat of a potential left-to-right reading bias (see also Section 5.2), we created two variants of the questionnaire. The variants would both alternate the left-to-right positioning of the PEOPLE- and THINGS-versions. One variant had the PEOPLE-version to the *left*³ in task T1, *right* in T2, *left* in T3, and to the *right* in T4. The other variant featured a horizontally mirrored layout (i.e., PEOPLE to the: *right*, *left*, *right*, and finally to the *left* in task T4). Half of the teachers were given one variant of the questionnaire; the other half of the teachers were given the other variant to forward to their students.

After execution of the experiment, but before statistical analysis, we eliminated the variants, by normalizing all data so that PEOPLE consistently appears to the *left* with a *strong preference* score of 1 and THINGS appears to the *right* with a *strong preference* score of 5. (The remainder of the paper consistently adopts this convention.)

3.6 Execution

Pilot experiments. To test our experimental setup, we conducted two pilot experiments physically on-location, as well as one pilot experiment virtually online. The two physical pilots were both conducted with three university students without programming experience from the University of Copenhagen.⁴ The virtual pilot was conducted with N=18 high-school students. The pilots led to clarifications on task instructions; in particular, emphasizing that the tasks were hypothetical in that participants were, in fact, *not* supposed to actually do the tasks, but merely pick the one that they imagined that they would prefer to do.

Adaptation to COVID-19. Our experiment was originally set to be carried out physically at the high-schools during March and April 2020. However, with Denmark entering national lockdown on March 13 (in response to the COVID-19 pandemic), we were forced to adapt from a physical to a virtual (online) experiment. We assumed that participating in an online survey (amidst a pandemic) would be a low priority for high-school students. And, since the literature reports that: “providing incentives for online quizzes increased participation” [27], we expanded the online pilot with a question on how likely students thought they would be to participate in the presence/absence of participation incentives; e.g., a chance to win a prize. Affirmative responses in the pilot led us to include an (optional) chance to win a 300 DKK⁵ gift card (see also Section 5.2).

Experiment. Four⁶ out of six high-schools approached agreed to participate and to forward the questionnaires to their students.

³Hence, the THINGS-version would be to the right in task T1.

⁴Note that: University of Copenhagen ≠ IT University of Copenhagen.

⁵300 DKK is approximately 40 EUR or 47 USD (as of October 2020).

⁶Nærum, Nørre G, Virum, and Ørestad Gymnasium.

The students were given three weeks to complete the questionnaire during the spring of 2020. The experiment ran with only one minor technical problem: In one of the surveys, the PEOPLE- and THINGS-versions in task T4 were erroneously identical when the questionnaire was sent out. The mistake was corrected within four hours of sending out the survey link to the teacher. In response to this mistake, we have eliminated all responses to task T4 from that questionnaire variant with timestamps within the affected time frame, but kept the responses for the other tasks (T1–T3), as they were unaffected by this problem. We have obviously also kept all responses from all tasks (T1–T4) from after the correction. Thus, the only interference caused by this mistake will be a slightly lower number of data points for task T4 (see below).

3.7 Analysis Method

In our study, each student is asked to rate four tasks using an ordinal scale with levels 1 to 5 (see Table 2). For the statistical analysis of the collected data, we use a cumulative logit model with a random intercept [2, 33], which is a standard model for multinomial data with an ordinal response variable and repeated measurements, such as ours. We note that categories of the ordinal scale 1 to 5 are merely labeled to indicate the order of categories and that their numerical values do not play a role in this kind of model; they could be renamed without impacting the results.

The statistical model describes, for each of the four tasks, the probability distribution of the score assigned by a specific student to the given task. This probability distribution consists of five probabilities, one for each of the five scores in the scale. The model is, however, formulated in terms of *log-odds*. For the sake of completeness and reproducibility, we here give a brief account of the cumulative logit model in the context of our study and refer to Agresti and Alan [2], and Larsen et al. [28] for details.

The scale has five categories. For any split between two neighboring categories j and $j + 1$, where j is one of $\{1, 2, 3, 4\}$, we can consider the binary decision of whether to choose one of the categories 1 to j (*in the direction of PEOPLE*) vs one of the higher categories $j + 1$ to 5 (*in the direction of THINGS*).

The inclination of the individual can then be quantified in terms of the *probability* of choosing a category in the direction of PEOPLE; more formally, as the *cumulative probability*: $P(Y \leq j)$. From these we can always find the probabilities for each of the five categories, so no information is lost by considering these cumulative probabilities instead.

Equivalently, we can use the four *odds* of choosing a category in the direction of PEOPLE, defined as the *cumulative odds*:

$$P(Y \leq j)/P(Y > j).$$

These odds describe how many times *more likely* the person is to respond among categories in the direction of PEOPLE than in the direction of THINGS.

Note that the binary decisions created by splits are not indicative of the preference per se; in fact, when we split as 1–4 vs 5, then choosing a category in the direction of PEOPLE also includes a slight preference for THINGS option. Of specific interest are perhaps the odds of stating an *explicit preference* for PEOPLE (score categories 1–2) as compared to a response of indifference or an explicit preference for THINGS (score categories 3–5).

An effect of gender, say, may then be quantified by an *odds ratio*, which expresses how many times higher the odds of leaning towards PEOPLE are for *women* than they are for *men*.

The cumulative logit model describes each of the four cumulative odds (at log scale⁷), i.e., the log-odds of rating between 1 and j :

$$\log\{P(Y \leq j)/P(Y > j)\} = \alpha_j + \beta_{\text{woman}} + \beta_{\text{inexperienced}} + U_s \quad (1)$$

Here, α_j is a *baseline* level of the log-odds (in fact, the log-odds for an experienced man), β_{woman} is the effect of being a woman rather than a man, and $\beta_{\text{inexperience}}$ is the effect of being inexperienced rather than experienced with programming. The term U_s is a student-specific component (*random effect*), which follows a Gaussian distribution with mean 0 and a variance estimated from data.

The student-specific component allows some students to naturally have a stronger preference than other students with the same combination of gender and experience – think of randomly sampling an *individual* level from a Gaussian distribution that is then added to the log-odds for that specific student: if the individual level is high, then the student has high odds (and probability) of rating in the direction of PEOPLE compared to students of same gender and experience. The random effect can also be seen as modelling the dependence between tasks regarding a person’s preferences.

We develop and motivate the structure of our model in Section 4 below, but further to this we note that extensive statistical model checking has been carried out to verify that the model adequately describes the data collected in our study.

In particular, we note that data supports the assumption of *proportional odds* implicit in the right hand side of (1).

Recalling that (1) is, in essence, a specification for each of the four cumulative odds of rating in the range 1 to j , the model is complex and can be hard to summarise. However, the proportional odds property offers a simple interpretation of the model: Clearly the probability (and thus the odds) of giving a score in a broad range (e.g., 1–4) is higher than the probability of giving a score in a narrow range (e.g., 1–2). However, the *effects of gender and programming experience—the odds ratios—are the same regardless of which of the four cumulative odds we consider*.

All significance tests are carried out in terms of likelihood ratio tests to give the most reliable results; in particular, all reported p -values are from such tests. For the analysis, we have used the software implementation available in R through the ordinal package [15].

4 RESULTS

In Section 4.1 below we give an overview of the collected data. Then, in Section 4.2, we establish the statistical model and investigate the effects of gender and programming experience. Finally, Section 4.3 gives a detailed overview of the distribution of student ratings.

4.1 Collected Data on Task Preference

Number of responses. Our recruitment strategy secured $N=558$ participants from the four high-schools that agreed to participate. There were 70 participants, who entered only background information without rating any of the four tasks. Our analysis is based

Table 3: The total number of responses for the different parts of the questionnaire.

	BACKGROUND INFORMATION	TASK T1	TASK T2	TASK T3	TASK T4
N =	558	488	463	441	374

on the remaining $N=488$ respondents, who rated at least one of the four tasks. The number of responses to each of the four tasks is detailed in Table 3. From the first task (T1) through the subsequent tasks, the number of responses drops, as students progressively give up on responding to the questionnaire. The somewhat larger drop for task T4 is due to the technical problem mentioned earlier (cf. Section 3.6). Importantly, the statistical framework used for our analysis does not require complete data from each respondent.

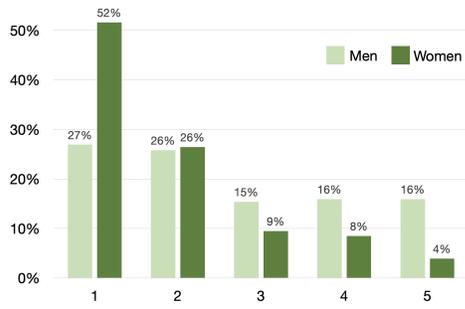
Gender. Among the 488 students included in our study, 63% (306) were women and 37% (182) were men. The percentage is representative of Danish high-schools, where the national gender distribution of high-school intake in 2019 was 61% women and 39% men, according to Statistics Denmark [38]. The experiment did not assume any specific definition or concept of gender, and the distinction we used is based purely on the self-assessment from the respondents. No participants entered a non-binary gender. All students came from the common academic high-school in Denmark.

Figure 3 shows the distribution of response scores from 1 (strong PEOPLE preference) to 5 (strong THINGS preference) for each of the four tasks (T1–T4). Responses from women are shown in dark green vs light green for those of the men. Histograms are shown to the left. To the right, we show a smoothed visualization to highlight the overall trend: responses from women are further left, towards PEOPLE, than those of the men.

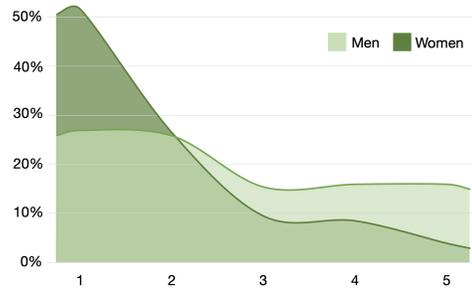
In Figure 3a, we see that more than half of the women (52%) strongly prefer the PEOPLE-version (score 1), whereas this was only the case for about a fourth of the men (27%). On the THINGS end of the spectrum (score 5), we see that almost none of the women (only 4%) strongly prefer the THINGS-version, whereas this was the case for about a sixth of the men (16%). In general, we see a tendency for the women to gravitate towards the PEOPLE end of the spectrum (left side), compared to the men whose responses appear to fall more uniformly on the preference scale. (This is most perceptible on the smoothed visualization; the right column of Figure 3.) We see this pattern for all tasks, albeit more pronounced for Task 1. We note also that the shape of the histograms seems to vary across tasks, so some task-dependent variation in preferences is to be expected.

To illustrate how we may summarise and quantify systematic differences in the distribution of responses between groups of people, consider again the histogram for Task 1 in Figure 3a. The (empirical) odds of a woman rating the task with a score of 1 is $52\% / (26\% + 9\% + 8\% + 4\%) = 1.07$, whereas for a man the odds are $27\% / (26\% + 15\% + 16\% + 16\%) = 0.368$. The odds of a woman rating the task with a score of 1 are thus $1.07/0.368 = 2.9$ times higher than those of a man; the (empirical) *odds ratio* quantifying

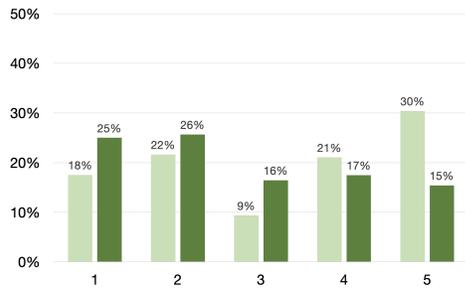
⁷Multiplicity of odds correspond to additivity at log-odds scale.



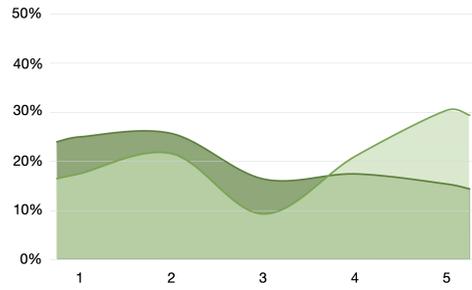
(a) Histogram (T1).



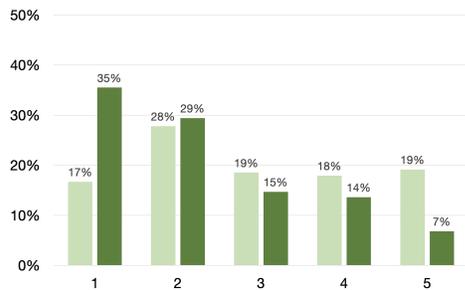
(b) Smoothed visualization (T1).



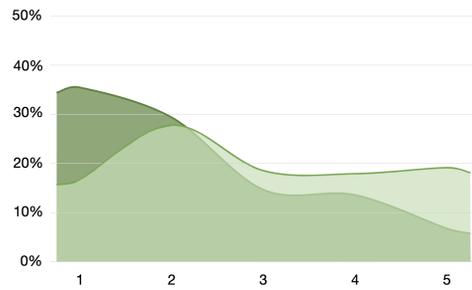
(c) Histogram (T2).



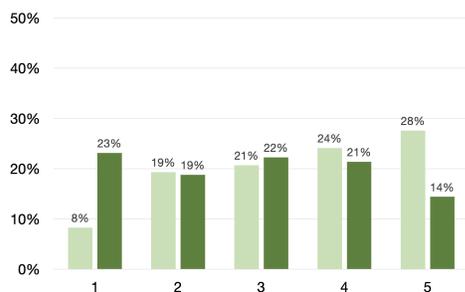
(d) Smoothed visualization (T2).



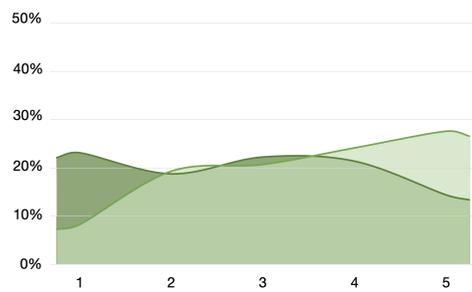
(e) Histogram (T3).



(f) Smoothed visualization (T3).



(g) Histogram (T4).



(h) Smoothed visualization (T4).

Figure 3: Distribution of responses for women (dark green) vs men (light green) for each of the four tasks on a score from 1 (strong preference for the PEOPLE-version) to .. 3 (indifference) .. to a score of 5 (strong preference for the THINGS-version). Histograms are shown to the left. To the right, we show a visualization that superimposes the answers from women and men in semi-transparent color and connects smoothed lines between the score categories. (Note that this visualization serves purely to illustrate the trend that: women answer left of the men.)

Table 4: Self-reported levels of prior programming experience for women vs men.

	PRIOR PROGRAMMING EXPERIENCE:			
	NONE	LITTLE	YES	DON'T KNOW
Women	72% (220)	21% (65)	4% (11)	3% (10)
Men	49% (89)	40% (73)	8% (15)	3% (5)

this gender effect is 2.9. Now, since we have a 5-point scale, we may also consider other odds. Take, for instance, the odds of scoring the task by either 1 or 2 (preference for PEOPLE) vs 3, 4, or 5 (indifference or preference for THINGS). For women these odds are $(52\% + 26\%) / (9\% + 8\% + 4\%) = 3.57$ whereas for men they are only $(27\% + 26\%) / (15\% + 16\% + 16\%) = 1.12$; this gives an odds ratio of 3.2, i.e., a gender effect similar to before. The odds themselves are higher when we consider the odds of giving scores 1 or 2, rather than the odds of giving a score of 1, but the odds ratio stays roughly the same. This property is referred to as *proportional odds*; more elaborate considerations along the lines of this example suggest that the proportional odds model, described formally in Section 3.7 is suitable for our data.

Programming experience. Table 4 shows the data on self-reported prior programming (in)experience for women vs men. As expected, we see that fewer women than men in high-school have prior experience with programming. Almost three-quarters of the women report no prior programming experience, whereas this was only the case for about half of the men. Of course, it has to be taken into account that men are more inclined than women to overestimate their own competence [6, 8, 16].

In the statistical analyses below, we simplify the self-reported levels of experience into only two categories: *No programming experience* (the leftmost “none” column of Table 4) vs *programming experience* (the three rightmost columns of Table 4, lumped together). We do so, because it is unclear whether “yes” and “little” programming experience, as self-assessed by high-school students, would amount to substantially different levels of experience. Further, we include the responses of “don’t know”, with the motivation that a student not explicitly reporting that they have no experience (“none”) would seem likely to have had some previous encounter with programming. An overview of how respondents are distributed among the four groups corresponding to combinations of gender and programming experience is seen in Table 5.

As for the effect of gender, we may get an idea of the difference in preferences between students with no prior experience with programming and those with experience by comparing various odds. For example, from the distributions of scores shown in Figure 3a we can conclude that the odds of an inexperienced student rating Task T1 by 1 or 2 (i.e., showing a preference for PEOPLE) are 2.68, whereas those of an experienced student are only 1.59; this gives an odds ratio of 1.7. Overall, data clearly suggests that the preferences in version of a task is dependent on prior experience with programming.

Table 5: Number of respondents according to gender and programming experience.

	INEXPERIENCE	EXPERIENCE	TOTAL
Women	220	86	306
Men	89	93	182
Total	309	179	448

It is thus paramount to *control* for the effect of programming inexperience in quantifying the effect of gender (and vice versa). To illustrate the care that needs to be taken, note from Table 4 that there is a markedly higher percentage of inexperienced students among the women; combined with a strong preference for PEOPLE among inexperienced students, this could imply a higher preference for PEOPLE among women even without any direct effect of gender. We, of course, need to separate the two effects appropriately, and do so by standard statistical machinery.

4.2 Statistical Analysis of Task Preferences

Motivated by investigations outlined in Section 4.1 and the statistical framework of Section 3.7, we propose a statistical model for the student rating of a given task and use this to address further the main questions of this paper. The model is a cumulative link mixed model, which describes the odds of rating *in the direction of PEOPLE*⁸ in terms of the respondent’s gender (female/male), programming experience (no/yes), and the given task (T1–T4).

Our statistical model takes into account a significant heterogeneity among students ($p < 2.2 \times 10^{-16}$), which is not attributable to either their gender or their programming experience, and that their preferences (as quantified by the probability of rating the task in the direction of PEOPLE) may thus generally be at the higher or at the lower end for all four tasks. We may think of it in terms of the model capturing personal preferences as a variation of the probability of rating in the direction of the PEOPLE version around some overall probability within each combination of gender, level of experience, and task considered.⁹

In our interpretation of the model, we shall focus on the odds of stating an explicit preference for the PEOPLE version (i.e., rating the task by score 1 or 2) vs indifference or preference for THINGS (i.e., rating the task by scores 3–5). Note, however, that the proportional odds property means that the effects we see are the same regardless of which “binary” decision is considered. So, for instance, whether we consider the odds of rating the task by a score in 1–2 vs a score in 3–5 (i.e., the odds of an explicit preference for PEOPLE) or the odds of rating the task by a score in 1–4 vs a score of 5, we see the same odds ratio quantify how the odds change between genders, levels of programming experience, or tasks.

Because of the Gaussian variation in the student preferences at log-odds scale, any odds (and associated probabilities) we consider

⁸Given a threshold of score j , a rating in the range 1 through j is considered *in the direction of PEOPLE*, while a rating in the range $j + 1$ through 5 is considered *in the direction of THINGS* (see Section 3.7).

⁹as implied by the assumed Gaussian variation at the log-odds scale.

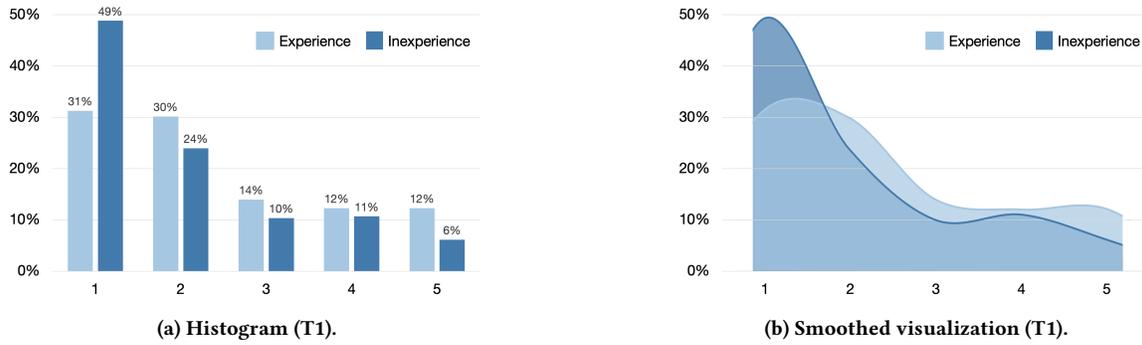


Figure 4: Distribution of responses for inexperienced students (dark blue) vs experienced students (light blue) for task T1 on a score from 1 (strong preference for PEOPLE) to 5 (strong preference for THINGS). The smoothed visualization on the right serves purely to illustrate that inexperienced students answer to the left of experienced students.

are, in fact, random. This, in turn, means that also the odds ratio formed by comparing two specific students will vary between pairs of students. We report the median odds ratio and further indicate the variation in odds ratios by an interval based on upper and lower 10th percentiles comprising the middle 80% of the odds ratios that would be obtained by comparing two random students (see Larsen et al. [28], who introduce this concept of an *Interval Odds Ratio*).

OBSERVATION 1: Women have a stronger preference for PEOPLE than men do. The median odds ratio obtained by comparing two random students with the same level of programming experience—one female and one male—is 2.7 (Table 6, top row). This means that the odds of the woman rating in the direction of PEOPLE are 2.7 times higher than the odds of the man rating in the direction of PEOPLE.

The effect is strongly significant ($p = 3.9 \times 10^{-13}$), as also apparent from its confidence interval 2.1 to 3.5, which indicates the uncertainty about the estimate and only contains values well away from 1 (corresponding to equal odds for the two genders).

Although it may seem from Section 4.1 that the effect of gender varies between tasks, we note that this variation is, in fact, *not* statistically significant ($p = 0.12$). We have investigated whether the effect of gender is different for inexperienced and experienced students, but, again, the data does *not* suggest this to be the case ($p = 0.24$).

To understand the impact of student heterogeneity, consider the fact that 80% of odds ratios comparing a woman to a man would be in the range 0.48 to 15. In particular, this implies that—while the overall preferences for PEOPLE is indisputably stronger among the women—certainly we may experience pairs of students where the male exhibits a stronger preference than the woman.

OBSERVATION 2: Students without programming experience have a stronger preference for PEOPLE than students with programming experience have. The median odds ratio quantifying the effect of programming inexperience is 1.4, and compares the odds for a student *without* programming experience to the odds for a student of the same gender, but *with* prior programming experience (Table 6, bottom row). The effect is significant (7.8×10^{-3});

Table 6: Effects of gender and programming inexperience, as quantified by the median odds ratio that would be obtained by comparing two students in terms of their odds of expressing a preference for PEOPLE. Also shown, a 95% confidence interval (CI) indicating the range of odds ratios supported by data, and a p -value for the test of no effect (i.e., an odds ratio of 1).

	ODDS RATIO	95% CI	P-VALUE
Female vs male	2.7	2.1 – 3.5	3.9×10^{-13} ***
Inexperience vs experience	1.4	1.1 – 1.9	7.8×10^{-3} **

this is reflected also in the confidence interval of 1.1 to 1.4, which indicates the range of odds ratios supported by our study and notably does not cover the value 1 (equal odds). The effect of programming inexperience does not appear to vary between tasks ($p = 0.39$). The individual variation in student preferences imply that 80% of the odds ratios, we may observe in practice, fall in the range 0.25 to 8.1.

Note, that while clearly an important factor in understanding the student preferences, the precise effects of programming experience may depend somewhat on how the experience categories are defined. We do not dive further into this as it is beyond the scope of this paper.

What about women with no programming experience? Since female students exhibit a higher preference for PEOPLE than male students do, and also inexperienced students exhibit higher odds than experienced students do, we may rightly ask how *inexperienced female students* compare to *experienced male students*. This question is particularly interesting from a recruitment perspective, not least noting from Tables 4 & 5 that this group contains 72% of the female respondents (45% of the total number of respondents).

The odds ratio comparing two students who differ in *both gender and experience* is obtained by multiplying the two effects for gender and experience in Table 6; this gives an odds ratio of $2.7 \times 1.4 = 3.8$, which is an even more pronounced discrepancy in preferences.

That is, women without experience have 3.8 times higher odds of preferring PEOPLE tasks than men with programming experience.

OBSERVATION 3: Preferences depend on the task at hand. We have seen that the effects of gender and programming experience both do not depend on the task at hand. Now we turn to investigating the effect of the individual tasks on the student preference, as quantified again by how the odds of a preference for the PEOPLE version changes when a different task is considered. Overall, there is a significant difference between tasks in their effect on student preference ($p < 2.2 \times 10^{-16}$).

If we look into scores given in Task T1 (see e.g., Figure 3), students appear to have a more pronounced preference for PEOPLE version than is seen for Tasks T2-T3. Indeed, considering the ratings that a specific student gives to each of the four tasks, the odds of the student rating in the direction of PEOPLE for Task T2 are only between 24% and 39% of the odds found for Task T1. Similarly, the odds found for Task T3 are between 42% and 69% of those for Task T1, while the odds found for Task T4 are only between 19% and 32% of the odds for Task T1.¹⁰

4.3 Distribution of Student Ratings

We have up until now considered the ratio of cumulative odds so as to quantify the *change* in odds between groups of students, but this gives no indication of the absolute size of the odds; is a student more likely to rate in the direction of PEOPLE than not? The statistical model does, in fact, also specify the probability of a student rating a task by a certain score. From these five probabilities for scores 1–5, we may readily compute the probabilities of *preferring the PEOPLE version* (scores 1–2), being indifferent (score 3), and *preferring the THINGS version* (scores 4–5). Figure 5 is a visualisation of the probabilities implied by the model. The probabilities can be seen as a "statistically smoothed" version of the frequencies in the histograms of Figure 3; a major advantage of looking at data through a statistical model rather than by raw frequencies is that it helps distinguishing noise from systematic patterns. Estimated probabilities are shown in Figure 5 for three different people: A median student (bold solid line), one in the 10th percentile (low preference for PEOPLE, dashed line), and one in the 90th percentile (high preference for PEOPLE, dotted line).

Consider first the median student rating Task 2 as shown in Figure 5 (bold solid line): For an inexperienced woman (top panel), the probability of preferring the PEOPLE version is noticeably higher than the probability of preferring the THINGS version. For the experienced male (bottom panel) the pattern is the opposite; the probability of him preferring the THINGS version is noticeably higher than the probability of preferring the PEOPLE version. Both students have about a 20% chance of being indifferent.

Now look at the two extremes for the inexperienced women (top panel): While the student at the 90th percentile (dotted line) exhibits an even stronger preference for PEOPLE, the student at the 10th percentile actually has a tendency to prefer the THINGS version.

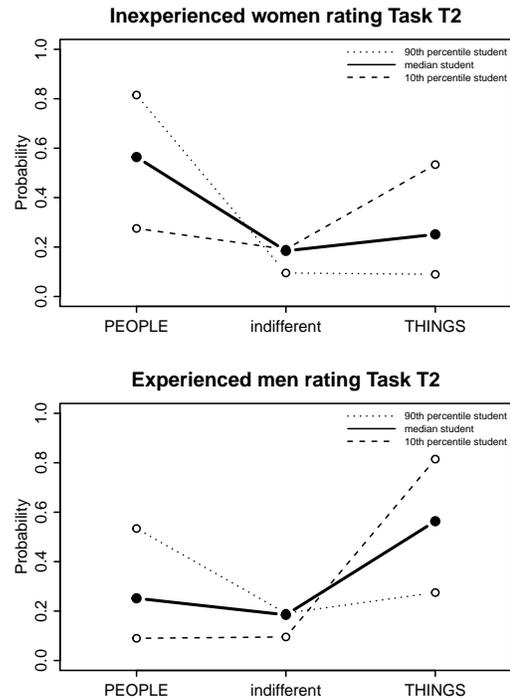


Figure 5: Probabilities of stating an explicit preference for PEOPLE (scores 1-2), indifference (score 3) or an explicit preference for THINGS (scores 4-5). Shown are the probabilities for a median student (solid), and an indication of the middle 80% of students via the probabilities of students in the upper (dotted) and lower (dashed) 10th percentiles.

The combined odds ratio of 3.8 comparing an inexperienced woman to an experienced man is reflected also in Figure 5; indeed, the probability of a (median) inexperienced woman preferring PEOPLE version of Task T2 is 0.564 where for the (median) experienced man this probability is 0.252, and so their odds are $0.564 / (1 - 0.564) = 1.29$ and $0.252 / (1 - 0.252) = 0.336$, respectively. This gives exactly the odds ratio of $1.29 / 0.336 = 3.8$ reported. Comparing the two 10th percentile students of opposite gender or the two 90th percentile students would give the exact same odds ratio.

Figure 6 gives the complete overview of how the student preferences change according to gender, programming experience, and the task considered.

In all four tasks, the median woman—whether experienced or not—exhibits a preference (highest probability) for the PEOPLE version (see columns 1&2 of Figure 6). The males, in comparison, exhibits a less consistent preference (see columns 3&4 in Figure 6): The median man—experienced or not—leans towards the PEOPLE version for Task T1, but towards a the THINGS version in Tasks T2 and T4. For Task T3, the inexperienced median man prefers the PEOPLE version, albeit less pronounced than all of the women, whereas experienced men express a slight preference for THINGS. We note that the probabilities of a student being indifferent is generally low.

¹⁰The ranges stated are the 95% confidence intervals for the three odds ratios comparing each of Tasks T2-T4 to Task T1.

Figure 6 reflects also the effects we have already seen regarding the *relative strength* of preferences: Women are *more* likely to prefer PEOPLE than are the men, students with no programming experience are *more* likely to prefer the PEOPLE version than are the experienced students, and the preference for PEOPLE is more pronounced in Task T1 than in the other three tasks. Indeed, all the odds ratios of Table 6 may be computed directly from the probabilities in Figure 6.

We refer to Figure 10 in Appendix B for a detailed view of how responses are distributed across ratings 1 through 5, rather than across the three-point scale (PEOPLE, indifference, THINGS) considered in Figure 6.

5 THREATS TO VALIDITY

We now consider the validity of our observational study in terms of *construct*, *internal*, and *external* validity.

5.1 Construct Validity

Establishing gender? We asked participants to self-assess as being of a particular gender; we included the options other (with a free text field), female, and male. The notion of the gender we have used may have been too simplistic.

Measuring preference? We quantify the preference of an individual student on a five-item Likert scale, where one is a stronger preference for the PEOPLE-version and five is a stronger preference for the THINGS-version. As this study measures hypothetical preference, it was important to measure a lack of preference, as well as the difference between weak and strong preference for a certain theme. The five-item scale allows us to do this.

Measuring experience? In the questionnaire given to the students, we asked the question “*do you have programming experience?*”. Here, we added options to answer “little” and “don’t know”, in addition to “yes” and “none”, to get a more nuanced picture of their level. For the analysis, however, we have grouped answers in two categories: *programming inexperience* and *programming experience*. Of course, a programming test would have been a more objective way of measuring experience, but that would have taken more time for the participants.

5.2 Internal Threats

Subjects not representative of intended population? We believe that the participants recruited were indeed representative of the intended population: We deliberately recruited high-school students as proxies for potential future university student population. About three-fifths were women and two-fifths were men which coincides with the gender composition of Danish high schools [44].

Bias introduced by imperfect response rate? Because of the COVID-19 lockdown, the questionnaire was distributed indirectly via teachers who were not able to physically meet with their students. For this reason, some teachers distributed the questionnaire via email, others via online school platforms. We thus unfortunately cannot infer exactly how many percent decided (not) to participate. Our best defense against this threat is the sheer number of respondents (about 500) along with their variation in programming experience level (see also 3.4).

Subjects deliberately interfering with study outcome? As for all studies, respondents may attempt to influence the outcome

of a particular study. This risk is heightened for “sensitive topics” such as gender issues. However, we generally mitigate this threat by not revealing that our study was, in fact, about gender. The high-school teachers, who forwarded the questionnaires to the students, obviously wanted to know what this study was about. However, we asked them to not reveal the nature of the study to their students. Of course, some students may have guessed what the study was about, but we further obfuscated the questionnaire by alternating, for each task, the left/right positioning of PEOPLE and THINGS. Aside from this, technically, SURVEYMONKEY uses cookies to attempt to rule out respondents issuing multiple responses. We believe this threat is unlikely to compromise the validity of our study.

Gender question induces subject self-stereotyping? Research has established that experimental subjects may self-stereotype if their identity is salient in the context [37]. Stereotypes are amplified by subjects responding to questions about their gender; e.g., women score lower on math and higher on language, *after* being asked to identify as females on a test.

To avoid this threat completely, we ought to have asked the gender question at the end rather than at the beginning of the questionnaire. However, it would have meant that we only had usable data from participants who made it to the end of the questionnaire. The considerations about not revealing the nature of our study apply here as well (see previous threat). This threat ought to diminish if the questionnaire is not perceived, by the subjects, to have anything to do with gender; after all, the PEOPLE–THINGS dimension is not so well known.

Bias from participation incentives? With the chance to win a prize comes the threat that opportunistically entrepreneurial students may be tempted to fill in the questionnaire very fast with bogus responses just to secure a chance to win a prize. Some participants did complete the questionnaire fast, but since they could just as well be fast readers and decision-makers, we did not attempt to classify and filter out “fast students.” Besides, the questionnaire completion times distribute continuously with no discernible temporal gap.

Images biased towards one gender? The image in task T2 (see Figure 2a) coincidentally features a woman. In hindsight, it would have been better to have used both a woman and a man on the image. Again, we do not believe this to have a major effect on the validity of the results specifically relating to task T2.

Bias from left-to-right reading? We mitigated the left-to-right reading bias by creating two left/right mirrored variants of the questionnaire, as explained in Section 3.5.

5.3 External Threats

Beyond Computing? We hypothesize that our results generalize to STEM; including physics and mathematics. Many of the PEOPLE–THINGS considerations ought to apply just as much to the other STEM topics. However, more investigation is warranted.

Beyond Danish students? We expect our results to generalize to other western societies. However, the results may not necessarily generalize to regions of the world where other socio-economic factors apply. In particular, to cultures where financial stability and/or family expectations may stronger impact vocational choices than individual preferences.

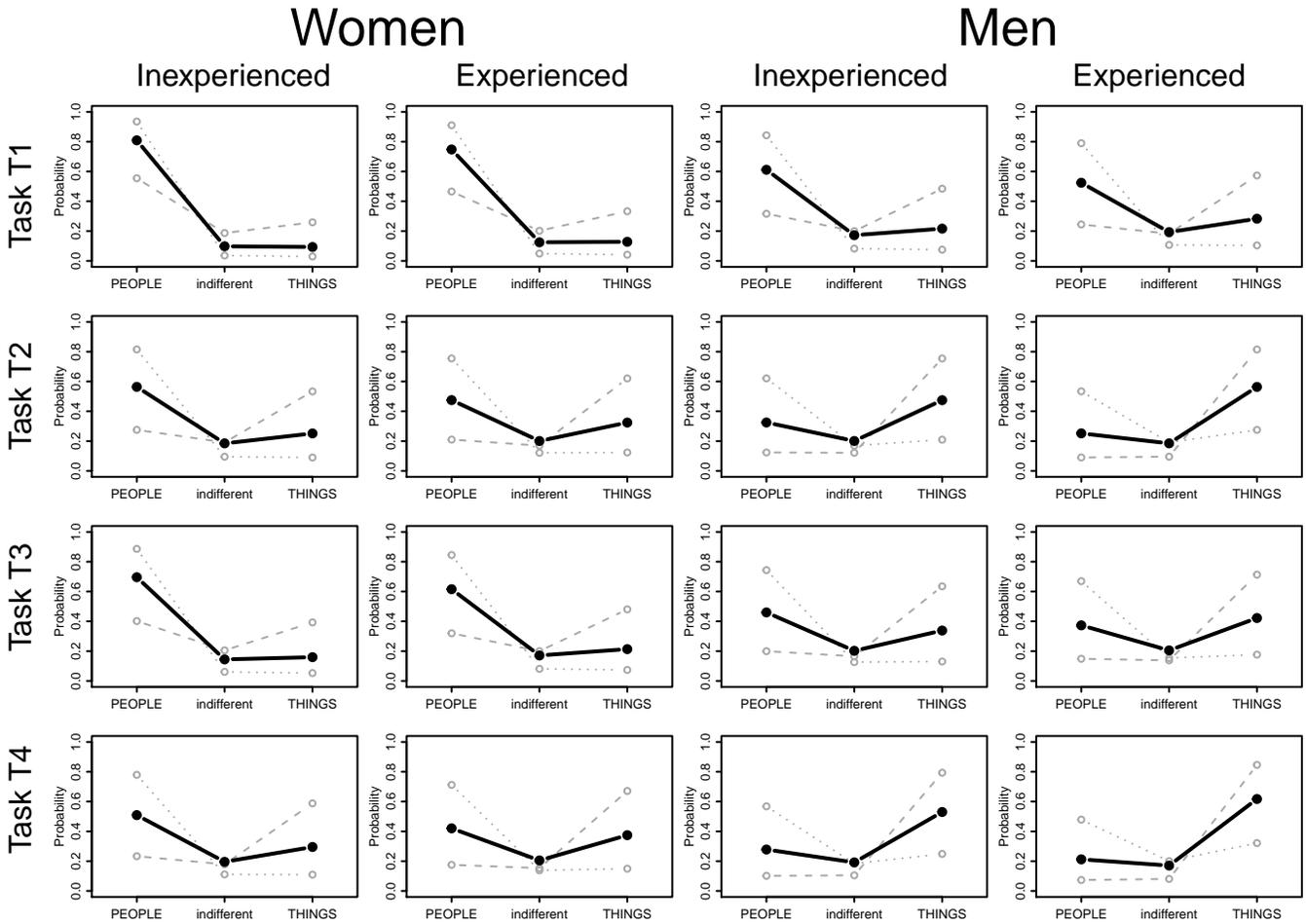


Figure 6: Probabilities of stating an *explicit preference for PEOPLE* (scores 1-2), *indifference* (score 3) or an *explicit preference for THINGS* (scores 4-5). Shown are the probabilities for a median student (solid lines), and an indication of the middle 80% of students via the probabilities of students in the upper (dotted) and lower (dashed) 10th percentiles.

Beyond high-school students? The focus of this study was to investigate high-school students as a proxy for potential future university students. Whether the results generalize beyond high-school students is beyond the objectives and scope of this study. The correlation with programming inexperience leads us to believe that the importance of favoring educational themes involving PEOPLE might diminish as students get exposure to computing and acquire experience with programming. Presumably, technology-trained students will pay more attention to the technology itself and how it is made than to the scenarios/themes of the educational activities. However, this is speculative; further investigation is needed.

Beyond hypothetical activities? In a study done later the same year [31], we tested whether our results would generalize to real (non-hypothetical) tasks; i.e., tasks that the students actually have to carry out rather than just imagine that they would carry out. We found very similar results to those in this study, although the participants in the later study were university students, rather than high-school students.

Beyond mere preference? This study was about the correlation between gender (and experience) and preferences according to the PEOPLE–THINGS dimension. We do not know if the results extend also to performance; i.e., whether women and inexperienced programmers perform “better” on educational activities with scenarios involving PEOPLE than THINGS. We intend to find out.

6 RELATED WORK

The relation between gender and vocational interests has been discussed extensively in the literature [4, 11, 19, 24, 26, 34–36, 41]. Similarly to Lippa’s work, some studies connected this with a broader idea of the connection between gender and personality traits [17, 20, 45]. It is important to note that our study does not presume this connection and tests the correlation between gender and the interest in the PEOPLE–THINGS dimension directly, similarly to Su et al. [39]. Apart from the study of the general correlation, many studies emphasize and explore the imbalanced gender distribution within STEM fields, especially mathematics and engineering

[7, 12, 14, 18, 25]. Some of the later papers relate directly to computer science [9, 32, 43] or discuss the correlation in the context of high-schools [42].

The key difference between these studies and our approach is that they did not focus on themes for activities, but rather on the general characteristics of STEM disciplines, their social history, or public perception [13].

To the best of our knowledge, this is the first study that focuses on gender differences on the PEOPLE-THINGS dimension specifically in teaching/learning activities within computing education.

The results of our experiment are to a large extent compliant with more general studies. Lippa [30] summarized data from two meta-analyses and three cross-cultural studies on gender differences in personality and interests. He found large gender differences on the PEOPLE-THINGS dimension of interest, with women more PEOPLE-oriented and less THINGS-oriented than men.

7 CONCLUSION

Our study aimed to see if Lippa’s general findings extrapolate to the specific context of computing education (for contemporary Danish high-school students).

One crucial caveat that we should keep in mind is that the tasks the respondents were choosing were only hypothetical. This methodological choice was a necessity as the majority of the respondents did not have the experience needed to perform the tasks but it has a significant impact on the implications of our results. The only thing that the experiment measured was the appeal of the tasks. For this reason, whenever we say that the respondents “chose” or “preferred” a given version of the task, it should be understood in a sense that they found it more appealing. The upshot of this is that our results can be most usefully applied as a part of a recruitment strategy where examples of tasks are often used to give the future candidates a “taste” of what they can expect to find during the studies. In addition to this, our results clearly show that to attract more women to computing it is important to emphasize a PEOPLE-oriented image of computing.

As can be seen in Section 4, the results are to a large extent compliant with the general trend discovered by previous authors as there is a clear tendency for women to favor PEOPLE-themed tasks—the odds of women preferring the PEOPLE-themed versions are 2.7 times higher than those of a man.

As can be seen from our results the correlation varies across different tasks. This is especially visible in the case of Task 1 in which the THINGS-oriented version was perceived as less appealing by the majority of students, regardless of gender. It is difficult to say why this particular task elicited these reactions. One possibility is that it could have been hard to discern the illustration of the spare parts.

An additional factor that played no role in previous studies due to their general nature is the level of experience of the respondents. Our study shows a visible correlation between the self-assessed programming experience and preference on the PEOPLE-THINGS scale. Inexperienced respondents found PEOPLE-themed examples more appealing. What is more, since the level of experience and

gender are independent factors the multiplied odds make inexperienced women 3.8 times more likely to see PEOPLE-themed tasks as more appealing.

In contrast to previous results, our experiment establishes that men’s preference towards THINGS-themed computing educational activities is actually much lower than what would be predicted from the broad trends discovered in the previous literature. (This is possibly due to computing’s historical reputation of dealing with computers and gadgets, inclining towards the THINGS end of the spectrum?) A strong preference for THINGS-oriented tasks was visible only in the case of experienced men. This correlation needs further study as its causes are far from clear. One possible suggestion (that needs to be further examined) is that it is an effect of the materials the students were exposed to earlier. It is possible that they have become “conditioned” with THINGS-oriented examples and exercises? Another possibility is that they perceive THINGS-oriented examples as more “advanced” or “professional?” Combined with the general preference for PEOPLE-oriented tasks visible amongst respondents lacking programming experience (regardless of their gender).

It goes without saying that the correlation we studied is just one of the factors that contribute to the current landscape of computing. We believe that the intervention we suggest can help to bridge the gender gap but it should be treated as orthogonal to many other relevant initiatives, such as fighting stereotypes, increasing the number of female role models, creating inclusive environments, and favoring collaboration over competition.

Crucially, the only aspect tested in our experiment was the effect of the (superficial) thematic incarnation of educational activities (as opposed to underlying characteristics of the computing discipline itself or specific individual personality traits). This is good news for those interested in diminishing the gender gap in computing degrees (possibly for STEM degrees, in general): High-school computing educators should, whenever possible, favor PEOPLE-themes in educational activities. Also, we encourage communication departments to emphasize the relation to and relevance for PEOPLE in the recruitment of potential university candidates.

All of this will, in general, benefit women and students without prior programming experience; thereby, computing education, and, on a longer time-scale: the area of computer science and software engineering. We hope this will contribute to improving the gender imbalance.

ACKNOWLEDGMENTS

We thank the high-school teachers for agreeing to forward the questionnaires to their students; and, of course, the many high-school students for taking their time to participate in this experiment and respond to our questionnaire. We also thank the Communication Department at the IT University of Copenhagen for facilitating contact with the high-schools. We thank Michael Caspersen for providing feedback on a preliminary version of this work.

REFERENCES

- [1] Janet Abbate. 2012. *Recoding gender: Women’s changing participation in computing*. Mit Press, Cambridge, MA, USA.
- [2] Alan Agresti. 2013. *Categorical data analysis* (3rd ed.). John Wiley & Sons, USA.
- [3] Armbrrecht, Arwen. 2015. Where are the missing women in tech? <https://www.weforum.org/agenda/2015/12/where-are-the-missing-women-in-tech/>

- [4] Jesse R Aros, George A Henly, and Nicholas T Curtis. 1998. Occupational sextype and sex differences in vocational preference-measured interest relationships. *Journal of Vocational Behavior* 53, 2 (1998), 227–242.
- [5] Catherine Ashcraft, Brad McLain, and Elizabeth Eger. 2016. *Women in tech: The facts*. National Center for Women & Technology (NCWIT).
- [6] Brad M Barber and Terrance Odean. 2001. Boys will be boys: Gender, overconfidence, and common stock investment. *The quarterly journal of economics* 116, 1 (2001), 261–292.
- [7] Camilla Persson Benbow, David Lubinski, Daniel L Shea, and Hossain Eftekhari-Sanjani. 2000. Sex differences in mathematical reasoning ability at age 13: Their status 20 years later. *Psychological science* 11, 6 (2000), 474–480.
- [8] Claes Bengtsson, Mats Persson, and Peter Willenhag. 2005. Gender and overconfidence. *Economics letters* 86, 2 (2005), 199–203.
- [9] Sylvia Beyer. 2014. Why are women underrepresented in Computer Science? Gender differences in stereotypes, self-efficacy, values, and interests and predictors of future CS course-taking and grades. *Computer Science Education* 24, 2-3 (2014), 153–192.
- [10] Philip Brey. 2010. Values in technology and disclosive computer ethics. *The Cambridge handbook of information and computer ethics* 4 (2010), 41–58.
- [11] Kingsley R Browne. 2006. Evolved sex differences and occupational segregation. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior* 27, 2 (2006), 143–162.
- [12] Sapna Cheryan. 2012. Understanding the paradox in math-related fields: Why do some gender gaps remain while others do not? *Sex Roles* 66, 3-4 (2012), 184–190.
- [13] Sapna Cheryan, Victoria C Plaut, Caitlin Handron, and Lauren Hudson. 2013. The stereotypical computer scientist: Gendered media representations as a barrier to employment for women. *Sex roles* 69, 1-2 (2013), 58–71.
- [14] Sapna Cheryan, Sianna A Ziegler, Amanda K Montoya, and Lily Jiang. 2017. Why are some STEM fields more gender balanced than others? *Psychological bulletin* 143, 1 (2017), 1.
- [15] R. H. B. Christensen. 2019. ordinal—Regression Models for Ordinal Data. R package version 2019.12-10. <https://CRAN.R-project.org/package=ordinal>.
- [16] Shelley J Correll. 2001. Gender and the career choice process: The role of biased self-assessments. *American journal of Sociology* 106, 6 (2001), 1691–1730.
- [17] Paul T Costa Jr, Antonio Terracciano, and Robert R McCrae. 2001. Gender differences in personality traits across cultures: robust and surprising findings. *Journal of personality and social psychology* 81, 2 (2001), 322.
- [18] Amanda B Diekman, Elizabeth R Brown, Amanda M Johnston, and Emily K Clark. 2010. Seeking congruity between goals and roles: A new look at why women opt out of science, technology, engineering, and mathematics careers. *Psychological science* 21, 8 (2010), 1051–1057.
- [19] Amanda B Diekman and Alice H Eagly. 2008. Of men, women, and motivation. *Handbook of motivation science* (2008), 434–447.
- [20] Sif Einarsdóttir and James Rounds. 2009. Gender bias and construct validity in vocational interest measurement: Differential item functioning in the Strong Interest Inventory. *Journal of Vocational Behavior* 74, 3 (2009), 295–307.
- [21] European Centre for the Development of Vocational Training. 2016. Skill shortages in Europe: Which occupations are in demand – and why. <https://www.cedefop.europa.eu/en/news-and-press/news/skill-shortages-europe-which-occupations-are-demand-and-why>
- [22] Eurostat. 2018. ICT specialists: statistics on hard-to-fill vacancies in enterprises - Statistics Explained. https://ec.europa.eu/eurostat/statistics-explained/index.php/ICT_specialists_-_statistics_on_hard-to-fill_vacancies_in_enterprises#Employment_and_recruitment_of_ICT_specialists
- [23] Eurostat. 2019. ICT specialists in employment. [https://ec.europa.eu/eurostat/statistics-explained/index.php/ICT_specialists_in_employment#ICT_specialists_by_sex](https://ec.europa.eu/eurostat/statistics-explained/index.php/ICT_specialists_in_employment#ICT_specialists_by_sex)
- [24] Clifford D Evans and Amanda B Diekman. 2009. On motivated role selection: Gender beliefs, distant goals, and career interest. *Psychology of Women Quarterly* 33, 2 (2009), 235–249.
- [25] Diane F Halpern, Camilla P Benbow, David C Geary, Ruben C Gur, Janet Shibley Hyde, and Morton Ann Gernsbacher. 2007. The science of sex differences in science and mathematics. *Psychological science in the public interest* 8, 1 (2007), 1–51.
- [26] Jo-Ida C Hansen, Rose C Collins, Jane L Swanson, and Nadya A Fouad. 1993. Gender differences in the structure of interests. *Journal of Vocational Behavior* 42, 2 (1993), 200–211.
- [27] Jonathan Kibble. 2007. Use of unsupervised online quizzes as formative assessment in a medical physiology course: Effects of incentives on student participation and performance. *American Journal of Physiology - Advances in Physiology Education* 31, 3 (2007), 253–260. <https://doi.org/10.1152/advan.00027.2007>
- [28] Klaus Larsen, Jørgen Holm Petersen, Esben Budtz-Jørgensen, and Lars Endahl. 2000. Interpreting Parameters in the Logistic Regression Model with Random Effects. *Biometrics* 56 (2000), 909–914.
- [29] Richard Lippa. 1998. Gender-related individual differences and the structure of vocational interests: The importance of the people-things dimension. *Journal of Personality and Social Psychology* 74, 4 (1998), 996–1009. <https://doi.org/10.1037/0022-3514.74.4.996>
- [30] Richard A. Lippa. 2010. Gender Differences in Personality and Interests: When, Where, and Why? *Social and Personality Psychology Compass* 4, 11 (2010), 1098–1110. <https://doi.org/10.1111/j.1751-9004.2010.00320.x>
- [31] Melissa Høegh Marcher, Ingrid Maria Christensen, Pawel Grabarczyk, Therese Graversen, and Claus Brabrand. 2021. Computing Educational Activities Involving People Rather Than Things Appeal More to Women (CS1 Appeal Perspective). In *Proceedings of the 17th ACM Conference on International Computing Education Research (ICER 2021), August 16–19, 2021, Virtual Event, USA*. ACM, New York, NY, USA, 13 pages. <https://doi.org/10.1145/3446871.3469761>
- [32] Jane Margolis and Allan Fisher. 2002. Unlocking the clubhouse: The Carnegie Mellon experience. *ACM SIGCSE Bulletin* 34 (6 2002), 79–83. Issue 2. <https://doi.org/10.1145/543812.543836>
- [33] Peter McCullagh. 1980. Regression Models for Ordinal Data. *Journal of the Royal Statistical Society. Series B (Methodological)* 42, 2 (1980), 109–142. <http://www.jstor.org/stable/2984952>
- [34] Stephen L Morgan, Dafna Gelbgiser, and Kim A Weeden. 2013. Feeding the pipeline: Gender, occupational plans, and college major selection. *Social Science Research* 42, 4 (2013), 989–1005.
- [35] Katja Paessler. 2015. Sex differences in variability in vocational interests: Evidence from two large samples. *European Journal of Personality* 29, 5 (2015), 568–578.
- [36] Eva H Shinar. 1975. Sexual stereotypes of occupations. *Journal of vocational behavior* 7, 1 (1975), 99–111.
- [37] Stacey Sinclair, Curtis D. Hardin, and Brian S. Lowery. 2006. Self-stereotyping in the context of multiple social identities. *Journal of Personality and Social Psychology* 90, 4 (apr 2006), 529–542. <https://doi.org/10.1037/0022-3514.90.4.529>
- [38] Statistics Denmark. 2020. Gymnasiale uddannelser. <https://www.dst.dk/da/Statistik/dokumentation/statistikdokumentation/gymnasiale-uddannelser>
- [39] Rong Su, James Rounds, and Patrick Ian Armstrong. 2009. Men and Things, Women and People: A Meta-Analysis of Sex Differences in Interests. *Psychological Bulletin* 135, 6 (2009), 859–884. <https://doi.org/10.1037/a0017364>
- [40] SurveyMonkey Inc. 2020. www.surveymonkey.com San Mateo, California, USA. www.surveymonkey.com
- [41] René T. Proyer and Joachim Häusler. 2007. Gender differences in vocational interests and their stability across different assessment methods. *Swiss Journal of Psychology* 66, 4 (2007), 243–247.
- [42] Mei Tang, Wei Pan, and Mark D Newmeyer. 2008. Factors influencing high school students' career aspirations. *Professional School Counseling* 11, 5 (2008), 2156759X0801100502.
- [43] Fay Pedersen Tveranger. 2017. *Programming in school. An insight to the Norwegian programming pilot and the inclusion/exclusion of girls in computer programming education*. Master's thesis. The University of Bergen.
- [44] Undervisningsministeriet. 2017. Rapport fra Udvalget om ligestilling i dagtilbud og uddannelse.
- [45] Yanna J Weisberg, Colin G DeYoung, and Jacob B Hirsh. 2011. Gender differences in personality across the ten aspects of the Big Five. *Frontiers in psychology* 2 (2011), 178.

A TASKS (T2–T4)

See figures 7, 8, and 9.

B ESTIMATED PROBABILITIES OF STUDENT RATINGS

See Figure 10.

Task 2: You must write an assignment on an article concerning IT

Imagine that you have to write an assignment in the programming class about something you read in an article.

You can choose to write about one of two articles:

Article A



Artificial intelligence gets new breakthrough in text reading

A new wave of text-to-speech AI has been made to understand the content of a text and thus adapt its reading to the mood using advanced machine learning. [Read more](#)

Article B



Artificial intelligence gets new breakthrough in factories

A new wave of AI for factory robots has been made to recognize objects and thus work more independently using advanced machine learning. [Read more](#)

Figure 7: Task T2 (summarize article).

Task 3: You hear about a project that another group in your class has done

Imagine that the class has been divided into groups and all groups have done a project. You hear about two projects that two different groups in your class have done:

Project A

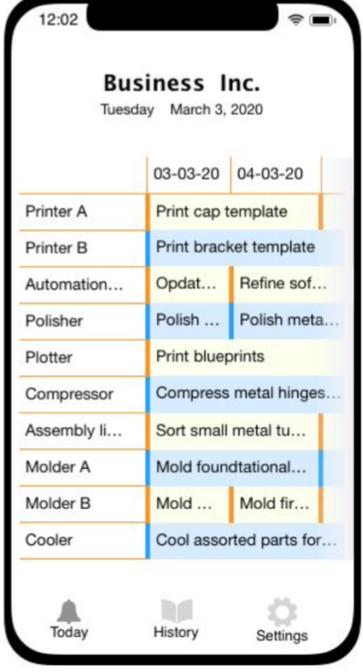
We have created an app for Business Inc. that keeps track of many employees at the same time. In order to optimize the workflow, it is important for the manager to always have an overview of the employees. We have developed an app that gives an overview of the employees. The app makes it possible to see which employees solve what tasks and when.



	03-03-20	04-03-20
Recruiter A	Hire project manager	
Recruiter B	Train student intern	
Accountant	Distribu...	Monthly sala...
IT Support	Updat...	Meeting with...
Expert con...	Participate in interdiscipli...	
HR	Update employee handb...	
Assistant pr...	Review filing structure	
Project Ma...	Approve Q3 budget	
Project Ma...	Test ne...	Update ...
PR Respon...	Participate in internationa...	

Project B

We have created an app for Business Inc. that keeps track of many machines at the same time. In order to optimize production, it is important for the factory to always have an overview of the production machines. We have developed an app that gives an overview of the machines. The app makes it possible to see which machines produce what components and when.



	03-03-20	04-03-20
Printer A	Print cap template	
Printer B	Print bracket template	
Automation...	Opdat...	Refine sof...
Polisher	Polish ...	Polish meta...
Plotter	Print blueprints	
Compressor	Compress metal hinges...	
Assembly li...	Sort small metal tu...	
Molder A	Mold foundational...	
Molder B	Mold ...	Mold fir...
Cooler	Cool assorted parts for...	

Figure 8: Task T3 (evaluate presentation).

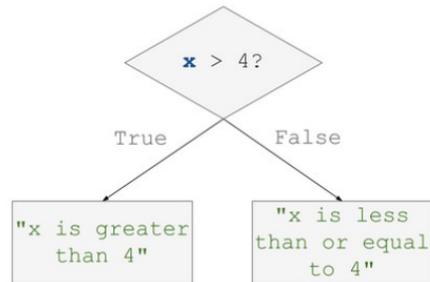
Task 4: You are taught how to program

Now imagine that you are taught programming in your class. You are introduced to what is called an "if-else" statement.

In programming, one can use something called an "if-else statement". It is used to tell the computer to do something if a statement is true, otherwise (else) do something different.

For example, you can write a small program that tells you if a variable called "x", is greater or less than 4.

```
variabel x = 5
if (x > 4)
    "x is greater than 4"
else
    "x is less than or equal to 4"
```



So the program checks if the statement $x > 4$ is true or false. In this case, it would be true ($5 > 4$) and the program would, therefore, say "x is greater than 4". If we changed the variable x to be 3, it would be false ($3 > 4$) and the program would say "x is less than or equal to 4".

Imagine that you have to solve a programming assignment in class. You can pick between these two:

Assignment A	Assignment B
<p>Your first programming task is to understand the code of a small program. The program should be used for an email service that sends reminders. The program must check if a couple has a "gold" anniversary, which they have after being married for 50 years.</p> <p>If 'years' is 50 years, the program should say "Anniversary", if not, it should say "No anniversary".</p>	<p>Your first programming task is to understand the code of a small program. The program should be used in a sorting machine that produces screws. The program must check if a screw is the standard size, which it is when it is 50 mm.</p> <p>If 'screw' is 50 mm, the program should say "Approved", if not, it should say "Not approved".</p>
<pre>variabel years = 54 if (years = 50) "Anniversary" else "No anniversary"</pre>	<pre>variabel screw = 54 if (screw = 50) "Approved" else "Not approved"</pre>
<pre> graph TD A{years = 50?} -- True --> B["Anniversary"] A -- False --> C["No anniversary"] </pre>	<pre> graph TD A{screw = 50?} -- True --> B["Approved"] A -- False --> C["Not approved"] </pre>

Figure 9: Task T4 (solve exercise).

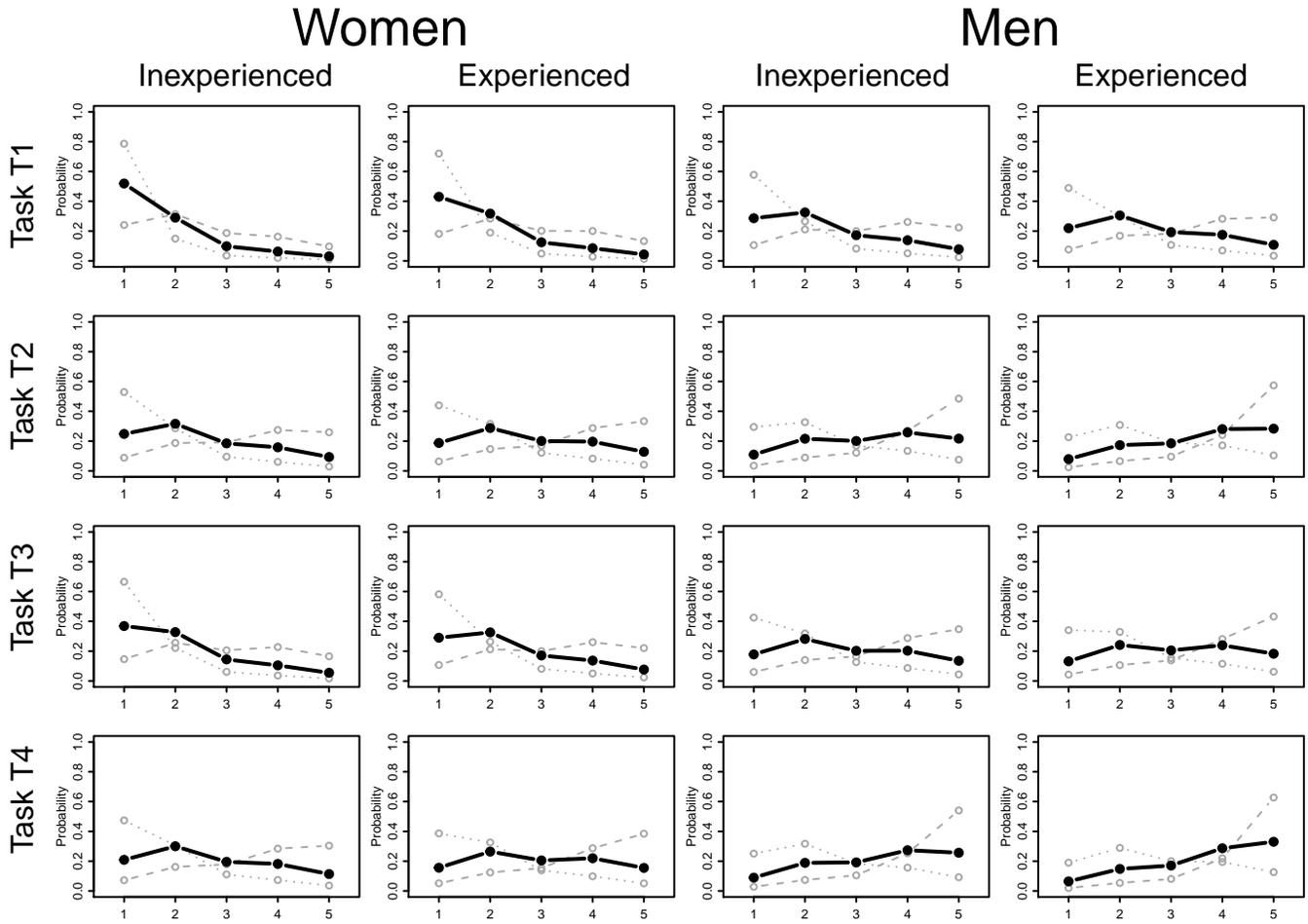


Figure 10: Estimated probabilities of ratings for each of tasks T1 to T4 for the four combinations of gender and programming experience. Shown are the probabilities for a median student (solid lines), and an indication of the middle 80% of students via the probabilities of students in the upper (dotted) and lower (dashed) 10th percentiles.