# An Empirical Analysis of Newcomers' Contributions to Software-Engineering Conferences

Rand Alchokr[1][0000−0003−0112−5430], Jacob Krüger[2][0000−0002−0283−248X], Yusra Shakeel[1,3][0000−0001−5135−4325], Gunter Saake[1][0000−0001−9576−8474], and Thomas Leich[4][0000−0001−9580−7728]

[1] Otto-von-Guericke University, Magdeburg, Germany
{rand.alchokr | shakeel | saake}@ovgu.de
[2] Eindhoven University of Technology, Eindhoven, The Netherlands
j.kruger@tue.nl
[3] Karlsruhe Institute of Technology, Karlsruhe, Germany
[4] Harz University & Metop GmbH, Wernigerode & Magdeburg, Germany
tleich@hs-harz.de

**Abstract.** Newcomer researchers play a key role in advancing research: They introduce new ideas and perspectives, have a high motivation, and can positively impact the performance of long-lasting teams. However, newcomers face obstacles when engaging in research—some of which they can overcome based on learning and mentoring (e.g., using research methods, scientific writing), but also potential biases of other researchers or unfair barriers (e.g., gate keeping, perceived expertise). In this paper, we report a study on newcomers' contributions to three major software-engineering conferences, and what these contributions may indicate regarding potential obstacles. Precisely, we investigated to what extent newcomers contributed to the main tracks of the highly reputable software-engineering conferences ASE, ESEC/FSE, and ICSE, analyzing a total of 4,620 papers and 7,337 authors. Furthermore, we investigated whether the reviewing model impacted the extent of newcomers' contributions, since all three conferences recently switched from single-blind to double-blind reviewing. The results indicate a decline in newcomer researchers contributing to the conferences, a trend that somewhat stabilized in recent years at a fortunately high level (i.e., more than 50 % of authors for all conferences). Furthermore, for ICSE, we found an indicator that the changed reviewing model mitigated the declining trend, but this was not visible for the other conferences, and that more newcomers are involved in high-reputation papers.

**Keywords:** Software engineering · Newcomers · Peer review

## 1 Introduction

*Newcomers* are those researchers of a venue who got their first paper at that specific venue accepted (they may have papers at other venues before). They can

initiate transitions within a research community since they can be the key drivers for novel research and modernization in that community. Newcomers usually start with a high motivation that may inspire their peers, potentially broadening their knowledge and perspectives. For such and other reasons, it is essential to understand newcomers' involvement in the research community, analyze their contributions (i.e., papers), and try to reveal hidden obstacles regarding their involvement. Some previous studies found such obstacles [9,21], for instance, bias of experienced researchers against newcomers during single-blind peer reviewing. Identifying such obstacles in different research communities is highly important to derive appropriate countermeasures, and thus improve the quality of as well as fairness in research. Moreover, the obtained insights can be transferred to other communities, if required.

In this paper, we report a multi-case study in which we analyzed the contributions of newcomers to three major software-engineering conferences. Our idea is to study historical changes that may be caused by obstacles newcomers face, particularly with respect to bias during peer reviewing (all three conferences recently switched from single-blind to double-blind reviewing). Unfortunately, we can only study the contributions (i.e., accepted papers), since detailed submission data (i.e., author data of rejected papers) is not available. Still, we investigated a factor that has been rarely studied, but can bias a reviewer's decision [17]: the *reputation* in terms of papers published at these conferences and the academic age of the author (in terms of years the author published at any venue and not restricted to those we are studying). Studying these factors allows us to understand whether it has become more challenging for newcomers to publish papers on their own (i.e., without previous experience or established co-authors). Based on our data, we analyze whether there are any pattern changes in the contributions of newcomers, particularly after the switch to double-blind reviewing. For this purpose, we extracted all paper (4,620) and author (7,337 distinct) data for the main tracks until 2020 of the: 1) International Conference on Automated Software Engineering (ASE); 2) Joint European Software Engineering Conference and Symposium on the Foundations of Software Engineering (ESEC/FSE); and 3) International Conference on Software Engineering (ICSE). These conferences have a high reputation, which is why researchers of any academic age and reputation aim to publish there. The recent switches to double-blind reviewing (ASE 2016; ESEC/FSE 2017; ICSE 2018) make them ideal subjects to analyze the differences between experienced researchers and newcomers with respect to the reviewing model. We publish all of our data in an open-access repository.[5]

Since it is impossible to study rejected papers, we use proxy measures and must be careful with interpreting our results. Still, our findings reveal important insights concerning the contributions of newcomers at major software-engineering venues. For instance, the findings show a changing trend for at least one of the three conferences (i.e., ICSE) towards more opportunities for newcomers after adopting double-blind reviewing. We hope that our results motivate the software-engineering as well as other communities to investigate the obstacles

[5] https://doi.org/10.5281/zenodo.8369616

for different groups of researchers in more detail and that they help implement appropriate countermeasures.

## 2 Background and Related Work

Newcomers are new members of a research community. Specifically, they are new to a certain venue, for instance, authors who publish their first paper at a certain conference [23]. Typically, such newcomers have a high motivation that can inspire others, improve the work atmosphere, and introduce innovative ideas. So, it is essential to understand how they are involved in the research community. For instance, one experiment revealed the existence of explicit bias in the peer-reviewing process towards newcomers (or novices) by comparing the outcome of a review when the paper is written by a prominent author (a Nobel laureate) or by a relatively unknown author [13]. Furthermore, we identified a list of publishing impediments that hinder newcomers and juniors [2]. These indicate that new researchers may face challenges in contributing to a community, with studies like ours helping to reveal and tackle such challenges.

At software-engineering conferences, the program committees carry the responsibility of performing peer reviews, while also dealing with the rapid development of science, higher submission numbers, and the demand for increasing quality [28]. So, many program committees have evolved in recent years (e.g., becoming broader and larger), which directly impacts how new contributions are selected at established conferences. A peer review should be above all fair, meaning that any paper is "judged on the merit of one's ideas, not on the basis of academic rank, gender, place of work, publication record, and so on" [17]. Researchers in different communities are debating the pros and cons of reviewing models with respect to such properties, aiming to further improve the quality and fairness of reviews [4, 8, 10, 12, 25, 32]. To this end, the blinding of papers is an important and regularly discussed property of the reviewing models. Namely, some models reveal author and reviewer information to varying degrees to each other and the chairs [19, 27]. Typically, a review can be 1) un-blinded, which means that the reviewers and authors know each other; 2) single-blinded, which means that the reviewers know the authors, but not vice versa; or 3) double-blinded, which means that reviewers and authors are completely concealed [22, 24]. Other forms of blinding (e.g., triple-blind) enforce even stricter rules, but these three forms are the most relevant ones [11]. Researchers have found that various biases may occur within different reviewing models, particularly single-blind reviewing. This violates the overarching goal of an unbiased and fair review [14]. For instance, interviews with experts on the pros and cons of implementing double-blind reviewing at software-engineering venues showed that the challenges involved are mostly logistical, and are outnumbered by the benefits in terms of fairness for authors and their scientific progress [4]. More precisely, reviewers will judge a contribution more objectively based on its scientific value if the authors' names are concealed. Similarly, reports of different software-engineering conferences on introducing double-blind reviewing (e.g., ICSE'18 [6]), indicate the same percep-

tion among their reviewers. However, these insights are based on the personal perceptions of involved researchers, not actual data. In contrast, a study shows a significant increase in the acceptance rate of female first-authors in the Behavioral Ecology Journal after introducing double-blind reviewing [5]. Two studies to determine the acceptance rates under double-blind reviewing, particularly for junior researchers within the ACM SIGMOD community, show no differences between the acceptance rates under single-blind and double-blind reviewing for the first study [15]. However, the second study dug into more details and found that double-blind reviewing benefits the fairness towards newcomers [29].

Lastly, author characteristics that have the potential to influence reviewers' judgment have been collected [4, 14]. We are concerned with three of the 17 characteristics mentioned: 1) reputation within the community, 2) number of prior publications, and 3) the number of (co-)authors. Such characteristics can be primary sources of bias during single-blind reviewing. Building on the aforementioned studies, we investigate these three characteristics and connect them to the academic age of researchers at three software-engineering conferences. So, we aim to elicit novel and complementary insights into how these characteristics and introducing double-blind reviewing impact newcomers.

## 3 Methodology

Next, we describe the goal and conduct of our multi-case study.

### 3.1 Goal and Research Questions

In this paper, we focus on potential reputation biases, implying that reviewers favor papers (under single-blind or un-blinded reviewing) with highly prestigious authors. Related studies indicate that authors, for instance, from prestigious affiliations, with high publication records, or more overall visibility receive less critical reviews, more grant funding, and are cited more often [14, 17, 18, 26]. This implies that newcomers who do not have a high reputation may be negatively impacted if they do not collaborate with the right co-authors. With our multi-case study, we aim to understand how the contributions of newcomers at three major software-engineering conferences have evolved. The results of such an analysis can hint at certain biases and provide a good overall impression of newcomers' situation, even though proving actual biases is hardly possible, due to the anonymity implemented in most peer-reviewing models used (e.g., we cannot obtain data for rejected paper or the involved reviewers). Concretely, we defined two research questions (RQs) to guide our multi-case study:

$RQ_1$ *How has the proportion of newcomers at the conferences evolved over time?*
$RQ_2$ *Has adopting double-blind reviewing impacted newcomers?*

### 3.2 Data Collection

We display an overview of our entire data collection process in Figure 1. Next, we explain the individual steps we employed.

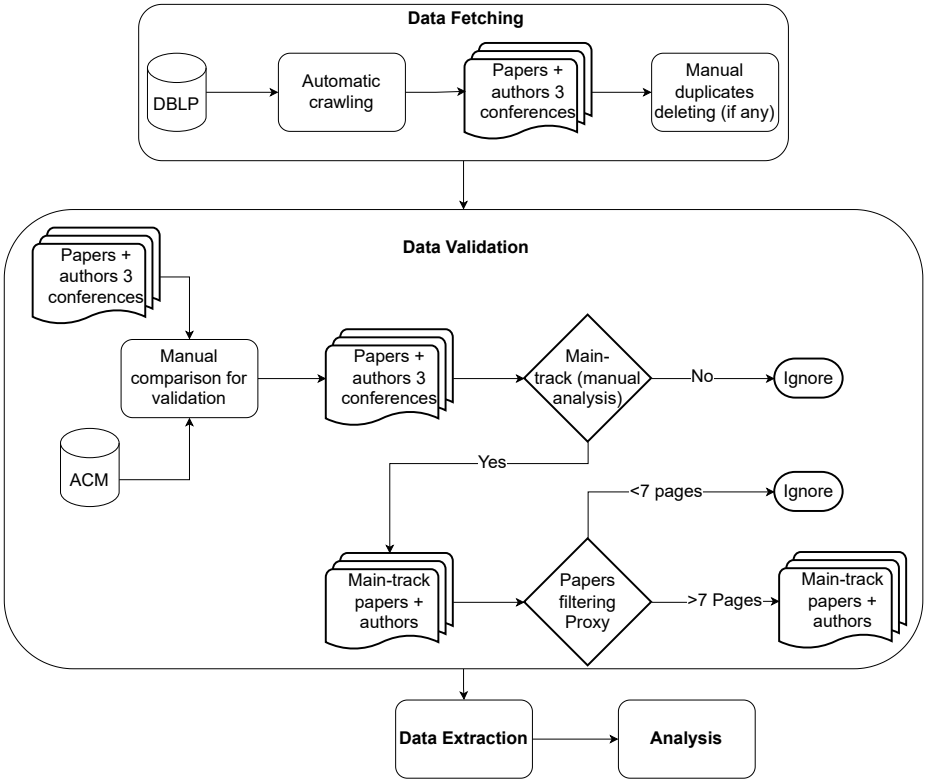



Fig. 1: Flowchart of our methodology.

**Subject Conferences.** We studied conferences since computer-science (and particularly software-engineering) research is generally more focused on those instead of journals compared to other communities [7, 16, 28]. Therefore, we analyzed the main tracks of three major software-engineering conferences (i.e., cases of our study): ASE, ESEC/FSE, and ICSE. We selected these conferences because they have a high reputation, indicating that newcomers and established researchers submit high-quality papers; and they recently introduced double-blind reviewing for their main tracks (ASE 2016; ESEC/FSE 2017; ICSE 2018). For our analysis, we considered all papers published since the first edition of each conference (ASE 1991; ESEC/FSE 1987; ICSE 1976) until 2020, the most recent year for which all data was available when we started our study.

**Data Fetching.** To collect our data, we automatically crawled *dblp*,[6] which provides bibliographic data structured (among others) by publication venues. We chose dblp because it covers all three conferences completely, is open-access, and has high data quality (e.g., distinguishing authors with the same names). Regarding ambiguities of author names, each author listed in dblp has a website that serves as an identifier to distinguish them from others with the same

[6] https://dblp.uni-trier.de/

Table 1: Overview of the data in our final dataset.

| Conference | Period | # Papers | # Authors | # Distinct Authors |
|---|---|---|---|---|
| ASE | 1991–2020 | 1,068 | 3,737 | 2,482 |
| ESEC/FSE | 1987–2020 | 1,252 | 4,312 | 2,614 |
| ICSE | 1976–2020 | 2,300 | 7,434 | 4,380 |
| Total | | 4,620 | 15,483 | 7,337 |

name. Even though this is likely not perfect, this property of dblp increases the robustness and reliability of our data.

**Data Validation.** To improve the comparability and quality of our data, we studied only main-track papers of all three conferences. For this purpose, we checked the extracted data manually and flagged each main-track paper (1 = main track, 0 = not main track). Then, we manually compared the labels to official information in the ACM Digital Library[7] to identify mislabeled papers; and matched the numbers of accepted papers to official statistics in the ACM Digital library (if available). We experienced that, particularly for older editions of the conferences, main-track papers are not clearly labeled across dblp and ACM. For this reason, we decided to enforce one more proxy criterion if the above validation was not conclusive: if we could not clearly label a paper as belonging to the main track, we excluded it if it comprised fewer than seven pages. Overall, our validation helped us to identify 4,620 main-track papers out of 11,106 papers in our initial dataset.

**Data Extraction.** For each paper, we extracted standard bibliographic data, namely its title, DOI, authors, publishing year, and page count. We used this data to validate the quality of our dataset and to investigate our research questions. For each author, we extracted their name, the identifier for their dblp website, as well as the first and last year they actively published at any venue (not only the three conferences we studied). We used a combination of paper (i.e., conference, year), author (i.e., start and end year), and combined (i.e., papers of an author) data to measure reputation and academic age (explained shortly). In Table 1, we summarize the properties of our final dataset. We extracted a total of 4,620 main-track papers that have been written by 15,483 authors. Note that this column (# Authors) includes authors multiple times if they wrote more than one paper. In contrast, we display the number of distinct authors who published at each conference's main track in the last column (7,337).

### 3.3 Data Analysis

**Newcomers.** Newcomers are researchers who are *new to a specific venue* [23]. So, for our study, we refer to a newcomer if the researcher did not publish a main-track paper before at that specific conference. We employ no restrictions on the academic age, meaning that a newcomer can be at any academic age. To get the number of newcomers at the main track of a conference (c) in a specific

[7] https://dl.acm.org/proceedings

year (y), we used an SQL query that is reflected by the following relation:

$$Newcomer_{y,c} \notin Authors_{year<y,c} \tag{1}$$

This relation fetches all authors in a specific year and conference who have not previously published at that conference.

**Academic Age.** The academic age is the time span for which a researcher has actively published [3]. We used this property to calculate an author's reputation. Consequently, we calculated the academic age individually for each author and published paper ($Year_{paper}$) based on the authors' first publication ($Year_{firstPaper}$) as follows:

$$Age_{academic} = Year_{paper} - Year_{firstPaper} + 1 \tag{2}$$

To exemplify our calculation, consider an author who published their first paper in 2010 and another one in 2012, thus the author's academic age when publishing this particular paper was three years.

**Author Reputation.** For each author in our dataset, we measured the reputation for a respective time and paper. There could be various formulations to measure an author's reputation. Some could include the author's own characteristics and related bibliometrics, such as the h-index or citation count. However, as we fetch the data from dblp, which lacks these metrics, and as we concentrate on newcomers of the three conferences, we consider the following equation to assess the reputation of an author on a specific paper:

$$Reputation_{author} = MT * Age_{academic}/Max(Age_{academic}) \tag{3}$$

$MT$ is the number of main-track papers the author has so far. Note that an author can have different reputation values for individual papers, depending on the academic age and previous productivity. We computed reputation based on the number of main-track papers an author published at *all three conferences* at that point in time. For comparability, we normalized the academic age based on the highest one in our dataset (54 years).

**Paper Reputation.** As paper reputation we refer to the sum of all reputations of its respective authors. We choose this measure to capture the co-authorships effect on newcomers and to investigate the level of collaboration newcomers have with other authors who have different reputation levels, resulting in high, average, or low reputation papers. In the previous Equation 3, we calculate one author's reputation. Now, we judge all authors as a whole rather than by picking one isolated authors. We calculate each paper's reputation separately by accumulating its authors' reputations as follows:

$$Reputation_{paper} = \sum_{author=1}^{N} Reputation_{author} \tag{4}$$

So, the result of Equation 3 reflects the reputation of a paper.

## 4 Results and Discussion

In the following, we report and discuss the results of our analysis structured based on our research questions. We managed and analyzed our data using the

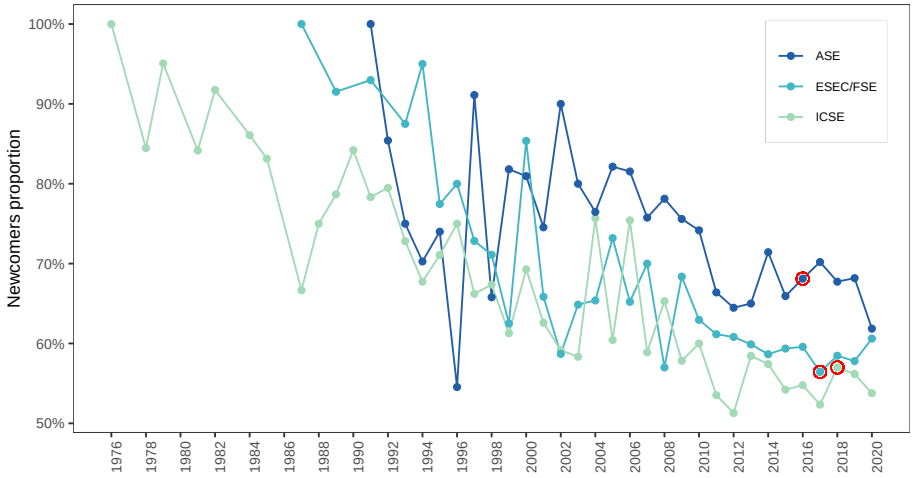



Fig. 2: Proportion of newcomers accepted each year at the three conferences.

R statistics environment using different libraries [20]. Note that in most areas we focus on describing our actual data and observations instead of statistical tests, which can be misinterpreted and misleading [30, 31].

### 4.1 $RQ_1$: Proportion of Newcomers

**Measurements.** To understand the extent of newcomers' involvement in the conferences over time ($RQ_1$), we measured for each conference individually the ratio of newcomers. We used Equation 1 to compute each conference's newcomers. Note that we look at newcomers by considering the number of papers they are involved in regardless of their position on the paper. This measurement provides an intuition as to what degree a conference's community involves and accepts newcomers, even though we cannot consider the submission data.

**Results.** In Figure 2, we display the proportion of authors who published their first main-track paper at each of the three conferences separately (i.e., newcomers). The red circles mark the years in which each conference adopted double-blind reviewing. As we would expect, the proportion of newcomers starts high and gets lower over time, somewhat stabilizing around 2010 for all three conferences. The high variance particularly in the earlier years may be caused by various reasons, such as the conference's locations, policy changes, or gained reputation. Interestingly, the proportion of newcomers never dropped below 50 % for any conference. Note that the empty data points for the early phases of ICSE (no regular/annual editions) and ESEC/FSE (ESEC started 1987 bi-yearly before FSE in 1993) are not missing, but the conferences did not happen.

We support these observations with statistical analyses to illustrate the stability of newcomers' proportions after 2010 at the three conferences combined. To this end, we calculated the standard deviation and variance to capture how far from the normal and how spread out these proportions are:

Table 2: Overview of the proportion of newcomers each year across the three conferences combined.

| | | Co(1999 & older) | Co(2000–2009) | Co(2010–2020) |
|---|---|---|---|---|
| +1STD (94.82%) | | | | |
| | STD1⟸ | 16 | 3 | 0 |
| +2STD (82.79%) | | | | |
| | STD2⟸ | 15 | 11 | 2 |
| Mean 70.76% | | | | |
| | STD3⟸ | 8 | 13 | 20 |
| -1STD (58.69%) | | | | |
| | STD4⟸ | 1 | 3 | 11 |
| -2STD (46.69%) | | | | |

Co = Number of conferences accruing at that period of time, STD = Standard Deviation

1) First, we calculated the mean: we summed up all proportions of newcomers at the three conferences in all years and divided that by the number of conference occurrences (102), resulting in a mean of 70.76 %.
2) Second, we calculated the standard deviation of 12.03 %.

For the standard deviation, we used:

$$Standard\_Deviation = \sqrt{\frac{1}{N-1}\sum_{i=1}^{N}(x_i - \overline{x})^2} \tag{5}$$

with $N$ being the total number of conference occurrences (102) and $x$ the newcomers' proportions for each occurrence. Our results show which proportions are within one standard deviation of the mean, enabling us to establish a notion of what is typical, what is more than the typical proportion, and what is less. We illustrate the results in Table 2.

**Observations.** From Figure 2 and Table 2, we can observe:

$O_1$ The overall trend is a decline in the proportion of newcomers.

$O_2$ Except for outliers, and particularly since around 2010, older conferences have a lower ratio of newcomers with accepted papers.

$O_3$ In contrast, the age of a conference apparently does not cause the proportions of newcomers to stabilize (i.e., they stabilize around 2010 for all conferences).

$O_4$ The introduction of double-blind reviewing seems to have a slight impact on the proportion of newcomers.

**Discussion.** Considering $O_1$, it is obvious that all conferences start with a 100 % proportion of newcomers that then decreases. However, it is positive that more than 50 % of the authors in each year published their first main-track paper at the corresponding conference. While the ongoing decline is concerning, combining this observation with the fact that more authors publish at the conference every year indicates that the absolute number of newcomers involved in each conference also increases, but more and more work seems to be driven by collaborations between experienced researchers [1,33]. $O_2$ supports this finding, indicating that a conference's growing community naturally decreases the proportion of newcomers. Therefore, we would argue that over time it becomes harder for newcomers to get their papers accepted, which may be caused by higher quality standards or a shift

in what is perceived as valuable research at each conference. Thus, newcomers first have to learn how to perform research on the expected quality. However, it may also be the result of gatekeeping and biases towards newcomers, since they lack a high reputation. Interestingly, these observations somewhat contradict $O_3$, which indicates that the proportions of newcomers stabilized at the same time for all conferences. Our results (cf. Table 2) show that most conferences held in the years from 2010 to 2020 had a proportion of newcomers that falls mostly within one standard deviation (12.07%) of the mean (20 conferences), whereas older conferences had more distributed or spread out proportions. This may indicate that the software-engineering community became more collaborative and international overall, or that the borders between its sub-communities and their conferences started to vanish (e.g., in terms of topics or perceived reputation). It may also indicate that the reputation and quality of the conferences could have a negative impact on the proportion of newcomers, since they may be discouraged to submit to highly selective venues. Still, we cannot explain this phenomenon in detail with our analysis, but argue that it deserves further research.

Finally, $O_4$ may indicate that introducing double-blind reviewing did only have a slight improvement on the proportion of newcomers. We can see that for ICSE and ASE the proportions of newcomers immediately increased; however, the proportion dropped low in 2020 for ASE (we shortly analyze this situation in more detail). Considering our previous observations, we actually argue that such stable or slightly increased proportions are already a success, since it means that the total number of newcomers increases.

### 4.2 $RQ_2$: Double-Blind Reviewing

Since our analysis, for now, was on an abstract level and did not consider other factors (e.g., the reputation of coauthors), we continue with a detailed analysis of the time period around the introduction of double-blind reviewing. More precisely, we analyze the period with the stabilized proportions (2010–2020) as well as the same number of years before and after introducing double-blind reviewing at each conference (2011–2020 for ASE, 2013–2020 for ESEC/FSE, 2015–2020 for ICSE). So, we aim to obtain a better understanding of the impact of double-blind reviewing. Therefore, we explore the conjecture that when the identity of the authors is revealed, their evaluation will be affected by the previous productivity.

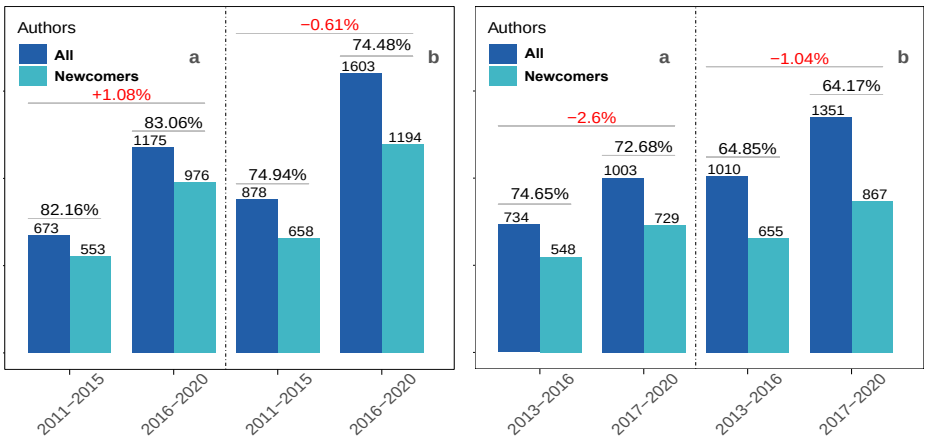
**Measurements.** We measured for each conference

- the proportion of newcomers at the main tracks in an equal period of time before and after double-blind reviewing has been adopted; and
- the proportion of newcomers involved in papers with a specific overall reputation (i.e., considering the reputation of coauthors);

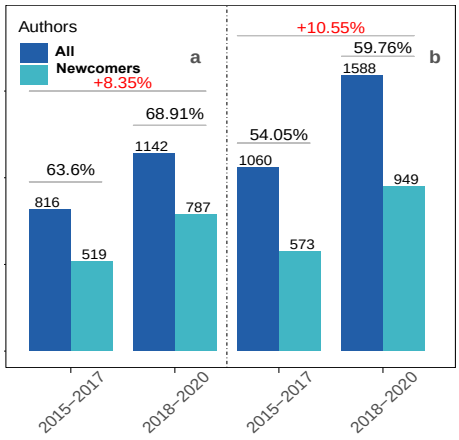
For the first measurement, we calculated the number of newcomers using Equation 1 at each conference during the specified period of time and compared it to the number of all authors at the same period. Then, we measured the percent change and whether its an increase or decrease by comparing the old percentage value to the new value using Equation 6 ($Val$: value, $abs$: absolute value):

$$Perc\_Change = (second\_Val - first\_Val)/abs(first\_Val) * 100\% \quad (6)$$

A positive percent change indicates an increase in the proportion of newcomers, whereas a negative change represents a decrease.




(1) ASE (2) ESEC/FSE




(3) ICSE

Fig. 3: Authors' contributions before and after double-blind reviewing has been introduced, distinguished by (a) the number of distinct authors and (b) number of paper contributions in the main tracks.

For the second measurement, we used Equation 4 to elicit the reputation of each paper. We exported the results into separate spreadsheets to manually calculate the top and bottom 25 % of papers in terms of their authors' reputation. Also, we elicited the proportion of newcomers each year within these two subgroups of papers.

**Results.** In Figure 3, we present an overview of the proportion of newcomers at each conference before and after double-blind reviewing has been adopted. We collected the data of each conference for a time span of 10, 8, and 6 years for ASE, ESEC/FSE, and ICSE, respectively. The three plots in Figure 3 show the

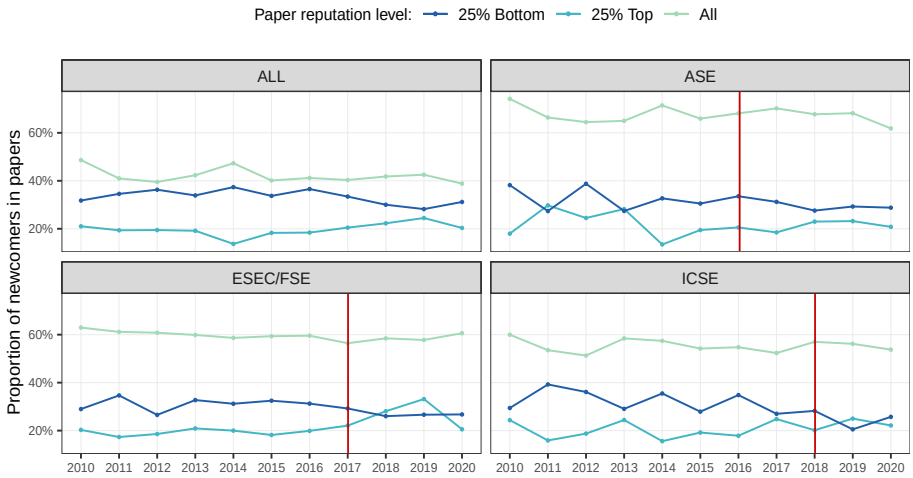



Fig. 4: Proportion of newcomers compared to papers' reputation.

proportion of newcomers in main-track papers. We distinguish two categories: **a)** the distinct number of authors in the main tracks; and **b)** the number of author's papers (since an author can have more than one paper). We perform this detailed analysis to gain more knowledge on the proportion of papers by newcomers compared to the overall research community. The results indicate a slight increase in the proportion of newcomers after switching to double-blind reviewing. To investigate the reasons behind this in more detail, we checked the **paper's reputation** based on coauthors' **academic age**. First, using Equation 4, we calculated each paper's reputation to study the bottom and top 25 % of papers. Then, we analyzed the number of newcomers in each of these categories (top 25%, bottom 25%) to understand their involvement. We display the corresponding results in Figure 4. As we can see, the proportion of newcomers involved in high-reputation papers increased after double-blind reviewing was introduced until a certain point, at which it declined again. Interestingly, newcomers seem to be underrepresented at the bottom and even more in high-reputation papers, but for ESEC/FSE and ICSE both groups became more similar after introducing double-blind reviewing.

**Observations.** From Figure 3 and Figure 4, we observe:

$O_6$ The proportion of newcomers after adopting double-blind reviewing at ICSE increased, whereas it remained stable at ASE, and decreased at ESEC/FSE.

$O_7$ More newcomers are involved in high-reputation papers after adopting double-blind reviewing.

**Discussion.** As we can see in Figure 3 ($O_6$), the proportions of newcomers contributing to the conferences vary. Namely, ICSE exhibits the lowest proportion of newcomers contributing in either of the two time periods (before and after double-blind reviewing) compared to ESEC/FSE and ASE. Still, considering the percentage change of newcomers in the considered time periods, surprisingly, ICSE has the highest increase. Notably, ASE has a remarkably high proportion of

newcomers, but it still exhibits a slight increase in the proportion of newcomers after introducing double-blind reviewing. Interestingly, the proportion dropped slightly for ESEC/FSE, which indicates that newcomers got fewer papers accepted under double-blind reviewing. These different trends may have various root causes that are subject to future work. For instance, the results for ICSE could indicate that it was particularly challenging for newcomers to publish there, but apparently only while author identities were revealed. This may be caused by potential biases towards newcomers, due to their missing reputation, or because of gatekeeping. In contrast, the ASE community apparently was always more inclusive, and thus the effect was smaller. However, we cannot explain the actual root causes for these trends, which require more analysis. To initiate research in that direction, we concentrated on investigating the potential impact of the academic age and reputation of coauthors. Regarding $O_7$, we can see in Figure 4 that the number of newcomers in top publications (high reputation authors' papers) increased slightly. We can observe the strongest change for ESEC/FSE, which increased to up to 33 % in 2019, the highest it has been in the covered period for the conference. In contrast, ASE and ICSE remained rather stable, even though there was a higher variance at the beginning of the covered periods. It is highly interesting that the overall proportion of newcomers contributing to high-reputation papers increased, while it decreased for bottom-reputation papers. This may indicate that more experienced and reputable researchers have become more encouraged or willing to collaborate with newcomers or to push their work towards high-reputation conferences after these introduced double-blind reviewing. Another reason for this trend may be the fact that newcomer researchers are less concerned about reviewers rejecting their papers simply for missing reputable authors—which may be not only an assumption but an actual positive effect of double-blind reviewing. Still, our findings are only a first step towards understanding the actual reasons. It is definitely important to better understand how newcomers are involved in papers, and whether potential biases of reviewers may transfer to coauthors.

## 5 Threats to Validity

Peer reviewing is a confidential process for outsiders who are not involved, meaning that we have no data on the number of submitted papers and their authors. Consequently, we can only use data from accepted papers, making our study vulnerable to survivorship bias. This also means that we cannot account properly for years with a lot of submissions of newcomers or with overall low acceptance rates, challenging our ability to derive insights on actual biases towards newcomers. To tackle such problems, we make the following reasonable assumption (considering the speed and growth of software-engineering research): If there is no bias towards newcomers, their prevalence in the accepted papers should stay stable to some degree from a certain point in time onwards. Building on this assumption, we derived measurements to reflect on the contributions of newcomers at the three conferences. Another threat to validity could be the metrics we used to compute the academic age and reputation, as these may be calculated differently by other researchers depending mainly on their perspective. Additionally, some

parts of our data may be incorrect. As we described, we extracted our data from dblp, which is a reliable data source. However, throughout the conferences' editions, the labeling of sessions and papers changed and during our validation, we found several mismatches between dblp and the ACM Digital Library. Since we focus most of our study on a more recent period, we argue that this threat is marginal for our actual findings—seeing that the data quality for more recent years is considerably better and more consistent.

Lastly, we focused our analysis on three software-engineering conferences. However, there are numerous other venues that researchers submit to, and that can impact a researcher's reputation the same way as these three conferences. So, our findings may be skewed toward a certain part of the software-engineering community. Still, we involved a large dataset in our analysis and derived several measurements to investigate our research objectives in a reliable way. Since other researchers may derive different findings, we make our dataset and analysis scripts publicly available.

## 6 Conclusion

In this paper, we reported a multi-case study on the involvement of newcomers in software-engineering research. For this purpose, we studied their contributions to three of the most prestigious software-engineering conferences, ICSE, ESEC/FSE, and ASE. Through analyzing the data of 4,620 main-track papers and their 7,337 authors, we were able to measure the involvement of newcomers, who have not previously published papers at the conferences' main tracks. From our data, we observed a decline in newcomers' overall proportion as authors at the conferences. Moreover, we analyzed the impact of double-blind reviewing, which was recently adopted by these venues, on newcomers. Even though only slightly, double-blind reviewing appears to have a positive impact on newcomers, mitigating or even reversing the observed decline to some extent. More precisely, our results indicate an increase in the proportion of newcomers at ESEC/FSE and ICSE, particularly in high-reputation papers after introducing double-blind reviewing.

As discussed before, our findings are only a first step in a comprehensive research topic concerning newcomer researchers. Further studies are required to explain our findings in more detail, potentially confirming or refuting some of them. A particular problem is the complexity of peer reviewing and potential biases that may occur in the process, involving numerous human factors that may be hard to impossible to study. So, for future research, we already started to expand our study with a more comprehensive dataset including additional years, mainly 2021, 2022, and later 2023. We believe this will provide a clearer look into double-blind reviewing and its impact. Furthermore, we plan to investigate other obstacles newcomers face and concentrate on different review models in more detail. Ideally, we could use actual submission data. Finally, approaching community members via survey research and experiments to corroborate our interpretations of the data with community perceptions, concerns, and experiences is an important direction to shed light on the research community.

## References


1. Alchokr, R., Krüger, J., Shakeel, Y., Saake, G., Leich, T.: A closer look into collaborative publishing at software-engineering conferences. In: International Conference on Theory and Practice of Digital Libraries, (TPDL). Springer (2022)
2. Alchokr, R., Krüger, J., Shakeel, Y., Saake, G., Leich, T.: Peer-reviewing and submission dynamics around top software-engineering venues: A juniors' perspective. In: International Conference on Evaluation and Assessment in Software Engineering (EASE). ACM (2022)
3. Alchokr, R., Krüger, J., Shakeel, Y., Saake, G., Leich, T.: On academic age aspect and discovering the golden age in software engineering. In: International Conference on Cooperative and Human Aspects of Software Engineering (CHASE). ACM/IEEE (2022)
4. Bacchelli, A., Beller, M.: Double-Blind review in software engineering venues: the community's perspective. In: International Conference on Software Engineering (ICSE). IEEE (2017)
5. Budden, A.E., Tregenza, T., Aarssen, L.W., Koricheva, J., Leimu, R., Lortie, C.J.: Double-blind review favours increased representation of female authors. Trends in Ecology & Evolution **23**(1) (2008)
6. Chechik, M., Harman, M., Zimmerman, T., Crnkovic, I.: ICSE 2018 PC Chairs Report (2018)
7. Chen, J., Konstan, J.A.: Conference paper selectivity and impact. Communications of the ACM (2010)
8. Cox, A.R., Montgomerie, R.: The cases for and against double-blind reviews. PeerJ **7** (2019)
9. Diem, A., Wolter, S.: The use of bibliometrics to measure research performance in education sciences. Research in Higher Education **54**(1) (2013)
10. Fox, C.W., Paine, C.E.T.: Gender differences in peer review outcomes and manuscript impact at six journals of Ecology and Evolution. Ecology and Evolution **9**(6) (2019)
11. Görögh, E., Ross-Hellauer, T.: Guidelines for open peer review implementation (2019)
12. Helmer, M., Schottdorf, M., Neef, A., Battaglia, D.: Gender bias in scholarly peer review. eLife **6** (2017)
13. Huber, J., Inoua, S., Kerschbamer, R., König-Kersting, C., Palan, S., Smith, V.L.: Nobel and novice: Author prominence affects peer review. Proceedings of the National Academy of Sciences **119**(41) (2022)
14. Lee, C.J., Sugimoto, C.R., Zhang, G., Cronin, B.: Bias in peer review. Journal of the American Society for Information Science and Technology **64**(1) (2012)
15. Madden, S., DeWitt, D.: Impact of double-blind reviewing on sigmod publication rates. SIGMOD Record **35** (06 2006)
16. Meyer, B., Choppy, C., Staunstrup, J., van Leeuwen, J.: Viewpoint research evaluation for computer science. ACM **52**(4) (2009)
17. Peters, D.P., Ceci, S.J.: Peer-Review practices of psychological journals: The fate of published articles, submitted again. Behavioral and Brain Sciences **5**(2) (1982)
18. Petersen, A.M., Fortunato, S., Pan, R.K., Kaski, K., Penner, O., Rungi, A., Riccaboni, M., Stanley, H.E., Pammolli, F.: Reputation and impact in academic careers. Proceedings of the National Academy of Sciences **111**(43) (2014)
19. Prechelt, L., Graziotin, D., Fernández, D.M.: A community's perspective on the status and future of peer review in software engineering. information and software technology. Journal of Computational and Graphical Statistics **95** (2018)

20. R Core Team: R: A Language and Environment for Statistical Computing (2018), https://www.R-project.org
21. Rørstad, K., Aksnes, D.: Publication rate expressed by age, gender and academic position – a large-scale analysis of norwegian academic staff. Journal of Informetrics **9**(2) (2015)
22. Saini, J.R., Sonthalia, N.R., Dodiya, K.A.: Identification of author and reviewer from single and double blind paper. International Journal of Computer and Information Engineering **8**(2) (2014)
23. Seeber, M., Bacchelli, A.: Does single blind peer review hinder newcomers? Scientometrics **113** (2017)
24. Snodgrass, R.: Single-versus double-blind reviewing: An analysis of the literature. ACM Sigmod Record **35**(3) (2006)
25. Soldani, J., Kuhrmann, M., Pfahl, D.: Pains and gains of peer-reviewing in software engineering. ACM SIGSOFT Software Engineering Notes **45** (2020)
26. Tahamtan, I., Safipour Afshar, A., Ahamdzadeh, K.: Factors affecting number of citations: a comprehensive review of the literature. Scientometrics **107** (2016)
27. Tomkins, A., Zhang, M., Heavlin, W.D.: Reviewer bias in single-versus double-blind peer review. Proceedings of the National Academy of Sciences **114** (2017)
28. Tran, H., Cabanac, G., Hubert, G.: Expert suggestion for conference program committees. In: 2017 11th International Conference on Research Challenges in Information Science (RCIS) (2017)
29. Tung, A.K.H.: Impact of double blind reviewing on sigmod publication: a more detail analysis. SIGMOD Rec **35** (2006)
30. Wasserstein, R.L., Lazar, N.A.: The ASA statement on p-Values: context, process, and purpose. The American Statistician **70** (2016)
31. Wasserstein, R.L., Schirm, A.L., Lazar, N.A.: Moving to a world beyond "$p < 0.05$". The American Statistician **73** (2019)
32. Weller, A.C.: Editorial peer review for electronic journals: current issues and emerging models. Journal of the American Society for Information Science **51** (2000)
33. Wren, J.D., Kozak, K.Z., Johnson, K.R., Deakyne, S.J., Schilling, L.M., Dellavalle, R.P.: The write position. EMBO reports **8**(11) (2007)