

Methodology Matters: How We Study Socio-Technical Aspects in Software Engineering

COURTNEY WILLIAMS, University of Victoria, Canada
 MARGARET-ANNE STOREY, University of Victoria, Canada
 NEIL A. ERNST, University of Victoria, Canada
 ALEXEY ZAGALSKY, University of Victoria, Canada
 EIRINI KALLIAMVAKOU, University of Victoria, Canada

Software engineering involves the consideration of both human and technical aspects, and although its origins come from the sub-disciplines of computer science and engineering, today the importance of the social and human aspects of software development are widely accepted by practitioners and researchers alike. Researchers have at their disposal many research methods they can choose from, but does software engineering research, at a community level, use methods that adequately capture the social and human aspects of the socio-technical endeavour that is software development? To answer this question, we conducted a categorization study of 253 ICSE papers and found a bigger emphasis on computational studies that rely on trace data of developer activity, with fewer studies controlling for human and social aspects. To understand tradeoffs that researchers make among their choice of research methods, we conducted a follow-up survey with the authors of the mapping study papers and found they generally prioritize generalizability and realism over control of human behaviours in their studies, sometimes for reasons of convenience or to appease reviewers of their papers. Furthermore, our findings surprisingly suggest a gap in knowledge about triangulation that could help address this gap within our community. We suggest our community, as a whole, diversify its use of research methods, to increase the use of methods that involve more control of the human and social aspects of software development practice while balancing our understanding of innovations on the technical side.

CCS Concepts: • **Computer systems organization** → **Embedded systems**; *Redundancy*; Robotics; • **Networks** → Network reliability.

Additional Key Words and Phrases: empirical methods, human studies, software engineering, meta-research, survey

ACM Reference Format:

Courtney Williams, Margaret-Anne Storey, Neil A. Ernst, Alexey Zagalsky, and Eirini Kalliamvakou. 2019. Methodology Matters: How We Study Socio-Technical Aspects in Software Engineering. *ACM Trans. Softw. Eng. Methodol.* 37, 4, Article 111 (August 2019), 22 pages. <https://doi.org/10.1145/1122445.1122456>

1 INTRODUCTION

Software engineering is at the forefront of innovation and research, and involves the consideration of both human and technical aspects. The origins of software engineering come from the 1950s

Authors' addresses: Courtney Williams, courtneywilliams@gmail.com, University of Victoria, Canada; Margaret-Anne Storey, mastorey@uvic.ca, University of Victoria, Canada; Neil A. Ernst, nernst@uvic.ca, University of Victoria, Canada; Alexey Zagalsky, alexeyza@uvic.ca, University of Victoria, Canada; Eirini Kalliamvakou, ikalliam@uvic.ca, University of Victoria, Canada.

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.

© 2019 Association for Computing Machinery.
 1049-331X/2019/8-ART111 \$15.00
<https://doi.org/10.1145/1122445.1122456>

and 1960s when the field emerged as a sub-discipline of computer science and engineering. As such, it was highly technical and focused on solving technical, logical and mathematical problems. Gradually, however, seminal works drew attention to the role of developers and social factors in software engineering [2, 4, 21, 32]. Nowadays, software development is recognized as a socio-technical endeavour [33], and many researchers consider both technical and human aspects of software development in their work.

To study human aspects, researchers can take advantage of several specific research methodologies that make use of empirical methods suitable for studying these aspects. Qualitative methods, described by Seaman *et al.* [17], are particularly valuable in highlighting human aspects. Sharp *et al.* [18] advocate using **ethnography** in software engineering studies, pointing to its potential to capture what developers do in practice and *why* they follow certain processes. Kitchenham [7] calls for incorporating approaches from social science, such as **case studies** and **quasi-experiments**, as she argues it will make findings more relevant to practitioners. Sjøberg, Dybå, and Jørgensen [26] argue that doing more empirical work in SE will provide us with the knowledge needed to develop better technologies for software development. And although empirical research in software engineering has increased, how many of our studies directly study developers or other stakeholders?

To understand how human aspects are studied (or not), we conducted a meta-study examining the *research strategies* and *data sources* reported in a cohort of papers published at the International Conference on Software Engineering (ICSE). We consider ICSE because it is seen by many as the flagship conference in software engineering and purportedly represents the breadth of software engineering research. Many of the papers presented at ICSE propose or evaluate technical tools and/or descriptive or predictive theories. We could expect that many ICSE papers would not directly study developers and may indirectly study developer behaviour through simulation or by mining and manipulating developer trace data, but at a community level we would also expect to find other papers that do directly study developer behaviours and evaluate new tools and interventions in real work developer work contexts.

We use Runkel and McGrath's research framework [16], originally developed to guide research on human behavior in psychology and sociology, as a lens to understand how research produced by the SE community captures *human and social perspectives*. McGrath [14] saw research methods as "*bounded opportunities*"—whereby choosing a specific method provides opportunities not available with other methods, but also introduces inherent limitations. Their model emphasizes that research strategy choice involves trade-offs in *generalizability*, *realism*, and *control*. By control, McGrath refers to control over the human subjects being studied. To address the inherent tradeoffs from one method to another, Runkel and McGrath recommend *triangulation* as the best mitigation strategy. The McGrath model has been used by other software engineering researchers to reflect on method choice and the implication of that on research design [5].

We consider the following research questions:

- RQ1:** What kinds of empirical studies are reported in papers submitted to ICSE? More specifically, we ask:
- 1.1: What *research strategies* are described in research published at ICSE?
 - 1.2: What *data sources* are described in research published at ICSE?
 - 1.3: How does the research presented in these papers prioritize *generalizability*, *control*, and *realism*?
 - 1.4: How is *triangulation* used in light of the prioritization of generalizability, control, and realism criteria?
- RQ2:** What are the *contributing factors* that led to method choice described in these papers?

To answer these questions, we conducted a two-phase study. First, we categorized 253 papers published in the technical track at ICSE (accepted papers over a recent 3-year period). As mentioned above, we chose to study ICSE because it is highly regarded and is assumed to not focus on a particular set of sub-topics or research methods. We classified the research strategies and data sources used in the research described in these papers according to Runkel and McGrath's model. From the perspective of focus on human and technical aspects, this phase revealed a tendency towards certain methods and data sources, and called for additional information about why researchers chose those methods. In the second phase of our study, we surveyed the papers' authors, asking them to classify their own papers based on the same terminology. We also asked the authors to reflect on their method choice, use of triangulation, and desirable research criteria they wished to achieve in their research.

In the papers we considered, we found a skewed use of research strategies and data sources, and that software engineering researchers prioritize generalizability and realism in their studies, for reasons including convenience or to satisfy reviewers' expectations. Our observations signal that software engineering studies, at the methodology level and at least in some publishing venues, may not adequately capture human and social aspects in software engineering. While triangulation is the recommended mitigation strategy, our findings surprisingly suggest a gap in usage of and knowledge about triangulation within our community.

The remainder of this paper is structured as follows. In Section 2 we discuss related work that has both informed and motivated this research. This is followed by Section 3 where we describe the methodology for both the categorization study and our survey. In Section 4 we present the findings of our studies, and discuss possible explanations for the results that are grounded in our data in Section 5. We also offer some recommendations for our research community to consider. We discuss limitations and threats to validity of this research in Section 6. Finally, we conclude by identifying areas for future work and reiterating important takeaways in Section 7. A number of traceability artifacts from our analysis and a replication package are published on our supplementary website at <https://bit.ly/2vKxXvg>.

2 BACKGROUND

Software development is a highly complex and technical process, and developers utilize a number of different technologies to design, develop, deploy, and maintain software. While much of our research is technical, the importance of considering human and social factors of software development has been recognized since the early days of software engineering. Books such as *The Psychology of Computer Programming* [32] and *Software Psychology: Human Factors in Computer and Information Systems* [21] drew attention to the role of developers and social factors in software development. Other authors drew on personal experiences to demonstrate the impacts of different social constructs and management practices in software development in their books, including *The Mythical Man-Month* [2] and *Peopleware: Productive Projects and Teams* [4]. While these are only a few of the many examples of early social research in software engineering, they still have impact today.

To study complex socio-technical systems, researchers in software engineering must employ a wide variety of techniques from a number of interdisciplinary fields. There are a number of seminal works that provide guidance for conducting and reflecting on empirical research in software engineering. One key example is the book *Empirical Methods and Studies in Software Engineering* [3], published in 2003. It offers an introduction of four major empirical methods: "controlled experiments, case studies, surveys, and post-mortem analyses" [34]. Another prominent research book, published in 2007, is the *Guide to Advanced Empirical Software Engineering* [22]. This book includes guidance for a number of specific techniques, including qualitative methods [17], focus groups [10], personal opinion surveys [8], and data collection techniques for field studies [24]. It also provides guidance

on general topics, such as how to design ethical studies involving humans in software engineering [31], a guide for building theories in software engineering [25] and a chapter explaining the benefits and drawbacks of different empirical methods in software engineering to assist in research design choices [5].

Other researchers offer guidance on specific methods for studying human subjects: Stol, Ralph, and Fitzgerald provide guidelines for grounded theory specifically in the context of software engineering [29] as they found that many papers that reported they used grounded theory lacked rigor. While Runeson and Høst adapt case study research guidelines to the software engineering domain [15], also in part to address the misuse of the term case study in our community. There are also a number of seminal works available that focus around experimentation and evaluations. Both Wohlin *et al.* [35] and Ko, Latoza, and Burnett [9] provide excellent resources for understanding how to conduct software engineering experiments with human participants. Sharp *et al.* [18] recently explained how ethnographic studies could show not only what developers do in practice but also *why*, and encouraged SE researchers to incorporate ethnography into their empirical studies.

These methods for directly studying human activities and behaviours are used across our community. Typically, there is at least one track on human aspects in the main research conferences, as well as special purpose workshops on the topic such as the CHASE series ¹. The papers presented at CHASE tend to address broad socio-technical topics but as a workshop focus on early results. The ESEM conference and EMSE journals also attract papers that consider human aspects, as their focus is on empirical methods of which many involve direct human involvement. But how frequently human aspects are considered in our main venues, particularly in papers that present technical innovations, is not at all evident. And some researchers feel that the coverage of human and social aspects is lacking [12, 19].

There are several meta-studies that reflect on papers published in our community and our use of empirical methods. Shaw [20] investigated the papers submitted to ICSE 2002, analyzing the content of the papers that were both accepted and rejected, as well as observing program committee conversations about which papers to accept. She found that there were very low rates of submission and acceptance of papers that investigated “categorization” or “exploration” research questions, or papers whose research results presented “qualitative or descriptive models”. A 2016 replication of Shaw’s methodology [30] showed that since 2002, reviewers have raised their standards, particularly with regards to empirical evaluations of research contributions. This is a good sign that empirical research is increasingly prominent in SE. The replication study also found that a new category of research papers, mining software repositories, was incredibly common. This new category of mining papers may study human behaviours, but often in an indirect way.

Zelkowitz [36] found that the community’s use of empirical validation techniques for research contributions was improving, but that researchers were using terms such as “case study” to refer to different levels of abstraction, making it hard to understand the communicated research. More recently, Siegmund *et al.*’s work [23] prompted discussions about validity within our community. Another recent paper, published by Stol and Fitzgerald, also builds on Runkel and McGrath’s research framework and use it to provide consistent terminology for research strategies. However, they adapt the dimension of control (using the term precision) to mean control over the study variables, rather than control of the human participants in the study [28]. They adapt the research framework to categorize research studies, but do not focus on human aspects.

These papers, although introspective about empirical research in our community, do not tease out how or to what extent social and human aspects are studied. Our aim is to understand how the software engineering research community currently approaches studying *human aspects* in

¹Cooperative and Human Aspects of Software Engineering, colocated with ICSE since 2011 <http://www.chaseresearch.org/>

software engineering. The software engineering landscape is constantly changing with the creation of new technologies and it has shifted in recent years with the addition of platforms and tools that make software development more collaborative and social (GitHub, StackOverflow, Continuous Integration and Slack are prominent examples). Therefore, we feel that an investigation of how social aspects are captured and discussed in current software engineering research is both relevant and timely.

3 METHODOLOGY

To investigate the research questions described in Section 1, we conducted a two-phased study. We manually analyzed and categorized three years of ICSE papers, and then followed this with a survey of the authors of those ICSE papers. We provide our methodological tools, anonymized raw data, and analysis documents on our supplementary website, <https://bit.ly/2vKxXvg>, for the purposes of replication and traceability.

3.1 Categorizing ICSE Paper Research Methods

To address RQ1, we manually analyzed ICSE technical research papers. We considered all technical research track papers from ICSE's 2015, 2016, and 2017 proceedings in the sample, collecting 84, 101, and 68 technical track papers from each year, respectively, for a total of 253 papers. We focused on three years of ICSE because it is the flagship SE conference and not focused on a specific type of SE research, and because we wanted to understand the current state of ICSE rather than show trends over time. Additionally, it was pragmatically easier for us to contact the authors to participate in our survey by using more recent papers.

After collecting the papers, we developed rules to use for our categorization. We iteratively refined Runkel and McGrath's descriptions of data sources and research strategies as we applied them to the ICSE papers in our sample, producing the adapted model described in Section 3.3. We then classified the papers according to these completed descriptions and recorded the classification of each paper, along with the reasoning for the classification.

3.2 Survey of Authors

We follow the reporting guidelines described by Jedlitschka and Pfahl [6] and used by Siegmund *et al.* [23] to describe the design and dissemination of our survey to ICSE paper authors.

3.2.1 Objective. With our survey, we aimed to find answers to the research questions presented in Section 1. In order to accomplish this, we asked the authors questions about their ICSE papers as well as their careers as a whole. We phrased questions using the terminology from the research lens so that the findings would triangulate with our findings from the categorization study.

3.2.2 Participants. Our survey participants consisted of the first author of each ICSE paper from our categorization study. We focused on the first authors since they were likely heavily involved in the research according to common publication conventions. We also chose this approach to avoid sending multiple invitations to the same author (i.e., many researchers contribute to a high number of papers).

At the time they conducted the research reported in their ICSE papers, participants were split fairly evenly between being university faculty (46.7%) or students (43.3%), with some industry involvement and researchers who were affiliated with more than one entity when they conducted their research. They conducted their research in a variety of countries: 15 different countries were represented, with the most prevalent being the United States (31.7%). Participants indicated that they had a wide range of experience conducting SE research, with a minimum of 2 years experience, a maximum of 25 years experience, and a mean of 7.1 years of experience.

| Question | Response Options |
|---|--|
| <i>ICSE Paper Questions</i> | |
| SQ1) Your ICSE paper title is: | Open (Auto-filled for participants) |
| SQ2) When this research was conducted, you were affiliated with: | A university (as faculty), A university (as a student), An industrial research laboratory, Other |
| SQ3) Please select all the research strategies that describe the research conducted in your paper. | 8 Research Strategies, Other |
| SQ4) Please elaborate on why you used those strategies | Open |
| SQ5) Were there any other research methods used as part of the study but NOT reported in the paper? If yes, please elaborate which ones and WHY they were not described in the paper. | Open |
| SQ6) Which data source(s) were used in the research reported in your paper? | 7 Data Sources, Other |
| SQ7) Please elaborate on why you chose those data sources. | Open |
| SQ8) Please indicate where you feel the research reported in your paper fits within the following criteria: | Generalizability, Control, Realism: Very low, Low, Neutral, High, Very high |
| SQ9) Please explain your selections above. | Open |
| <i>General Research Questions</i> | |
| SQ10) Considering the last five years, please rate how often you have USED the following research strategies in your work. | 8 Research Strategies: Never, Rarely, Sometimes, Often, Most of the time |
| SQ11) How do you define triangulation, and how do you use triangulation in your work? | Open |
| SQ12) In your own work, do you PRIORITIZE any criteria (generalizability, control, or realism) over another? Why? | Open |
| SQ13) Do you PERCEIVE A BIAS in the software engineering research community regarding certain criteria (generalizability, control, or realism) when it comes to publishing work? | Open |

Fig. 1. Survey questions and possible responses, abbreviated for clarity.

3.2.3 Questionnaire and Conduct. The survey was designed in three parts. First, we asked the authors basic demographic questions. Second, we asked questions specifically about the ICSE papers they authored. To ensure that our participants understood the terminology in the survey (the research strategies, data sources, and the three desirable research criteria), we included a set of definitions with questions that involved these terms. We also provided examples next to each provided response option. We then asked authors to classify their papers and explain why they made those choices for their work. Finally, we asked participants about their research careers as a whole, and their perceptions and experiences. All sections contained both closed-ended and open-ended questions, and all questions were optional. A summary of the questions and possible responses is shown in Figure 1.

To contact participants, we used email addresses from ICSE papers and public researcher websites, contacting second or third authors if an email invitation failed to deliver. Participants were contacted in March 2018, with a reminder email two weeks before closing the survey to responses in April. In total, we sent 253 survey invitations successfully and received 60 responses, for a response rate of 23.7%.

3.2.4 Survey Analysis. We analyzed the survey data in three ways. First, closed answers were cleaned and visualized using R. Second, the author-generated classifications were compared to our findings from the categorization study. Where there were discrepancies, we investigated possible causes and card-sorted these instances to determine common causes for discrepancies. To validate our suggested causes of discrepancies, we conducted member checking with participants who indicated they were willing to answer further questions. For this step we did not use the terminology from the research lens, as we found miscommunication of the research lens to be the most common cause of discrepancies. After member checking, we applied minor corrections to the categorization study classifications where appropriate.

Finally, we followed an open coding approach to analyze the answers to open-ended questions. The first author of this paper coded the responses and organized the codes into overarching categories. Another author conducted an independent coding task on a subset of the data using the coding scheme developed by the first author. There were minimal differences between the two code sets, which were discussed and resolved when found. Primarily, these were errors of omission

as there were many codes in the set. The first author then synthesized the contents of the codes in these categories by iteratively describing and summarizing the data to produce a set of findings.

3.3 Our Research Lens for Interpreting Socio-Technical Research in Software Engineering

To classify the ICSE papers, we adapted Runkel and McGrath’s model of research strategies and data sources [13, 16]. This model highlights how the choice of research strategy and data source impacts desirable research criteria, generalizability, realism, and control. *Generalizability* refers to how generalizable the findings are to the population outside of the specific actors under study. *Realism* is how closely the context under which evidence is gathered matches real life. *Control* is defined as having control of the measurement of human behaviors under study, as well as any extraneous factors not under study.² Acknowledging that all method choices have inherent weaknesses, Runkel and McGrath emphasized the importance of triangulation across research strategies and data sources as a mitigation strategy. As the model was originally created for the traditional social sciences, the complexities of SE research introduced a number of fringe cases, signaling the need to adapt and extend the model for use in SE. We describe the adapted model below.

3.3.1 Research Strategies. Runkel and McGrath [13, 16] placed eight different research strategies on a “circumplex” diagram as segments of a circle; the circle is separated into four quadrants containing two research strategies each, as seen in Figure 2. Strategies are mapped on the circumplex according to the level of particularity/universality of the behavior systems under study and how obtrusive the researchers are into natural settings experienced by the human subjects. The circumplex also includes three dimensions representing desirable criteria—generalizability, realism, and control—and shows where each of the three exist at their highest potential for maximization.

Researchers must triangulate across the circumplex, using complementary strategies that make up for each other’s weaknesses, aiming to create a collective body of work that is high in generalizability, realism, and control. It is important to note that while a study follows a single research strategy, a research paper can contain multiple separate studies, and thus a paper can describe a number of research strategies.

- *Field strategies* in SE involve researchers entering the natural setting of studied participants to conduct their research. In **field studies**, the researcher does not manipulate the setting and instead conducts their research in the “natural” environment. For example, a researcher may observe how agile practices are used in a startup company. **Field experiments** differ by introducing a controlled condition into the situation under study to understand the effects it creates—compromising some *unobtrusiveness* for higher *control* in the resulting study. An example of a SE field experiment could be introducing a novel automatic testing tool in a company and observing its effects on code quality.
- *Experimental strategies* in SE involve testing hypotheses in highly controlled situations. These strategies yield *high control in the measurements* and *control over extraneous factors* but at the cost of *reduced realism of context* and *narrowed generalizability*. **Laboratory experiments** refer to situations created by the researchers where participants take part in an experiment. This strategy is used when researchers *focus on a certain behavior* and wish to measure it with *considerable control*. For example, a researcher investigating the effects of a new debugging tool on programming task efficiency may invite graduate students to a lab and ask them to accomplish a set of *predetermined* debugging tasks with and without the tool. **Experimental**

²In his paper, McGrath refers to the terms “precision” and “control” synonymously. For the purpose of clarity, we use the term “control”.

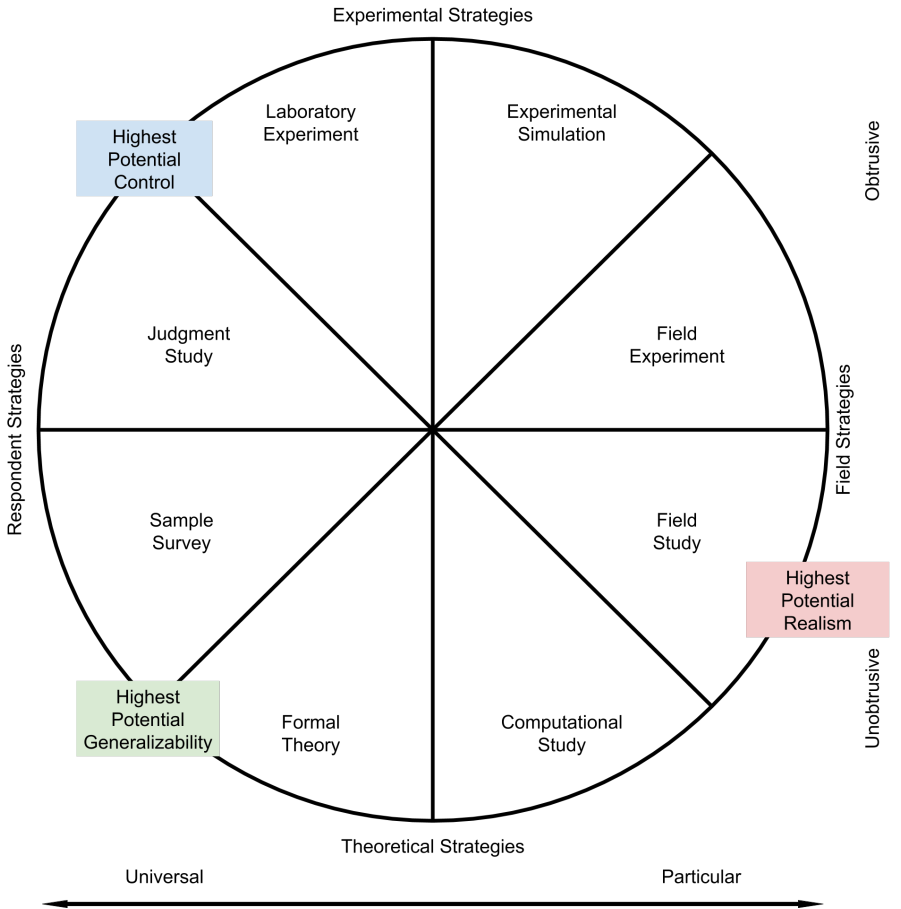


Fig. 2. McGrath’s Circumplex of Research Strategies

simulations in SE aim to *replicate some aspect of the participant’s natural environment* during a controlled experiment thus *gaining some realism*. For example, a researcher investigating project management meetings may conduct an experiment in a room with a similar setup to the one used at the company.

- *Respondent strategies* are used to systematically gather participant responses to questions posed by the researcher. **Sample surveys** aim to gather *information about the human behavior under a stimulus* while **judgment studies** aim to gather *information about the stimulus itself*. The main difference between **Sample Surveys** and **Judgment Studies** is whether the study aims to gather *information about the human behavior under a stimulus*, or *information about the stimulus itself*. These strategies make the participant’s physical setting and conditions irrelevant. Sample Surveys tend to use representative populations, making them *highly generalizable*, while Judgment Studies are typically done with “actors of convenience”, lowering the potential for *generalizability* but increasing *control*. **Sample Surveys** in SE are used to

investigate the effects that a phenomenon has on human behavior by surveying specific members of a chosen population, aiming at generalizing the findings to more of the population. For example, a researcher aiming to improve continuous integration tools may distribute an online survey, asking developers to describe how they use these tools and what challenges they face. Sample surveys are not limited to surveys in the traditional sense, and this method could also refer to interviews and focus groups. Sample surveys can be more convenient than field strategies because they often do not require physical access to an industrial environment and can be remotely conducted. **Judgment Studies** are commonly used in SE to evaluate the performance or utility of a new tool or technique. For example, in order to evaluate an API recommendation system, a researcher may invite developers to use the system and then survey them on the relevance and accuracy of the resulting recommended APIs. Judgment studies tend to be *high on control of measurement* of both the stimulus materials and the responses; however, they are often *low on generalizability* of population, as they are done with “actors of convenience” or relatively small population samples.

- *Theoretical strategies* differ from the previously described strategies as they are *the only methods not involving the inclusion of active human participation as part of the research* (but the studies may be based on past empirical data and studies). **Computational studies** refer to computer experiments using a *complete and closed system* to model operations without any human involvement or dynamic feedback from the outside world. The primary tool of the researcher is a computer. These studies are very common in SE and can be conducted using a wide variety of techniques, including experiments to evaluate software tools, data mining studies, computational analysis of big data, the creation and evaluation of prediction models, natural language processing techniques, and computer simulations. This strategy was originally named *Computer Simulation* by Runkel and McGrath, but we changed the name to Computational Study to reflect the varied nature of studies conducted using this strategy in SE. For example, a researcher aiming to evaluate a new bug detection technique may use version control history in an open-source project to see if their tool identified all the bugs that were fixed in subsequent versions of the project. Another example is running a series of experiments comparing the performance of various state-of-the-art static Android security analysis tools. Computational Studies may use methods for gathering and analyzing digitized data, which is common in data mining studies. **Formal theory** research focuses on the creation of models and theories based on previously gathered data or existing theories and models, instead of gathering new empirical data. In SE this includes qualitative synthesis studies, literature reviews, mathematical or logical research papers, etc. For example, by building on a previously formed model, a theory formulation study may aim to identify and describe underlying factors, which can explain why certain practices support alignment and coordination in software projects.

3.3.2 Adapting the Circumplex. As we conducted the categorization study, we realized that the circumflex model of Runkel and McGrath [16] had difficulty with completely characterizing the papers published at ICSE. Coming as it does from the social and behavioural sciences, the circumplex has no quadrant that maps directly to the many solution+evaluation papers we found. We therefore adapted the circumplex as shown in Fig. 3. Our high-level approach first separated empirical approaches from non-empirical approaches. For non-empirical papers, we created a category, Meta, for research that analyze the research papers themselves (such as a systematic literature review). We also moved Runkel and McGrath’s category Formal Theory for research strategies that use a mathematical epistemology.

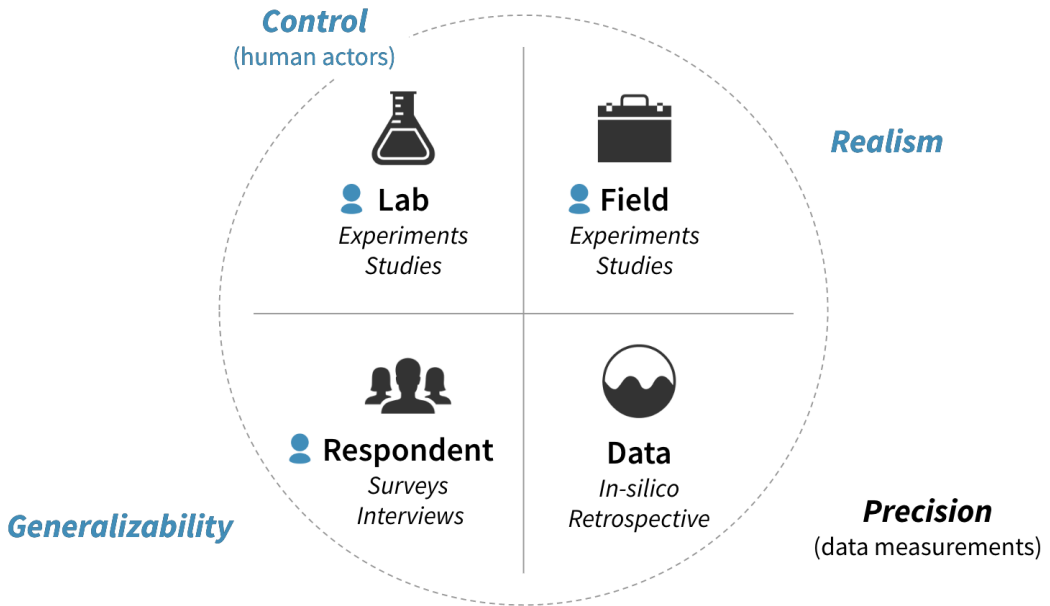


Fig. 3. Adapted circumplex for categorization study.

Simplifying the empirical categories of Runkel and McGrath, we created four quadrants. Three quadrants (all except the lower right) are primarily strategies that involve humans: Lab strategies, Field strategies, and Respondent strategies. These map directly to the categories Runkel and McGrath defined, and which we explained above.

Our new addition is to add a Data quadrant that reflects a logic of precision over data-based research strategies. As we will show, this quadrant captures the majority of papers published in our dataset. As the original circumplex nicely elaborates, one's choice of strategy maximizes the potential for either control, generalizability, realism, or precision, and reduces the potential for the other strategies. Thus, Data strategies maximize precision, while foregoing Control over human actors.

3.3.3 Data Sources. In addition to research strategies, McGrath [13] also describes a number of possible empirical data sources for behavioral research, and the benefits and drawbacks of each. These sources help us to determine the level of human involvement in the research. Self Reports and Observations are *active forms of human involvement*, where Archival Records and Trace Measures are *inactive forms of human involvement*. Additionally, research that uses logical constructs, mathematics, and proofs rather than empirical data has *no human involvement* outside of the researchers themselves.

Self Reports refer to instances where participants voluntarily report on their own behavior or perceptions for research purposes, usually responding to direct researcher questions through a questionnaire or an interview. They have the benefit of being able to determine a participant's perceptions about a topic from their own perspective, but they have the drawback of being at risk of bias from participants wanting to portray themselves positively, or tell the researcher what they want to hear.

Observations by a Visible Observer and **Observations by a Hidden Observer** are observations of human participants; either participants are aware they are being observed or measured

(Visible Observer) or not (Hidden Observer). Data gathered through these methods occurs in real time from a variety of techniques, including sensor data, video and audio recordings, and being physically present in the same room as a participant. Observational data has the potential to show how participants respond to different stimuli, which can be helpful for maximizing control. However, visible Observer data may be influenced by the participants *reacting* to the fact that they are being observed, potentially changing their behavior. Hidden Observer methods do not have this limitation, but instead present the ethical concern of researching someone without their consent. An example of Visible Observer data in SE are notes taken while observing a development meeting in a company, and one example of Hidden Observer data is entering a development team and observing their behavior for research purposes under the guise that you are a new team member. **Public Archival Records** and **Private Archival Records** are data about human behavior that is recorded by a third party for non-research purposes, but is used as the subject of research after the fact. The difference between them is that private records would be unlikely to become a matter of public record, like a diary entry. Both of these data sources are fairly uncommon in SE research; public records are often very easy to access and can be useful for showing trends over time, for example university graduation statistics could help to show trends in SE education. Private records, on the other hand, are often very difficult to access due to security and ethical concerns. One example of a private record in SE would be high-level production meeting minutes in industry.

Trace Measures are records indirectly created by humans as a result of their behavior. Humans create these measures on their own; they are not collected by a third party and they are not created for the purposes of research. Most software development artifacts fall into this category as they are *traces* created by developers as a result of software development behavior. For example, software is written by developers to fulfill some need, but later the source code (or its bugs, commits, or error logs) becomes a Trace Measure we can study in future research. Often, Trace Measures are publicly available and easily accessed, and are not influenced by the knowledge that the traces would be analyzed for research. However, there are drawbacks to using Trace Measures, particularly with the lack of control over measurement and lack of context available to explain such data.

McGrath discusses how data collection methods can be classified by type of human involvement.

Self reports refer to study instances where participants voluntarily report on their own behavior for research purposes. **Visible observer** and **hidden observer** data are observations of human participants; either they are aware they are being observed (visible) or not (hidden). **Public archival records** and **private archival records** are records of human behavior that are recorded by a third party for non-research purposes. The difference between them is that private records are unlikely to become a matter of public record. **Trace measures** are records indirectly created by humans as a result of their behavior. For example, the source code developers write to fulfill some need becomes a trace measure to be used in future research. Self reports and observations are considered *active forms of human participation*, while archival records and trace measures are *inactive forms of human participation*. **Formal/theoretical** is used to describe purely theoretical research that does not consider empirical data. It reflects the absence of human involvement beyond the researchers themselves.

4 FINDINGS

We structure this section around answering our research questions and include insights from both the categorization study and the survey. The categorization study data is presented after having made minor adaptations following our validation and member checking. To protect anonymity when presenting survey data, author (survey participant) quotes are identified using “Ax”, where the x corresponds to the order in which authors responded to the survey. We refer to survey questions

by the identifier “SQx” as seen in Figure 1. We refrain from presenting numerical data from the open-ended questions in our survey as quantifying this information may be misleading.

4.1 The Current State of Software Engineering Research from a Social Perspective

To understand how the SE research community addresses the social aspects of software development, we investigated the research strategies and data sources used by ICSE authors, as well as the corresponding balance of generalizability, realism, and control. We first present the research strategy and data source classifications from the categorization study. Numerical data here represents the number of papers that *include* a specific research strategy or data source.

RQ 1.1: What research strategies are described in research published at ICSE?

Among the 253 papers, we found a high use of data strategies (195/77.1%) compared to any of the other research strategies in the categorization study (shown in Figure 4). There were fewer instances of each of the other research strategies, with lab strategies being slightly less common than other strategies.

Data strategies were the most commonly used, but these data strategy papers reported on a variety of techniques. They include data mining studies, natural language processing experiments, computer simulations, computational experiments to evaluate tools and techniques, computational analysis of software artifacts, and computational prediction models.

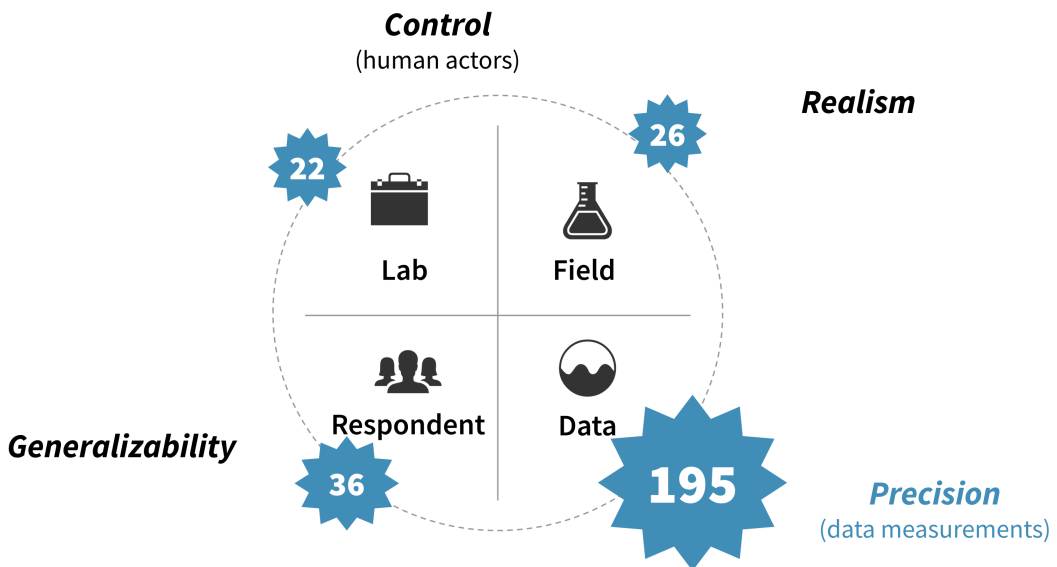


Fig. 4. Research strategies used in ICSE papers.

RQ1.2: What data sources are described in research published at ICSE?

We found a high use of trace measures as a data source (82.21%) in the categorization study (shown in Figure 5). About 20% of papers included self-reports or visible observer data, but other data sources were not featured prominently. Trace measures used in papers varied, with researchers reporting the use of log files, data from websites such as Stack Overflow, datasets of software bugs,

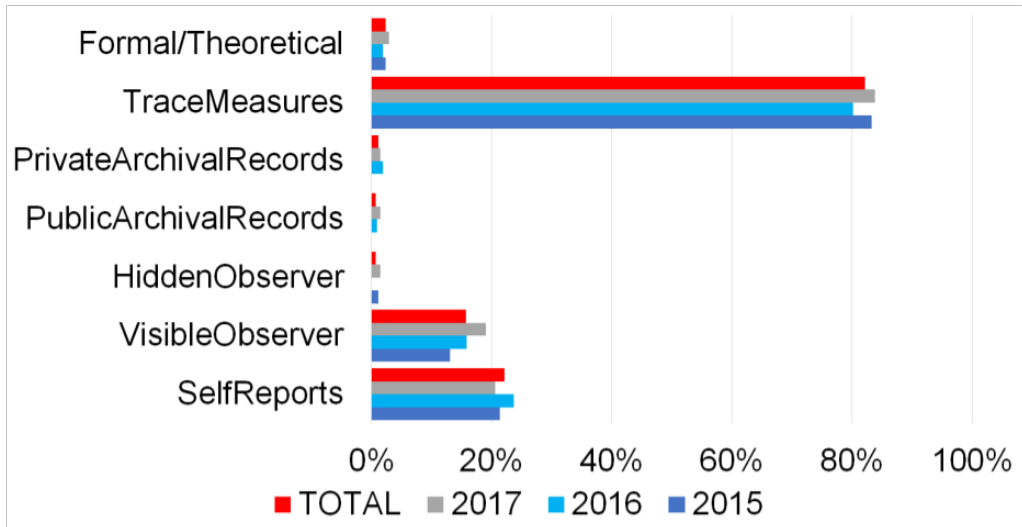


Fig. 5. Data sources included in ICSE papers.

open source repositories, code comments, research papers, and software programs such as websites and apps, among other sources.

RQ1.3: How does the software engineering research community prioritize generalizability, control, and realism?

To investigate the balance of generalizability, realism, and control in ICSE papers, Fig. 6 shows the absolute frequencies. Due to the high use of data strategies, our findings show a skew towards relatively high potential for precision over data, but low potential for control over human behaviour. Neither criteria's potential is fully maximized as it would be if we observed more field studies or sample surveys/formal theory.

In order to investigate the notion that levels of realism and generalizability may be higher than control in ICSE papers, we asked the authors to rate their papers according to these criteria (SQ8). The resulting distribution is shown in Figure 6. We see that authors rated their own papers more highly on realism and generalizability and lower on control. Overall, authors rated their papers highly on all three criteria; this is perhaps expected, as the papers were published in a top-tier conference and are likely to be of good quality, and authors are unlikely to respond in a way that reflects poorly upon themselves. Even so, it is the difference between each of the criteria that we would like to draw attention to, indicating that realism and generalizability may be more highly prioritized than control in ICSE papers.

We also asked the authors if they prioritize generalizability, realism, or control in their research in general (SQ12). While some authors indicated that they prioritized all criteria equally, others said that some approaches to research were better suited for prioritizing certain criteria over others. Further still, authors indicated some criteria were more important to them in their careers as a whole. Overall, realism was the highest priority for authors, followed by generalizability. A small minority of the authors indicated that they prioritized control.

When asked whether they perceived a bias in the community with regards to these criteria (SQ13), authors responded in a similar way; while some indicated they did not perceive a community

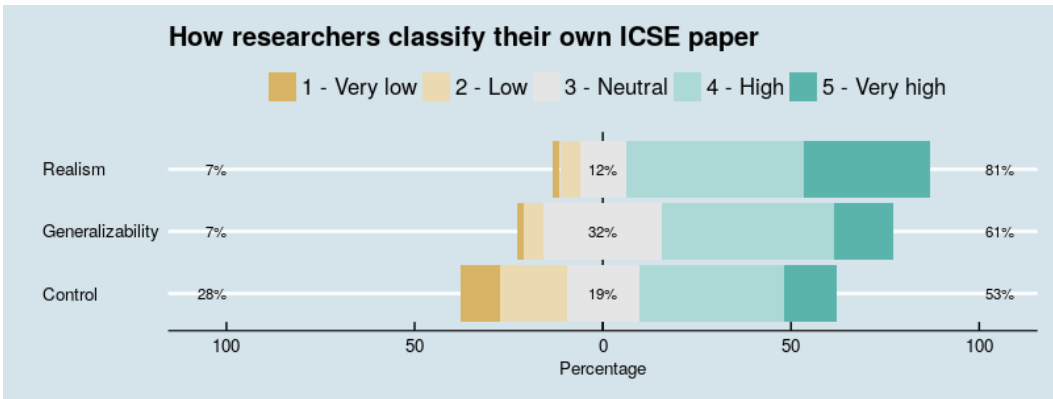


Fig. 6. Authors indicated that their papers are higher in realism, closely followed by generalizability, and lower in control.

bias towards any particular criteria, the majority indicated they believed the community was prioritizing some criteria over others. Authors shared their perception that the community was too focused on generalizability, followed by realism, and control was last. The responses to these two survey questions support the imbalance suggested by the categorization study data that realism and generalizability are prioritized over control in SE research from a social perspective.

RQ1.4: How is triangulation used in light of the prioritization of generalizability, control, and realism?

Given our findings of the high use of data strategy using trace measure data in ICSE papers, we chose to investigate triangulation. Computational studies can help researchers include large samples of varied data in their work to maximize generalizability. Trace measure data may often be readily available and is also not influenced by participant knowledge of research tasks, which can help researchers maximize realism in their studies. However, the combination of a data strategy and trace measure data includes a potential weakness: it does not allow for researchers to control for different confounding factors that influence developer behaviors as they cannot be well exposed with these methods. Triangulation across complementary research strategies and data sources is considered key for mitigating this weakness [16].

Papers in our categorization study reported up to three research strategies and up to four different data sources. We found that 48 papers (19%) reported more than one research strategy and 53 papers (21%) reported more than one data source. 71% of the papers published at ICSE in those three years do not report triangulation with different data source types or research strategies in a single paper. However, we recognize that these authors may have triangulated their research strategies and data sources and published this work in another venue, another year of ICSE, or not at all. We discuss author perceptions about triangulation in the following section.

4.2 Contributing Factors to the Current State of Software Engineering Research from a Social Perspective

Our findings from RQ1 helped us present an understanding of the current state of SE research from a social perspective. However, it is imperative that we also understand the factors that contribute to this state, which we addressed with RQ2. Such factors help us contextualize our understanding and illuminate potential issues for further discussion in Section 5. These factors were identified through

our qualitative analysis, primarily the author responses to SQ4-7 and SQ10-12. Each “box” below represents a unique factor contributing to the current state of SE research from a social perspective.

Authors focus on technical problems and choose methods that best fit these problems.

A common theme is that authors chose the research strategies and data sources that were the best fit for their research questions or topic area, and prioritized the criteria (generalizability, realism, or control) most relevant to their chosen approach. When authors communicated the focus of their research to us, unsurprisingly, the majority indicated that they addressed a technical problem with their work. While it is important to address technical problems in software development, we draw attention to the fact that there is *relatively less focus* on social problems in SE compared to technical issues.

Authors indicated they primarily used computational studies and formal theory in their work to study technical aspects of software development. Authors indicated that they used these strategies to evaluate an approach, tool, or algorithm that formed part of their research contribution. For example, authors said they conducted computational studies to “*prove [the] scalability of our approach*” (A10) and to “*[test] performance against previous benchmark suites*” (A45).

A minority of authors who were focused on technical research topics also indicated they thought that other strategies involving active human participation were beyond the scope of their work and they did not see them as an option. For example, an author commented that “*because we are working on algorithms [...] there is no need to conduct user study with human participants*” (A21) and “*user studies would have been beyond the scope of our work*” (A27).

A minority of authors indicated they used humans to evaluate their tools. For example, one author said they “*used a controlled environment for participants to use a tool we developed*” (A59), and another said they used “*self-reports to assess perceived usefulness*” (A18) of a tool.

Authors choose data sources opportunistically, by ease of use and access (or lack thereof) to certain data sources.

When analyzing why authors chose certain research strategies and data sources (SQ3-7), we found authors had practical concerns such as ease of use or availability. Authors said it was difficult to gain access to developers and software engineers to conduct studies involving them. They also highlighted that ethical concerns surface when accessing developers, making it difficult to use Hidden Observer data, for example. One author said, “*ideally, we also would have conducted a field experiment to answer questions regarding usability, but we didn’t have subjects readily available with the right training*” (A16). This suggests that lack of access to an appropriate population may be keeping authors from triangulating by using research strategies that involve developers.

At the same time, authors found it easier to gain access to trace measure data, saying it was publicly available (A34, A54, A17), easy to access (A53), or used by other researchers in related work (A34). One author indicated that Sample Surveys are an easy way to reach practitioners (A10), however, other authors explained they chose computational studies with trace measures because they were easier to conduct and more efficient at analyzing large sample sizes than other research strategies (A41, A57). Participants signaled that the use of readily available data may have hidden risks for the SE research community; as one author pointed out, “*just because data is convenient or available does not mean it reveals what we are looking for.*” (A4)

There is a potential lack of knowledge about triangulation in the software engineering research community.

After finding very little triangulation across research strategies and data sources in ICSE papers during our investigation into RQ1.4, we asked authors to define triangulation and describe how they use it in their work (SQ11). This was to gauge how authors understood triangulation (if at all), and whether or not it was a technique they applied to their work. For context, we understand triangulation as “*taking different angles towards the studied object and thus providing a broader picture*” of the phenomenon under study, as described by Runeson and Höst [15].

Some authors provided us with rich definitions of triangulation and indicated they utilized a number of different techniques in their work. For example, one author said, “*triangulation means integrating different sources of evidence. You can triangulate across data collected from different sites, using different methods, or analyzed in different ways. I do all of the above, depending on the study.*” (A12) Other authors did not provide a definition, but described their knowledge and practice of triangulation in the context of their own specific research domain, like experimentation: “*I define triangulation as reaching conclusions based on multiple data sources and/or multiple experiments that investigate some phenomenon using different techniques.*” (A60)

A surprising number of authors (17 out of 60) indicated they did not understand triangulation in the context of research. Some of the responses given were “*I’ve never used this word; I’m unsure what is being asked exactly (perhaps that answers something?)*” (A16) and “*I did not know about triangulation prior to this survey*” (A56). This finding may have been influenced by the presence of students in our sample; 11 out of 17 authors indicated they were students at the time they conducted the research published the paper. Still, these students may have joined the research community in a different role after the publication of their paper.

Authors believe that reviewers are too focused on generalizability, so they focus on having large sample sizes to be accepted for publication.

Authors indicated a perceived bias towards generalizability in the SE community, particularly with regards to the reviewing process. Authors reported that, in their view, generalizability is too heavily emphasized by reviewers because it is easy to criticize in papers, even if work is highly realistic or controlled (A19, A28). Authors thought this was leading reviewers to have “*totally unrealistic expectations regarding generalizability*” (A26) and sample size, and one author said, “*generalizability is highly demanded by reviewers. This is why there is an increasing number of subjects (software systems and developers) in studies over the last decade.*” (A34)

Authors noted this influenced their priorities, saying that they tried to maximize generalizability in their work because “*that’s what reviewers easily criticize*” (A19). The perception that large sample sizes are needed in a study for a paper to be accepted for publication may be keeping authors from conducting research that is highly naturalistic or controlled, such as field studies or laboratory experiments, where having large sample sizes is prohibitively difficult.

Authors prioritize having high realism and generalizability in order to be relevant to practitioners.

When asking authors about their priorities and perceived bias toward particular criteria (SQ12, 13), we found they emphasized a need for practical relevance in SE research. Those who prioritized realism in their work tended to have relationships with industry and were concerned with creating solutions to real-world problems that could be adopted by their industrial collaborators, as well as other developers with similar issues. For example, one author said, “*realism comes first. Given that I work [in] an industrial research lab and all my research needs to help practitioners in the real-world daily work.*” (A10) Another explained, “*I need to have impact on the process and save money for the company. Not being realistic means I will have zero impact.*” (A31)

Other authors believed that realism and generalizability were dependent on each other. For example, one author felt that it is *“important to make sure that the approach could generalize for subjects beyond the current study to ensure the applicability of the approach”* (A22). Another author commented that *“realism is an important contributor to generalizability; the two are not orthogonal”* (A9).

It was noted by some authors that this perceived need for practical relevance in SE research potentially influences research direction. One author noted that *“software engineering researchers are not interested in “boring” research; they are very solution-focused. As a result, fundamental research does not get done in my opinion, because it requires giving up realism for control.”* (A3) Another responded, *“I think in the last decade there is a growing and strong bias towards realism. It is a strength but also a weakness: bias towards short term impact, bias towards solutions that work although nobody cares why, results over-fitting available data.”* (A42) This view could explain why fewer studies are using methods that investigate causal factors of software development behavior.

5 DISCUSSION

Our investigation of ICSE from a social perspective leads to a number of valuable insights, each of which has several implications for the SE research community. We aim to generate reflection and discussion in the community by outlining the implications of our findings and raising questions in light of our collected evidence.

5.1 Replication in Software Engineering

The snapshot we produced through our study shows a preference towards computational studies and trace measures in current ICSE research. While there is merit in using these methodological tools in particular for studying technical research questions, as a community we should be aware of the trade-offs. Many papers that focus on technical solutions also may have to take into consideration social and human factors, but it is difficult to understand the social factors that cause the behaviors we measure through a computational approach alone. Similarly, trace measures—while often easy to obtain—may be missing social and contextual factors that led to the data’s creation [1]. Without understanding the social context surrounding the creation of software artifacts that we use in our research, we have no way of knowing how the rapid advancement of software development technologies and practices will affect our work.

Computational studies are typically considered highly replicable because they are often accompanied by packages including the software artifacts, algorithms, and tools used in the studies. However, without understanding *what about the social context of development made the artifacts the way they are*, it may be difficult to truly replicate the study on software that was created in a similar way. After all, software is continuously evolving and so diverse that the context surrounding software is more important than ever. One author commented to us that *“given the wide variety of programming languages, project processes and communities, open source vs proprietary projects, etc. I think true generalizability is often difficult to achieve.”* (A48) Thus, we call on the community to reflect on the *true* replicability of this type of work in SE.

5.2 The Importance of Understanding Causal Factors of Software Developer Behaviors

When mapped on Runkel and McGrath’s circumplex, we see that the community’s most popular choice for research strategy supports the criteria of generalizability and realism while neglecting control. The authors who responded to our survey indicated that this imbalance reflected their priorities, where realism and generalizability were seen as necessary to have practical relevance and be accepted for publication. However, highly controlled or naturalistic research in SE helps us to understand the specific factors that cause certain software development behaviors, which is

instrumental for developing foundational knowledge on which to base our future research. Having this type of knowledge helps us to make informed decisions and base our tools and technologies on a deeper understanding of the needs of developers that will use them.

5.3 Triangulation in Software Engineering Research

Triangulation *could* serve as a mitigation strategy to the high use of computational studies and trace measure data in SE. Runkel and McGrath recommend the use of various research strategies and data sources to address imbalance in prioritizing generalizability, realism, and control. While we could not determine whether authors triangulated their work outside of their paper, our findings show that the vast majority of the papers in our sample used a single research strategy and data source type. In addition, we found an alarming number of participants (17 out of 60) that indicated they did not understand the concept of triangulation in the research context. This suggests potential knowledge gaps in the SE research community about triangulation. Triangulation is a critical concept for empirical research, so we suggest that this issue requires further investigation.

However, the responsibility for triangulation does not need to be on the level of individual studies; while it is valuable to triangulate findings with multiple research strategies and data sources within a single paper, it may be impractical, and many studies have valuable insights that warrant an entire paper. Instead, we suggest that the community as a whole should be responsible for triangulation; it is unreasonable to expect each researcher to have the knowledge *and* the means to conduct research using all of the methods available. Thus, we call on the community to reflect on the value of studies that triangulate work conducted by other researchers. Harnessing our strengths and skills in particular methods as individual researchers to triangulate the work of others will help us to develop more impactful findings for the community as a whole[11], and help us to better understand the social, as well as technical, aspects of software development.

5.4 Reflecting on the Contributing Factors to the Current State of ICSE

Investigating the factors that are contributing to the current state of SE illuminated some potential issues, and we pose open questions to the research community based on our evidence.

The research community's overwhelming focus on technical research was a major factor for the high use of computational studies and trace measure data, as it is often the combination that is the best fit to address highly technical research questions. However, we argue that social factors are equally as important as technical concerns in SE, and require attention. We wonder if it is not possible to reframe some of the research questions that guide us in studying technical phenomena to instead consider them from a *socio-technical* perspective, incorporating both social and technical concerns. As a community, are the research questions we choose to investigate adequately addressing the socio-technical nature of software development?

Our analysis revealed that there is a dissonance between author priorities and what they perceive as the priorities of reviewers. Paper reviewers are authors themselves, so we suggest looking into other explanations of the current state. One author already suggested that the possible cause could be that reviewers do not have adequate time to properly review papers (A25), and that generalizability is easier to criticize in a paper than realism or control. If authors perceive that their paper will not be accepted if they cannot produce a large sample size they may refrain from conducting highly controlled or naturalistic research, where having large sample sizes is prohibitively difficult. This type of research is key to understanding the complexities of human behavior in software development, so we suggest that the issue of reviewer prioritization needs further study and reflection from the community. Do our peer review and publication processes bias the community towards certain types of research?

6 LIMITATIONS

In this section, we identify the limitations associated with this research and the measures we took to mitigate these issues through the research design.

6.1 Construct Validity

The model we adapted for this research was originally developed in the 1970s for traditional behavioral science domains such as psychology and sociology, not socio-technical domains like software engineering. One threat may be that the original model may not adequately capture the true potential for generalizability, realism, or control in software engineering studies. To mitigate this issue, we iteratively adapted and extended the original model to fit the current realities of software engineering research so that they accounted for the technical aspects of software development.

Also, although Runkel and McGrath's model is from the 1970s, it is generally considered as still relevant today in the fields of HCI (Human Computer Interaction) and CSCW (Computer Supported Cooperative Work) and applied to socio-technical research settings. We introduced this model to software engineering researchers back in 2008 [5] as a way to guide suitable research methods for research questions, and this paper is still being used today. Most notably, Stol et al., adapted it for their "ABC of software engineering research" paper [28]. However, they changed the interpretation of the obtrusiveness dimension to mean obtrusiveness in terms of how obtrusive a researcher may be on a research setting – a researcher that manipulates the data only would with their interpretation be obtrusive. We remain consistent with Runkel and McGrath and consider obtrusiveness in terms of obtrusiveness on the human subjects being studied. Thus, for Stol and Fitzgerald's interpretation, a lab experiment could be an experiment with no human subjects (and rely on data only that is perturbed by the researcher), whereas, we consider an experiment that uses and manipulates data only as a "computational study" rather than a lab experiment.

Indeed, this distinction was perhaps confusing to some of our survey participants as some authors confused "computational study" with "experimental simulation". An experimental simulation does involve human observation and participation with some aspects of their environment being simulated, whereas a computational study is conducted without any direct human subject interactions. To mitigate this issue, we triangulated the findings from the survey with the categorization study we did. We also conducted member checking with authors without relying on the use of the terminology from the research lens to determine a common ground between the authors and the researchers.

6.2 Internal Validity

The majority of the research tasks reported in this paper were conducted by the first author, which introduces a threat to internal validity. Because many of the analysis tasks rely on human judgment, such as open coding and classification of research papers, heavily relying on a single researcher introduces the potential for researcher bias. To mitigate this issue, we validated the classifications from the categorization study by having the authors of papers classify their own work as part of the survey. We also validated the open coding of long-answer questions with a second independent coder, and provided a number of analysis documents on our website to make the analysis process as transparent and traceable as possible.

6.3 External Validity

For our investigations, we chose to focus on three years of ICSE proceedings and the authors of those papers. This introduces a limitation in terms of external validity as different years, different

venues, and different tracks may have produced a different distribution of research strategy and data source use, and may have attracted researchers with different experiences and opinions to participate in our survey. As mentioned earlier, there are other venues that clearly focus on human aspects, including the CHASE workshop, co-located with ICSE since 2011, as well as other venues such as VL/HCC³ and CSCW⁴. Thus we recognize our findings are particular to ICSE, but we feel important to share as the ICSE publishing venue is recognized as being inclusive in terms of topics and methods, but it is also seen by many as the premier publishing venue in software engineering.

Researchers may wish to apply our approach to other venues (e.g., journals or ESEC/FSE conference) thus we provide a number of documents on our supplementary website designed to help other researchers follow our methodology, making it highly replicable. We encourage members of the community to conduct replication studies on additional years of ICSE and other venues to explore the differences that may exist between venues and time periods in SE.

6.4 Novelty

We know of only one other work that has applied the circumplex model to learn about the SE research community. Stol and Fitzgerald [27] described the benefits of using the model in SE, but as we mentioned above they rely on a very different interpretation of the model. In their interpretation an “actor” when studying the SE process could be either a human participant or a *software system*, deviating from the original model where humans and social constructs were the only potential “actors” for a study. We have developed an interpretation of the model from a *socio-technical* perspective. We believe that maintaining that actors must be human allows us to understand how SE research addresses social factors of SE.

7 CONCLUDING REMARKS

Through our categorization study and survey of ICSE authors, we identify a number of potential issues within the community for further study and consideration. With all of the implications of our findings in mind, we call the SE research community to action: software is designed, developed, and maintained by people. If our goal is to improve software development processes and tools, we must, as a community, adequately study humans to understand the social factors that influence software development. Understanding the complexities of human behavior requires the use of a diverse set of research methods and both active and inactive forms of human participation to produce a collective body of work.

Our findings present a platform for the SE research community to have an informed discussion around a number of issues, including the role of triangulation in SE, the effects of our current reviewing process, our choice of methods and topics, and our prioritization of generalizability, realism, and control with respect to the study of human behavior. We do not suggest that any of these issues should be acted on immediately; rather, we call the community to further investigate, reflect upon, and discuss these issues. The work that we produce as a community should be a reflection of our collective values and goals for the future of SE, and so it is time for the SE research community to engage in a discussion around our priorities for the future of the discipline, carefully considering the benefits and drawbacks associated with the current state of our research output. If we see our current state as being in conflict with our vision for the future of our research field, then we must then also discuss the changes we need to make moving forward to better consider the social aspects of the socio-technical system that is software development. We hope that this

³<http://conferences.computer.org/VLHCC/>

⁴ACM Conference on Computer Supported Cooperative Work, <https://cscw.acm.org>

paper sparks a change within the research community and that we begin to diversify our research choices to include more active human involvement in our work.

As this research investigated a number of issues in the SE research community, there are a number of possible areas for future work and follow-up studies. Our findings suggest a lack of knowledge about triangulation within the SE research community, which we believe should be studied further to determine the severity of this issue, as well as potential causes and solutions. We also believe that the disconnect between author and reviewer priorities calls for potential action to align our reviewing process with one that reflects our collective priorities.

REFERENCES

- [1] Jorge Aranda and Gina Venolia. 2009. The Secret Life of Bugs: Going Past the Errors and Omissions in Software Repositories. In *Proceedings of the 31st International Conference on Software Engineering (ICSE '09)*. IEEE Computer Society, Washington, DC, USA, 298–308. <https://doi.org/10.1109/ICSE.2009.5070530>
- [2] Frederick P. Brooks, Jr. 1995. *The Mythical Man-month (Anniversary Ed.)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.
- [3] Reidar Conradi and Alf Inge Wang. 2003. *Empirical Methods and Studies in Software Engineering: Experiences from Esernet*. Springer-Verlag, Berlin, Heidelberg.
- [4] Tom DeMarco and Timothy Lister. 1987. *Peopleware: Productive Projects and Teams*. Dorset House Publishing Co., Inc., New York, NY, USA.
- [5] Steve Easterbrook, Janice Singer, Margaret-Anne Storey, and Daniela Damian. 2008. *Selecting Empirical Methods for Software Engineering Research*. Springer London, London, 285–311. https://doi.org/10.1007/978-1-84800-044-5_11
- [6] A. Jedlitschka and D. Pfahl. 2005. Reporting guidelines for controlled experiments in software engineering. In *2005 International Symposium on Empirical Software Engineering, 2005*. 10 pp.–. <https://doi.org/10.1109/ISESE.2005.1541818>
- [7] Barbara Kitchenham. 2007. *Empirical Paradigm – The Role of Experiments*. Springer Berlin Heidelberg, Berlin, Heidelberg, 25–32. https://doi.org/10.1007/978-3-540-71301-2_9
- [8] Barbara A. Kitchenham and Shari L. Pfleeger. 2008. *Personal Opinion Surveys*. Springer London, London, 63–92. https://doi.org/10.1007/978-1-84800-044-5_3
- [9] Andrew Jensen Ko, Thomas D. LaToza, and Margaret M. Burnett. 2015. A practical guide to controlled experiments of software engineering tools with human participants. *Empirical Software Engineering* 20, 1 (2015), 110–141. <https://doi.org/10.1007/s10664-013-9279-3>
- [10] Jyrki Kontio, Johanna Bragge, and Laura Lehtola. 2008. *The Focus Group Method as an Empirical Tool in Software Engineering*. Springer London, London, 93–116. https://doi.org/10.1007/978-1-84800-044-5_4
- [11] Wendy E. Mackay and Anne-Laure Fayard. 1997. HCI, Natural Science and Design: A Framework for Triangulation Across Disciplines. In *Proceedings of the 2Nd Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques (DIS '97)*. ACM, New York, NY, USA, 223–234. <https://doi.org/10.1145/263552.263612>
- [12] John A McDermid and Keith H Bennett. 1999. Software engineering research: a critical appraisal. *IEE Proceedings-Software* 146, 4 (1999), 179–186.
- [13] Joseph E. McGrath. 1995. Human-computer Interaction. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, Chapter Methodology Matters: Doing Research in the Behavioral and Social Sciences, 152–169.
- [14] Joseph E McGrath. 1995. Methodology matters: Doing research in the behavioral and social sciences. In *Readings in Human-Computer Interaction*. Elsevier, 152–169.
- [15] Per "Runeson and Martin" Höst. 2008. Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering* 14, 2 (19 Dec 2008), 131. <https://doi.org/10.1007/s10664-008-9102-8>
- [16] Philip J. Runkel and Joseph E. McGrath. 1972. *Research on Human Behavior*. Holt, Rinehart, and Winston, Inc.
- [17] Carolyn B. Seaman. 2008. *Qualitative Methods*. Springer London, London, 35–62. https://doi.org/10.1007/978-1-84800-044-5_2
- [18] H. Sharp, Y. Dittrich, and C. R. B. de Souza. 2016. The Role of Ethnographic Studies in Empirical Software Engineering. *IEEE Transactions on Software Engineering* 42, 8 (Aug 2016), 786–804. <https://doi.org/10.1109/TSE.2016.2519887>
- [19] Helen Sharp and Hugh Robinson. 2005. Some social factors of software engineering: the maverick, community and technical practices. In *ACM SIGSOFT Software Engineering Notes*, Vol. 30. ACM, 1–6.
- [20] Mary Shaw. 2003. Writing Good Software Engineering Research Papers: Minitutorial. In *Proceedings of the 25th International Conference on Software Engineering (ICSE '03)*. IEEE Computer Society, Washington, DC, USA, 726–736.
- [21] Ben Shneiderman. 1980. *Software Psychology: Human Factors in Computer and Information Systems (Winthrop Computer Systems Series)*. Winthrop Publishers.

- [22] Forrest Shull, Janice Singer, and Dag I.K. Sjøberg. 2008. *Guide to Advanced Empirical Software Engineering*. Springer-Verlag, Berlin, Heidelberg.
- [23] Janet Siegmund, Norbert Siegmund, and Sven Apel. 2015. Views on internal and external validity in empirical software engineering. In *Software Engineering (ICSE), 2015 IEEE/ACM 37th IEEE International Conference on*, Vol. 1. IEEE, 9–19.
- [24] Janice Singer, Susan E. Sim, and Timothy C. Lethbridge. 2008. *Software Engineering Data Collection for Field Studies*. Springer London, London, 9–34. https://doi.org/10.1007/978-1-84800-044-5_1
- [25] Dag I. K. "Sjøberg, Tore Dybå, Bente C. D. Anda, and Jo E." Hannay. 2008. *Building Theories in Software Engineering*. Springer London, London, 312–336. https://doi.org/10.1007/978-1-84800-044-5_12
- [26] D. I. K. Sjøberg, T. Dyba, and M. Jorgensen. 2007. The Future of Empirical Methods in Software Engineering Research. In *Future of Software Engineering, 2007. FOSE '07*. 358–378. <https://doi.org/10.1109/FOSE.2007.30>
- [27] Klaas-Jan Stol and Brian Fitzgerald. 2015. A Holistic Overview of Software Engineering Research Strategies. In *Proceedings of the Third International Workshop on Conducting Empirical Studies in Industry (CESI '15)*. IEEE Press, Piscataway, NJ, USA, 47–54.
- [28] Klaas-Jan Stol and Brian Fitzgerald. 2018. The ABC of Software Engineering Research. *ACM Trans. Softw. Eng. Methodol.* 27, 3, Article 11 (Sept. 2018), 51 pages. <https://doi.org/10.1145/3241743>
- [29] Klaas-Jan Stol, Paul Ralph, and Brian Fitzgerald. 2016. Grounded Theory in Software Engineering Research: A Critical Review and Guidelines. In *Proceedings of the 38th International Conference on Software Engineering (ICSE '16)*. ACM, New York, NY, USA, 120–131. <https://doi.org/10.1145/2884781.2884833>
- [30] C. Theisen, M. Dunaiski, L. Williams, and W. Visser. 2017. Writing Good Software Engineering Research Papers: Revisited. In *2017 IEEE/ACM 39th International Conference on Software Engineering Companion (ICSE-C)*. 402–402. <https://doi.org/10.1109/ICSE-C.2017.51>
- [31] Norman G. Vinson and Janice Singer. 2008. *A Practical Guide to Ethical Research Involving Humans*. Springer London, London, 229–256. https://doi.org/10.1007/978-1-84800-044-5_9
- [32] Gerald M. Weinberg. 1985. *The Psychology of Computer Programming*. John Wiley & Sons, Inc., New York, NY, USA.
- [33] Brian Whitworth. 2011. Virtual communities: concepts, methodologies, tools and applications. , 1461-1481 pages.
- [34] Claes "Wohlin, Martin Höst, and Kennet" Henningsson. 2003. *Empirical Research Methods in Software Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg, 7–23. https://doi.org/10.1007/978-3-540-45143-3_2
- [35] Claes Wohlin, Per Runeson, Martin Host, Magnus C. Ohlsson, Björn Regnell, and Anders Wesslin. 2012. *Experimentation in Software Engineering*. Springer Publishing Company, Incorporated.
- [36] Marvin V. Zelkowitz. 2007. *Techniques for Empirical Validation*. Springer Berlin Heidelberg, Berlin, Heidelberg, 4–9. https://doi.org/10.1007/978-3-540-71301-2_2