Gender Influences On School Students’ Mental Models of Computer Science
A Quantitative Rich Picture Analysis with Sixth Graders

Philipp Brauner
RWTH Aachen University
Aachen, Germany
brauner@comm.rwth-aachen.de

Martina Ziefle
RWTH Aachen University
Aachen, Germany
ziefe@comm.rwth-aachen.de

Ulrik Schroeder
RWTH Aachen University
Aachen, Germany
schroeder@cs.rwth-aachen.de

Thiemo Leonhardt
RWTH Aachen University
Aachen, Germany
leonhardt@cs.rwth-aachen.de

Nadine Bergner
RWTH Aachen University
Aachen, Germany
bergner@cs.rwth-aachen.de

Birgit Ziegler
University of Darmstadt
Darmstadt, Germany
ziegler@bp.tu-darmstadt.de

Figure 1: Examples of the drawings of the school children’s mental models of computer scientists. The images were evaluated along multiple dimensions and linked to the drawers’ gender and self-efficacy in interacting with technology.

ABSTRACT
Despite great efforts, women are underrepresented in computer science and other science, technology, engineering, and mathematics disciplines. To understand the root of this we studied 112 high school students (51 female, 61 male) aged between 10 to 13 years. The questionnaire-based survey revealed a significant effect of gender on technical self-efficacy and on interest in computer science. Furthermore, we extracted the students’ mental models by letting them draw computer scientists. A rich picture analysis revealed significant effects of gender on the stereotypicality of the images. The gender gap revealed by this multi-method approach influences students’ career decisions and yields in declining participation of women in STEM. Consequently, future measures must focus on school students at an earlier age. These should be offering a variety of male and female role models and should strengthen the individual’s technical self-efficacy, as it profoundly impacts later career decisions.

CCS CONCEPTS
• Social and professional topics → Computer science education; K-12 education; Gender;

KEYWORDS
Gender, Stereotype, Mental Models, Self-Efficacy, STEM, Computer Science, Science and Technology Education, Rich Picture Analysis

ACM Reference format:
https://doi.org/10.1145/3196839.3196857

1 INTRODUCTION
The demand for science, technology, engineering and mathematics (STEM) experts is rising and the growth of STEM employment is considerably larger than the total employment growth [21]. Despite excellent job opportunities, many high school graduates do not consider studying STEM related topics at university and therefore neglect pursuing a career in this challenging, but also profitable and rewarding domain [35]. Although the work of women had a major impact on computer science and shaped its history, present, and future [23], women are massively underrepresented in computer science and most STEM domains [38–40]. While the shortage of STEM professionals can be argued controversially [12],
the gender gap is unacceptable and threatens gender equality in these domains.

Many initiatives and research projects aim at increasing the overall number of students and graduates in the STEM disciplines and some specifically address women as the largest underrepresented minority in these areas [15, 19, 31, 36, 41, 43]. However, most measures address this problem around the time of high school graduation or at a university level. Yet, several studies suggest that the pivotal point may be much earlier [8, 34].

In this article we present a field study with school students from sixth grade in the age of 10 to 13 years. We show that gender differences are already present at this age and that these affect technical self-efficacy, interest in computer science, and the mental models of computer scientists (see Figure 1).

2 BACKGROUND

With their seminal work “Unlocking the Clubhouse”, Fisher and Margolis identified a complex interconnected network of reasons why women refrain from studying computer science [18, 19]. For example, they found that many male computer science students enjoyed “hacking for hacking’s sake”. In contrast, many women were rather interested in the application of computer science in other domains, such as arts, medicine, or physics and the societal consequences of digitalization. In other words, women prefer domains, where the benefits of applying computer science are clearly visible. In addition, they identified lower computer self-efficacy as one key barrier for pursuing a career in computer science.

2.1 Self-efficacy Believes

Following Bandura’s self-efficacy theory, self-efficacy refers to the belief of a person to have the competency to carry out a desired action [4]. Self-efficacy believes are rooted in own personal experiences (“experience of mastery”), observations of peers and models (“vicarious experience”), the social support (“social persuasion”), and also physiological reactions (“emotional arousal”).

Personal experience: If a person successfully handles a difficult situation, this will strengthen the belief that he or she will be able to master comparable situations successfully in the future. Failures, on the other hand, can lead to a situation in which the person is more likely to doubt his or her skills in the future.

Observation of models/peers: For the purpose of estimating one’s own self-efficacy, people also consider the behavior of similar persons and role models in these situations. If a situation is mastered by a model – for example, a fellow student – this increases the self-efficacy expectation of the person her- or himself.

Social support: If a person is reassured by others to be able to judge a situation, this increases their faith in this ability. Conversely, doubts lead to a lesser belief in one’s own competence.

Physiological reactions: According to Bandura, physiological reactions of the body are also a factor of self-efficacy expectations. If the thought of a situation leads to increased palpitations, tremors, tension or even fear, this reduces the self-efficacy maintenance.

Self-efficacy believes affect the choice of possible actions, the performance while doing these actions, as well as the endurance during these action.

Choice of action: People will rather choose tasks from areas where they have a high self-efficacy expectation. Similarly, tasks in areas with low self-efficacy expectations are avoided.

Performance: Not only the choice of an action is influenced, but also the achieved performance. People with a high self-efficacy expectation achieve a higher performance than people with a lower self-efficacy expectation.

Endurance: In addition to the increase in performance, people with high self-efficacy expectation tend to occupy themselves with a task for a longer period of time despite the occurrence of obstacles.

Gender differences in computer self-efficacy have been discussed for many years and most often women report much lower scores than men [3, 8, 10, 11, 44]. Consequently, women often show lower performance when interacting with interactive systems and have lower endurance if difficulties emerge. Also, self-efficacy believes are closely related to the locus of control from Weiner’s theory of causal attribution [22, 45]. Theory of causal attribution has recently gained traction in computer science and human-computer interaction research, as it explains whether students attribute success or failure to their own action or to external factors [9, 27, 37]. Most importantly, this also affects self-efficacy and participation in university level computer science degree programs [2, 19]. Figure 2a illustrates the relationships within the self-efficacy model.

2.2 Self-efficacy and Career Choices

In a broader scope, the concept of self-efficacy is also a key constituent of Lent et al.’s Social Cognitive Career Theory (SCCT) [28, 29]. Self-efficacy believes are embedded into a model that describes how professional or academic interest is generated, how career-relevant decisions are made, and which factors determine the achievement and endurance of the service provided. Independent variables of the model are the self-efficacy believes already discussed, the expected results, as well as personal goals.

Self-efficacy believes: The self-efficacy expectation is analogous to Bandura’s model of the expectation, justified in the assessment of one’s own competence, to be able to carry out a certain action successfully.

Outcome expectations: The expected results describe the results or consequences expected from an action performed. For example, a person may have the expectation that a computer science degree will lead to a high income. The expected results are influenced either by personal experience or by observing peers or the environment. Self-efficacy expectations are generally given a higher priority in the selection of possible actions than expected results.

Personal goals: Personal goals are defined as the intention to pursue a certain activity or to achieve a certain result. In this context, goals are considered to be an important instrument with which people can organize, direct and sustain their efforts over a longer period of time without external
reinforcement. The objectives set by a person are largely dependent on self-efficacy expectations and the expectation of results.

Similar to the self-efficacy model, these factors influence the development of career interests, the choice between possible career paths, and the performance and achievements in the career.

**Interest**: Interest in an activity develops when a person feels competent in an area, and thus has a high degree of self-efficacy, and when the presumed consequences of this activity are assessed as positive. This interest, coupled with a high degree of self-efficacy and good expectations for results, manifests itself in the setting of personal goals that are either mastered or missed by the person.

**Choice of career paths**: Should people be prevented from doing what they are interested in—for example because of economic difficulties, the lack of formal qualifications, discrimination, or physical handicaps—SCCT provides options for choosing between different possible career paths. The choice will be based on the possibilities available, whether the person has confidence in them, and whether the expectation of results is sufficiently positive.

**Performance**: In the SCCT model, performance includes both success in handling tasks and staying power in the event of problems. The performance is influenced by the actual ability, the perceived self-efficacy expectation, the expectation of results, but also by personal goals (e.g., a good intermediate diploma) that are set by the client.

SCCT predicts that an optimistic, albeit appropriate, self-efficacy expectation promotes good performance. However, career performance suffers if the required skills are not available, or if the self-efficacy does not match the existing abilities. If one’s ability is underestimated, lower goals are set and the attained performance is reduced in case of obstacles. Figure 2b illustrates the SCCT.

The (perceived) achievements play a decisive role in future career decisions, because a positive experience—for example experiences of success in a robot programming course—leads to an increased self-efficacy expectation and thus higher goals [30].

Gender (i.e., the social gender role), ethnicity, and other existing or constructed distinguishing features can also be located in the theoretical model of the SCCT: Due to negative expectations of results, e.g., due to the lack of role models of one’s own group, people may choose different career paths or show different performances despite comparable self-efficacy expectations due to different expectations of results. Similarly, the self-efficacy expectations of boys and girls will develop differently if playing with gender-specific toys.

The next section presents the concept of mental models that serve as proxies for role models and therefore shape self-efficacy believes and career choices.

### 2.3 Stereotypes, Mental Models, and Rich Picture Analyses

Mental models are cognitive representations of complex objects, processes, or structures and enable us to evaluate the consequences of our actions. These simplified mirrors shape our behavior and, if aligned with reality, they enable us to efficiently and effectively interact with our environment [13, 20, 24]. However, incorrect mental models prevent correct inferences, limit the evaluation of our abilities, and may therefore prevent us from selecting optimal actions and achieving what we could have.

Mental models have been studied in various domains and contexts. For example, the school students’ mental models of scientists were first analyzed by Mead and Metraux in 1957 [33]. In their study with school children from different schools and of different ages, scientists were seen as a “man who wears a white coat and works in a laboratory. He is elderly or middle aged and wears glasses”. In the context of computer science education, Denham used a rich picture analysis to analyze how children’s drawings of computers [14] and found that the understanding of computer components and their relationship increases with age and experience. Borthwick [7] used a rich picture analysis and subsequent interviews to learn why disaffection for mathematics with school students is high. She found circumstantial evidence that disaffection might be caused by both a perceived lack of applicability of mathematics as well as limited collaboration. In a review article, Finson concludes that mental models of scientists got less stereotypical in recent years and that students’ perception of scientists can be positively influenced by providing adequate role models and personal experiences with science and scientists [17]. Miller et al. conducted a meta-analysis of five decades of Draw-A-Scientist studies in the U.S. [34]. While the results show that the overall gender bias has decreased in recent decades, the study also found that the
proportion of men in the images increases with the students’ age. Consequently, the students’ mental models of scientists get more stereotypical with age.

Also, a recent study by Völkel et al. found gender differences regarding the expectations towards computer science at a university and men reported a higher fit between a prototypical image of computer scientists and their self-image than women [43].

In conclusion, mental models have been found to relate to observation and experience and are expected to guide our decision making. To understand if mental models relate to the low participation in computer science (and presumably other disciplines from the STEM spectrum) and the very low share of girls and women in these, we conducted an empirical study guided by the following research questions:

1. Does gender influence computer self-efficacy and interest in computer science of sixth graders?
2. Does gender influence school children’s mental models of computer scientists?
3. Does self-efficacy relate to school children’s mental models?

3 METHOD

First, this section briefly presents the evaluation frame in which the school students’ images were acquired, the sample of the study (Sec. 3.1), and the measured personality states and traits of the students (Sec. 3.2). Next, the framework for analyzing and quantifying the images is presented (Sec. 3.3) and the actual image quantification process and its validity are presented (Sec. 3.4). Last, a stereotype score is calculated for each image (Sec. 3.5). The actual results of the rich picture analysis are then presented in the Section 4.

3.1 Procedure and Description of the Sample

The subjects of this study were participants of robotic courses for increasing vocational interests and interest in computer science (CS). These courses are based on the Roberta workshops [30, 41] and are usually organized as two-day workshops for girls aged 12 to 14 years. The positive short term effects of this measure on technical self-efficacy and vocational interest have been discussed in [8, 30, 41]. The present study was aimed at identifying the predictors for successful course outcomes and positive long term effects of these interventions. Yet, organizational constraints required us to study slightly younger children in gender balanced courses.

Targeting measures to increase children’s interest in technology design and development is difficult, as two different forces must be balanced: On one side, the children’s cognitive abilities must be sufficiently developed to understand the required fundamental programming concepts [30]. On the other side, deeper personal interest is build and maintained only in a few subject areas and opportunities for the development of interests must be offered at an early stage [27]. Also, the perception of the usefulness of technology decreases with age and negatively influences the development of interest in this field. Specifically, differences in technical self-efficacy and interest in technology are already evident at the age of 13/14 years [47]. Consequently, measures for increasing technical self-efficacy and for generating vocational interest in design and development of technology must begin at the latest at the age of ten.

In total, 113 children from the sixth grade from three schools in Aachen, Germany and four different classes participated in this study. 51 are girls (46%) and 61 are boys (54%). Most school students were aged 11 (n = 43) or 12 years (n = 66); two students were aged 10 and one was 13.

3.2 Measured Variables

A few days before the robotic workshop, the school children completed pen and paper questionnaires of about 45 minutes with their demographic data and their personality states and traits. Specifically, we asked for the following aspects:

- To investigate the influence of the students’ sex on computer science, we captured his or her gender (binary choice).
- We also asked for the kids’ age, although they were all from the sixth grade and consequently in an age range from 10 to 13 years.
- Self-efficacy in interacting with technology (SET) is measured using an 8 item scale [5].
- We captured the students’ interest in computer science. Interest was subdivided into the dimension interest in technical computer science, applied computer science, theoretical informatics, and practical computer science.
- As the final questionnaire had a free page, we asked the students to draw a picture of a computer scientist. This task was carefully phrased to avoid gender biases.

All constructs (besides gender and the image) were captured on 6-point Likert-scales, recoded, and aggregated into a mean score ranging from 0 – 100% if suitable.

3.3 Quantification of the Drawn Pictures

Of the 113 children who have participated in the study, 99 submitted a drawing and 14 did not submit a picture. Some children wrote additional comments on their drawings, ranging from additional explanations of their drawings (e.g., “I envision computer scientists as nice people but a bit hyper smart. Difficult to draw”, “a typical girl”, or “also without glasses”) to aspects they found not easy to portray, such as computer scientists being “nerds” or “only computers”. In four cases, the children added thought bubbles with mathematical formulas or a light bulb, probably suggesting an extraordinary intellect or a stroke of genius. The hand written comments probably influenced the raters’ decisions, but beyond that the comments were not further investigated in this study. In addition, one student commented in huge letters “I hate drawing” instead of submitting a picture.

To rule out biases that might influence the picture analysis, we asked nine adult reviewers aged 25 to 38 years to rate the drawings along multiple dimensions. Six reviewers had a STEM background and three had a background in social sciences. Four are female and five are male. All reviewers had to evaluate every drawing.

They rated the pictures along multiple dimensions on 10-point semantic differentials (a scale with two opposing poles). These dimensions were defined in an expert workshop by analyzing similar

---

1The questionnaire was in German and the question “Zeichne einen Informatiker” (“draw a computer scientist”) might prime towards male computer scientists. Hence, we used the phrase “Zeichne einen Menschen, der im Bereich Informatik arbeitet” (“Draw a person working in computer science”), which should not induce a bias.
pictures gathered at a previous Girls’ Day event. The evaluated dimensions were sex (does the drawing picture a male or a female person; ranging from male to female), “mood” (is the mood presented in the image negative to positive), attractiveness (is the appearance of the person not at all attractive to very attractive), situation (is the person pictured as isolated or shown in a social context), intelligence (the person looks less intelligent or very intelligent), and geek (does the person look like a geek or not like a geek).

Furthermore, they indicated as binary choices whether computers, additional technical gadgets or mobile devices were present in the drawing and whether the person in the picture was wearing glasses. Lastly, the reviewers reported if the presented image was actually a drawing.

The criteria used relate to the widespread belief that it is mainly men who work in the field of computer science and that their activities there would require little social interaction [19, 25]. Some criteria, such as the dimension “geek”, may be controversial, but the use of the term in both pop and subculture justifies its use for this study. In addition, it is also possible to reevaluate the present image set with different criteria.

We used the application shown in Figure 3 for gathering the ratings. It presents the drawing on the right hand side of the screen and sliders and check-boxes for entering the judgments on the left hand side.

3.4 Rich Picture Analysis
First, we checked if the reviewers rated the pictures consistently, i.e., if the reviewers agreed on the same ratings for each dimension of each picture. Thereto we calculated the Intra-Class-Correlation coefficient (ICC) (3, 1) for every dimension assessed (see [16] for a detailed explanation). We found a high to very high and always significant agreement on all dimensions between the reviewers, ranging from $ICC_{avg} = .961$ [$p < .001$] for the dimension sex to $ICC_{avg} = .720$ [$p < .001$] for the dimension intelligence. Consequently and as shown in Table 1, there was little to no disagreement between the reviewers’ evaluation of the pictures.

Due to the reviewers’ agreement it is permissible to calculate the arithmetic mean of the individual ratings for each image. Consequently, each image is assigned with a score for each assessed dimension (e.g., sex, sociality, ...). Each score has a theoretical range from $-100$% to $+100$%. For example, a score of $+100$% on the social scale would indicate that the depicted situation is highly social, a score of $0$% would indicate a neutral situation, whereas a score of $-100$% would indicate a very unsocial or isolated situation.

This mean score was further split into a three-stage variable, allowing us to differentiate between pictures clearly showing a specific characteristic (e.g., a picture clearly showing a man or a woman) and pictures lacking this clarity (e.g., pictures with both a man and a woman or pictures that do not allow our reviewers to identify the sex of the depicted persons). Also, we aggregated the individual ratings for the categorial dimensions into one categorial variable by using the majority of raters (e.g. if six evaluators indicated that a mobile phone was present, then the image is considered as containing a mobile phone). As above, the ratings were consistent for all dimensions. Just in one picture did the raters disagree about the occurrence of computers: A computer was visible in the drawing, but only as a t-shirt print.

3.5 Calculation of a Stereotypical Score
For a better understanding of the picture space, we also calculated a principle component analysis (PCA) for the four variables mood, sociality, geek, and attractiveness. This method is able to identify, if the underlying data feature an uni- or multidimensional structure. However, the PCA identified only a single dimension for the picture space. As the internal reliability of this dimension was high ($\alpha_n=104 = .770$), we were able to calculate a stereotypical score for each image, as the average score across each of the four dimensions. Pictures scoring high on this scale show very stereotypical depictions (less social, less attractive, negative mood, and geeky appearance of the people in the image) related to literature research [26].

In the final step, these scores for the images are added to the survey responses of the children (e.g., gender, set, ...) and the relationships between the variables from the surveys and the drawn images can now be investigated in the following section.

4 RESULTS
The first part of the results section presents the effects of gender on self-efficacy in interacting with technology and the childrens’ attitude towards computer science. As this perspective is not particularly new, readers seeking in-depths information about this are advised to consult additional articles (e.g., [8, 19, 47]).

The second part is dedicated towards the analysis of the drawn images and the relationship between the images and the attitudes of the children.

4.1 Childrens’ Attitudes Towards Computer Science
We analyzed the students’ self-efficacy in interacting with technology (set) and their interest in computer science. Both scales achieved an excellent internal reliability [set: Cronbach’s $\alpha_{n=104} = .879$, interest: Cronbach’s $\alpha_{n=113} = .972$].
With an average of 69 ± 18% the reported SET scores were considerably above the center of the scale, whereas the average scores for INTEREST in computer science were at the center of the scale (49.9 ± 22.7%).

We found a significant effect of the student’s GENDER on the SET, $F_{1,106} = 19.857, p < .001, \eta^2 = .158$. As Figure 4 illustrates, the girls reported a significantly lower technical self-efficacy (61 ± 21%) than the boys (77 ± 17%).

Also, a one-way ANOVA with GENDER as independent variable and overall INTEREST in computer science as dependent variable shows a significant effect of GENDER, $F_{1,101} = 4.013, p = .045 < .05, \eta^2 = .039$. The girls reported a significantly lower INTEREST (44.8 ± 19.8%) in computer science than the boys (53.8 ± 24.1%). Again, Figure 4 illustrates this effect.

### 4.2 Overall Analysis of the Mental Models of Computer Scientists

First, we wanted to understand the general perception of computer scientists through the eyes of the children:

On average, the children drew computer scientists as predominantly male (~32.7±58.6%). When the GENDER of the drawn person is seen as categorical variable (male, undefined, female), then 67 of the 99 drawings show male computer scientists (67.7%), 19 drawings show female computer scientists (19.2%) and 13 drawings are rated as indecisively (e.g., because of multiple people visible in the picture or because of a very small drawing).

Regarding the MOOD of the images, the average score across all images was ~12.4 ± 31.2% and thus below a neutral rating of 0%.

On average, the score for the ATTRACTIVENESS of the people drawn achieved a score of +22.4 ± 43.8% and can be interpreted as positive.

The SITUATIONS depicted in the images were – on average – evaluated as less social (~23.2 ± 33.6%).

The average INTELLIGENCE score for the drawn persons were positive (17.6 ± 17.0%), hence the drawn persons are perceived as above average intelligent.

The last considered dimension, how “geeky” the drawing appears, achieved a score of ~17.0 ± 24.8% and thus below the center of the scale.

Table 1 and Figure 5a show the average scores. As the 95% confidence intervals indicate, each of the six dimensions is significantly different from 0% and thus significantly different from the center of the scale.

Computers are present in 55 and thus in about half of the 99 drawings (53.3%). As stated above, one picture was hard to classify as a computer was present as a t-shirt print (counted as computer is present).

Although it is impossible to imagine everyday life without them, smart phones or mobile phones are not represented in any of the pictures. However, additional gadgetry is visible in 10 of the 103 drawings (9.7%).

In about half of the pictures people wear glasses (50.5%). This is actually a good estimate and slightly below the actual number of people wearing glasses in Germany of 62% [1].

In summary, the persons in the pictures were drawn as predominantly male and in rather isolated situations. The persons are perceived as looking like a geek, although also perceived as rather attractive and rather intelligent. Still, the mood in the images was perceived as slightly negative.

### 4.3 Intra-Image Correlations

In the next step, we calculated a correlation analysis to evaluate if the assessed dimensions are interconnected. As the upper part of Table 2 shows, some of the criteria are closely connected; whereas others are not. The GENDER of the drawn persons is linked to the MOOD of the images and rather female images were evaluated as more positive ($\rho = .359$). GENDER is also linked to the SITUATION and pictures with women were rather associated with social situations ($\rho = .381$). Persons in images with a more positive MOOD were rated as more ATTRACTIVE ($\rho = .457$) and the situations were evaluated as more social ($\rho = .546$). Strikingly, the ATTRACTIVENESS of the drawn persons was also linked to the presented SITUATION and how social it appeared ($\rho = .501$).

Surprisingly, the GEEK-dimension is associated with all other dimensions: Images with a more male appearance were classified as more GEEKY ($\rho = .452$). Also, a negative MOOD in the images was strongly associated with the level of Geekism in the images ($\rho = .613$). Furthermore, the scale for ATTRACTIVENESS and GEEK was associated ($\rho = .357$) and the more attractive persons in an image were evaluated, the less geeky they were perceived. The perception of the SITUATION depicted in an image was likewise linked to the GEEKNESS ($\rho = .557$). Persons drawn in social situations were rather not perceived as geeks by the evaluators. Finally, we found a weaker, but still significant, relationship between the perceived INTELLIGENCE of the drawn persons and the GEEK-score ($\rho = .181$). Although the last effect is considerably lower than previous, more intelligent persons were more often considered geeks.

### 4.4 Effects of the Drawer on the Pictures

Now we wanted to understand if the participants’ GENDER influences their mental model of computer scientists, represented by the drawn pictures. Consequently, we calculated a MANOVA (multivariate analysis of variances) with GENDER as independent variable and the six evaluation criteria SEX, ATTRACTIVENESS, MOOD, SITUATION, INTELLIGENCE and GEEKNESS as dependent variables.
### Table 1: Scoring criteria and their inter-rater-reliability. All dimensions were evaluated consistently by the reviewers.

<table>
<thead>
<tr>
<th>Category</th>
<th>Scale</th>
<th>Reliability</th>
<th>Overall</th>
<th>Girls</th>
<th>Boys</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>IRR</td>
<td>M SD</td>
<td>M SD</td>
<td>M SD</td>
</tr>
<tr>
<td>Sex</td>
<td>(male–female)</td>
<td>.961</td>
<td>p&lt;.001</td>
<td>-32.7%</td>
<td>58.6%</td>
</tr>
<tr>
<td>Mood</td>
<td>(negative–positive)</td>
<td>.837</td>
<td>p&lt;.001</td>
<td>-12.4%</td>
<td>31.2%</td>
</tr>
<tr>
<td>Situation</td>
<td>(unsocial–social)</td>
<td>.836</td>
<td>p&lt;.001</td>
<td>-23.2%</td>
<td>33.6%</td>
</tr>
<tr>
<td>Attractiveness</td>
<td>(not intelligent–very intelligent)</td>
<td>.950</td>
<td>p&lt;.001</td>
<td>22.4%</td>
<td>43.8%</td>
</tr>
<tr>
<td>Intelligence</td>
<td>(not intelligent–very intelligent)</td>
<td>.720</td>
<td>p&lt;.001</td>
<td>17.6%</td>
<td>17.0%</td>
</tr>
<tr>
<td>Geek</td>
<td>(very geeky–not geeky)</td>
<td>.735</td>
<td>p&lt;.001</td>
<td>-17.0%</td>
<td>24.8%</td>
</tr>
<tr>
<td>Computer</td>
<td>Computer / No Computer</td>
<td>.979</td>
<td>p&lt;.001</td>
<td>53.4%</td>
<td>–</td>
</tr>
<tr>
<td>Mobile</td>
<td>Mobile phone / No mobile phone</td>
<td>.724</td>
<td>p&lt;.001</td>
<td>0.0%</td>
<td>0 / 103</td>
</tr>
<tr>
<td>Gadgets</td>
<td>Gadgets / No gadgets</td>
<td>.890</td>
<td>p&lt;.001</td>
<td>9.7%</td>
<td>10 / 93</td>
</tr>
<tr>
<td>Glasses</td>
<td>Glasses / No glasses</td>
<td>.964</td>
<td>p&lt;.001</td>
<td>50.5%</td>
<td>52 / 51</td>
</tr>
<tr>
<td>No Picture</td>
<td>Picture / No Picture</td>
<td>.987</td>
<td>p&lt;.001</td>
<td>3.8%</td>
<td>99 / 4</td>
</tr>
<tr>
<td>OVERALL</td>
<td></td>
<td>.922</td>
<td>p&lt;.001</td>
<td>3.8%</td>
<td>99 / 4</td>
</tr>
</tbody>
</table>

**Figure 5:** Mean classification of the images and mean classification separated by the artists’ gender (whiskers show the SE).

Table 2: Spearman’s $\rho$ correlations of the children’s gender and set on the six assessed dimensions of the images (gender, mood, look, social, intelligence, and geek). Gender in the images is coded with $-100\%$=male and $+100\%$=female and drawers’ gender is coded as boy=2 and girl=1. [$^{*} p < .1, * = p < .05, ** = p < .001$]

<table>
<thead>
<tr>
<th>Gender</th>
<th>Mood</th>
<th>Attractiveness</th>
<th>Social</th>
<th>Intelligence</th>
<th>Geek</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (male - female)</td>
<td>.359**</td>
<td>.381**</td>
<td>.452**</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mood (negative - positive)</td>
<td></td>
<td>.457**</td>
<td>.546**</td>
<td>.613**</td>
<td></td>
</tr>
<tr>
<td>Attractiveness (not at all - very)</td>
<td></td>
<td></td>
<td>.501**</td>
<td>.357**</td>
<td></td>
</tr>
<tr>
<td>Situation (unsocial - social)</td>
<td></td>
<td></td>
<td></td>
<td>.557**</td>
<td></td>
</tr>
<tr>
<td>Intelligence (low - high)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.181*</td>
</tr>
<tr>
<td>Geek (very geeky - not at all)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

$\text{Gender}$ had a significant overall effect on the drawings regarding stereotypes [$V = .319, F_{6,96} = 7.481, p < .001, \eta^2 = .319$], meaning that the pictures drawn by boys differ significantly from the images drawn by girls.

First, the drawers’ $\text{gender}$ had a significant influence on the gender depicted in the images [$F_{1,101} = 20.035, p < .001$]. Boys clearly preferred to draw images depicting their own gender ($-53.6 \pm 38.8\%$), whereas the girls’ images were more balanced, but still
lead towards the "male side" (−5.7 ± 70.0%). However and as the large standard deviation suggests, the images were not strictly balanced, but girls drew male and female looking computer scientists equally often. The same picture emerges if gender of the drawn images is analyzed as a categorical variable (male, undefined, female) \( \chi^2 = 22.8, p < 0.01 \): Girls drew men (n = 20) and women (n = 17) about equally often, whereas boys more often drew men (n = 47) instead of women (n = 17). Eight of the girls drawings and five of the boys drawings were not clearly classifiable.

The mood of the images was also significantly influenced by the drawers’ gender \( F_{1,101} = 6.481, p = .012 < .05 \). The boys’ images had a slightly more negative atmosphere (−19.1 ± 29.0%) than the images of the girls (−3.8 ± 32.1%).

The situation was not influenced by gender \( F_{1,101} = 2.497, p = .117 > .1 \) and the girls’ images (−17.3±37.1%) were evaluated as about equally social as the boys’ images (−27.8 ± 30.2%).

Likewise, the attractiveness of the drawn persons was not affected by the drawers’ gender \( F_{1,101} = 1.784, p = .185 > .1 \) and the girls’ persons were evaluated as about equally attractive (28.9 ± 42.2%) as the boys’ persons (17.3 ± 44.6%).

However, the perceived intelligence of the drawn computer scientists was evaluated significantly higher for the images drawn by girls (25.7 ± 14.9%) than for the boys’ images (11.2 ± 15.9%) \( F_{1,101} = 22.37, p < .001 \).

Lastly, there was no significant effect of the drawers’ gender on the geek score of the images \( F_{1,101} = 3.423, p = .062 < .05 \) and the persons drawn by boys were evaluated as about equally geeky (−20.9 ± 25.9%) as the persons from the girls (−11.9 ± 22.6%).

Regarding the self-efficacy in interacting with technology – which was found to be significantly lower for the girls than for the boys – a significant effect was only found for the situation \( \rho = .182, p < .05 \) and students with higher self-efficacy in interacting with technology drew more images depicting more social environments.

An amusing trifle is the significant effect of gender on whether the drawn person wears glasses or not \( \chi^2 = 6.62, p < 0.05 \). The girls more often drew pictures with people wearing glasses (66.6%) than the boys (37.9%). The illustration of computers on the other hand is not influenced by the gender \( \chi^2 = 1.7, p = 0.1 \).

In summary, boys have mainly drawn persons of their own gender, whereas girls’ were less biased towards their own gender or towards male computer scientists. Boys had a more negative perception of the work of computer scientists and girls imagine computer scientists to be more intelligent than the boys do. Figure 5b and Table 1 shows the effect of gender on the images.

5 DISCUSSION

In this work, we used the rather unusual evaluation technique to analyze images of computer scientists hand-drawn by sixth graders and to investigate the relationship between the images and gender and self-efficacy in interacting with technology of school children. As discussed in Section 3.2, we must admit that the analysis of the children’s mental models and the concluded findings are serendipitous, because the image analysis was not part of a formal research model. Nevertheless, we made some astonishing findings that relate to the image of computer science.

First, we found that girls reported significantly lower levels of technical self-efficacy and lower interest in computer science than boys. This corroborates previous findings that women often have lower self-efficacy in interacting with technology than men [3, 8, 10]. Nevertheless, we are profoundly concerned that this effect emerges that early and can empirically be measured at the age of 11 or 12 years.

Second, the analysis of the images drawn by the children offers one possible explanation for this effect. Tversky and Kahnemann showed that if precise information on a topic is not available, we use heuristics that are based on observations, examples, or assumptions and that these heuristics can yield in systematic misconceptions on the topic (availability heuristic) [42]. The present study sheds light on the mental models and assumptions (and possible observations?) the children have of computer scientists that may then shape their short-term interest and self-efficacy, as well as their long-term career paths. Bandura’s self-efficacy theory states that self-efficacy is fed, among other constructs, by role models and peers [4]. Of course, this relates to the drawn images, as the mental models and archetypes of computer scientists shape whether they are perceived as a positive and thus an admirable role model, or as negative and thus as a rather fearful counter-example. This study shows that most children have a misaligned stereotypical image of computer scientists: They are seen as predominantly male – especially by the boys – and they were depicted in rather isolated situations with limited social interactions. Furthermore, the presented situations were classified as slightly negative and the persons were drawn to be less attractive.

On the positive side, we noticed that the share of images showing women in computer science is above the number of women enrolled in IT study programs and currently working in IT [38, 39].

As outlined at the beginning of this article, there are myriads of measures to address the low retention rate of women in STEM. Yet, most of these measures target girls in the middle of secondary school education and also frequently near the transition phase between school and university, or at university level. This article has shown that gender differences in regard to mental models, self-efficacy, and interest have already developed by the age of 12 years and possibly even earlier. Consequently, measures that address the quality of higher education programs in science, technology, engineering, and mathematics, are still a very valuable contribution to address the retention rates and increase the number of students successfully completing the programs. Yet, they fail to address the pivotal points in a students career development process, when mental models, interests, and self-efficacy begin to diverge and the majority of students, predominantly but not exclusively female, start to neglect pursuing a career in STEM.

Thus, measures to increase the low interest in computer science (and other areas of the STEM spectrum) in general and to increase the interest of women in particular need to address younger and youngest children. The theories this study builds on provide multiple anchors that might be addressed: For example, Banduras’ self-efficacy theory [4] postulates that personal experiences, model observation, social support, and physiological reaction are key constituents for the development of self-efficacy.

Successful examples for measures that provide model observation, social support, and personal experiences include computer
programming courses and re-attribution trainings for young school children, especially for girls [8, 30, 32]. Another playful and self-directed example to create technical self-efficacy might be Electronic Bricks [46]: Conventional building blocks are augmented with sensors, actuators, and logic functions. Children can combine them with ease to build simple cyber-physical systems. This enables them to develop an understanding of computing principles on the one hand, and to increase their technical self-efficacy on the other. In regard to mental models, Finson et al. have shown that children’s mental models can successfully be adjusted to better capture the social and collaborative reality of people working in STEM [17]. For example, bringing guests from STEM disciplines into the classrooms, inviting boys and girls into universities for research internships, or organizing Girls’ Days and similar events may help to align the mental models with reality. A further possibility are courses in school labs, such as the InfoSphere at RWTH Aachen University [6]. These extracurricular learning environments provide practical experiences of interacting with science and technology and may lower the barrier towards pursuing a career in STEM.

In summary, the responsibility of academia and industry must be to develop measures that address ill-aligned mental models, low interest in computer science and technology, as well as differences in self-efficacy at an even earlier age. As the Digital Revolution is reshaping every aspect of our society, the importance of these measures cannot be stressed enough: The ability to recognize, reflect, and understand computer science concepts will be one of the key skills of the 21st century. Besides increased employment prospects, the broad dissemination of technical self-efficacy and interest in computer science will be vital for participation in society and must therefore become a part of early childhood education.

6 SUMMARY, LIMITATIONS, AND OUTLOOK

The study revealed significant gender differences of sixth graders in regard to interest in computer science and self-efficacy in interacting with technology, as well as in regard to the image of computer scientists expressed through drawings. Although this study revealed interesting and relevant insights into the mental models of school children, it is limited and offers room for improvements by 1) assessing the current picture set with different criteria, 2) by extending the methodology to different work-contexts and domains, and by 3) including additional personality states and traits into the analysis.

Using the same set of images, different evaluation criteria might yield different or deeper insights into the children’s mental models of computer scientists. In addition, one could get a broader and more diverse view of the mental models by letting the children draw a set of different scenes that depict computer scientists at work. Currently, we have acquired the single and probably most dominant view on computer scientists, whereas a more fine-grained view on different work-contexts could provide us with a richer and more pronounced view of the children’s perspective of computer scientists. For example, we could ask to draw a computer scientist during a meeting, while discussing with a customer, or during a product roll-out. Also, the current study solely focused on computer scientist. Beside a transfer to different domains from STEM, the career and vocational choice community would profit from a broad analysis of children’s mental models across a multitude of different domains from science and engineering, over arts and philosophy, to medicine and elderly care. Lastly, the current study only addresses gender and technical self-efficacy as factors that might influence the children’s mental models. Subsequent studies with a larger sample size might also investigate other variables, such as the parents’ professions, the availability of toys with a technical character, etc. The findings from this and subsequent studies may also serve as a starting point for in-depth interviews or surveys that illuminate the perceived barriers and promoters on developing interest in STEM and other disciplines.

In summary, this article shows that school children as young as 11 years have stereotypical views of computer scientists that influence the development of career interest towards STEM.

ACKNOWLEDGMENTS

We thank all school children and the picture evaluators for their willingness to participate in this study. The active support of “Sean” Chantal Lidynia and Fabian Comanns is highly acknowledged. We would like to thank Hiltrud Westram for her endless dedication to ignite enthusiasm for computer science in children. The authors would also like to thank the anonymous referees for their very valuable comments and helpful suggestions. The work was funded by the German Research Foundation (DFG) as part of the German Excellence Initiative, the Cluster of Excellence “Integrative Production Technology for High Wage Countries” (EXC 128), and the German Federal Ministry of Education and Research within the project “TABULA” (16SV7574K).