Action Escrows for Social Change in Online Communities

Pranav Khadpe

CMU-HCI-
August

Human-Computer Interaction Institute
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213

Thesis Committee:
Geoff Kaufman, Co-Chair, CMU HCII
Chinmay Kulkarni, Co-Chair, Microsoft AI
Jason Hong, CMU HCII
Karrie Karahalios, UIUC

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy.

Keywords: escrows mechanisms, online communities, norm misperceptions, critical mass

Abstract

Many long-standing problems in the online communities literature, such as groupthink, silent majorities, bystander effects, and collective action failures, can all be traced back to first-mover disadvantages: it is risky to be the first to speak up. Isolated individuals deciding whether or not to raise a diverging perspective or to initiate a collective action effort often cannot be sure whether others would welcome it, and back them up. As a result, people may fail to surface an opinion even though it is privately held by many, refrain from publicly speaking up against misbehavior even though many privately think it is unacceptable, and fail to act in response to concerns that are not voiced publicly but are widespread. First mover disadvantages don't just impede one-time prosocial actions; they also act as a brake on positive norm change.

This dissertation proposes that designers of online communities can lower these first-mover disadvantages through a design pattern that I call an *action escrow*—a mechanism where people deposit a socially risky action with an intermediary system that only executes the action if a prespecified trigger criterion is met. For example, an action escrow for encouraging authentic opinions might allow a user to place a comment into escrow with the instruction that it be posted publicly only if the escrow system receives similar comments from two other users. Although action escrows are not new—they feature in some existing systems and are inspired by traditional escrows in legal and economic scholarship—this dissertation gives name to this loosely applied pattern and formalizes its scope, and utility for addressing persistent challenges in online communities.

To show that action escrows are effective, I introduce two systems, Nooks and Empathosphere, each of which advances action escrows further as an approach to addressing first-mover disadvantages. Nooks shows how action escrows can lower first-mover disadvantages in bringing up new topics in a community. Empathosphere shows how action escrows can be applied to give communities translucence into privately held opinions. Through a field deployment of Nooks and an experimental evaluation of Empathosphere, I show how action escrows do indeed address first-mover disadvantages and by doing so, encourage prosocial action, reveal suppressed perspectives, and improve inclusion.

To inform future implementations, I describe a broader design space of action escrows, outline the limitations and risks of introducing action escrows into online communities, identify how these risks can be mitigated, and synthesize broader opportunities for escrow mechanisms to address long-standing challenges in HCI.

As activities in online communities increasingly flow beyond digital boundaries to shape political movements, social institutions, and civic discourse, I envision a future where thoughtfully designed mechanisms can re-engage dormant voices, counter false polarization, and foster inclusion, in digital spaces and ultimately in broader society.

Acknowledgments

It is hard work writing a dissertation. Many people have helped me knowingly or unknowingly, along the way and it is nice to be able to thank them publicly here. But I will inevitably forget some. If I do forget you, I apologize. I still appreciate you; don't be sad.

I want first of all to thank my advisors Chinmay Kulkarni and Geoff Kaufman. It is hard to enumerate all the ways in which Chinmay has contributed to this dissertation, to my growth, and to how I think about the world. To name just a few: he has refreshingly brilliant ideas, is fearless at executing, learns dangerously quickly, and is a fierce champion for the people he works with. I want to be like him when I grow up, so I'm glad I get to continue turning to him for advice on research and all my other endeavors. Working with Geoff has been extremely fun because he takes play seriously, in every way: he has shown how you can design games that can increase voter turnout, and he has beaten me every time we've played pickleball. Beyond his social psychology expertise, Geoff has brought mischief, play, and subversion to our work—things I hope to never outgrow. I hope there is a sense of optimism about people and about technology that runs through this dissertation, and if there is, it comes from Chinmay and Geoff.

Thank you also to my committee members, Karrie Karahalios and Jason Hong. A big part of working on completing a PhD and launching yourself into a field is the act of making yourself legible, to yourself and to your community. Karrie was ahead of me in understanding who I was as a researcher. She was always able to sense what really excited me, who I'd probably read, and what would be intellectually exciting for me to read and think about. Jason was always able to forecast how my work, at least as written, would be received by the community. His feedback is predictive, honest, and challenging—all reasons why I am grateful to have him on my committee. He was also always the first person on my committee to get back to me with feedback on anything I wrote.

Much of the writing in this dissertation reflects my own intellectual work, but research, at least the kind I like, is a collaborative effort, and the papers I've used as starting points for some of the chapters have invariably been shaped by my collaborators. Chinmay and Geoff, of course, have provided help with framing and naming things in the papers that became this dissertation. Thanks to Lindsay Popowski, Kyzyl Monteiro, and Lindy Le for their help with the action escrows paper that would become the spine of this dissertation. Thanks to Shreya Bali for her work on Nooks (Chapter 4), Olivia Xu and Anna Gu for their work on Hug Reports (mentioned in Chapter 3), and Kimi Wenzel and George Loewenstein for shaping some ideas that appear in Chapter 7.

On my way to and through grad school, I've been fortunate to have the mentorship of many inspiring researchers. I'm forever grateful to Michael Bernstein and Ranjay Krishna for responding to a cold email from an undergraduate halfway around the world and giving me that formative summer at Stanford HCI, and for

teaching me how to do research, present, write, and strive for excellence. None of this would have been possible without Monojit Choudhury, who gave me the golden ticket opportunity to do an undergraduate internship at Microsoft Research India, introducing me to HCI research and putting me on the track towards a career in research. I later had the opportunity to return to Microsoft to work with Jina Suh and Shamsi Iqbal, both of whom have been incredibly generous—with time, ideas, encouragement, and support—during and beyond my internship.

I'm also indebted to Sauvik Das, Ken Holstein, Danaé Metaxa, Vineet Pandey, and Koustuv Saha for their guidance at different stages of my academic journey, from applying to grad school to the job search at the end. While I hope to pay back your generosity, I will most certainly pay it forward!

Thank you also to my collaborators on other projects I got to work on during grad school: Yasmine Kotturi, for showing me that sometimes the most effective way to design communication platforms that support safe, inclusive participation is by adopting design processes that mirror those values; and Alicia DeVrio and Myra Cheng, for including me in your adventures and making this final stretch so much more fun!

Thank you to all of my other peers in the field who have been sources of inspiration, pillars of support, and partners in venting: Adinawa Adjagbodjou, Lea Albaugh, Karan Ahuja, Alex Cabrera, Julia Cambre, Hao-Fei Cheng, Yi-Fei Cheng, Hyunsung Cho, Erica Principe Cruz, Wesley Deng, Nathan DeVrio, Lisa Egede, Will Epperson, Morgan Evans, Pedro Ferreira, Kapil Garg, Luke Guerdan, Amber Horvath, Jane Hsieh, Anna Kawakami, Seyun Kim, Isadora Krsek, Tzu-Sheng Kuo, Andrew Kuznetsov, Hank Lee, Vimal Mollyn, Kyzyl Monteiro, Katelyn Morrison, Tonya Nguyen, Soya Park, Napol Rachatasumrit, Vivian Shen, Venkat Sivaraman, Cella Sum, Vedant Swain, Jordan Taylor, and Nur Yildirim.

Many thanks to the following friends who did all sorts of things to help me along the way. Bart Duisterhof, Sam Speer, Sarah Costrell, Rayna Hata, Tejus Gupta, and Rishi Veerapaneni for making sure grad school had playtime. Alicia DeVrio for her sharp observations about the everyday, that make me jealous I didn't think of them myself. Nathan DeVrio for teaching board games with the kind of patience and calm that has become my template for the kind of teacher I want to be. Anna Kawakami for writing advice like "all the sections need to talk to each other". Luke Guerdan for reminding me repeatedly and by example that a peaceful mind is the best kind to aspire to have. Vivian Shen for showing me that you can win at anything if you really care, even raffles. Jordan Taylor for our Sunday long runs during which he guided me through the landscape of critical scholarship. Wesley Deng for faithfully attending my social computing reading group, often as the sole participant beyond the organizers. Friends from MSR India across the US—Adithya Pratapa, Simran Khanuja, Pratik Joshi, Vidhi Jain, and Prakhar Agarwal—for making navigating grad school, and a new country a collective journey. Aamnah Khalid for being the most charismatic person, and for making everything feel like an adventure. Nithin Kannan for saying wise things like "efficient markets make people sad because no one gets a good deal." Prathamesh Dandekar for showing me that no project is too crazy. So-

ham Kamat for getting it. Marc Mascarenhas for always planning the next reunion. Kimi Wenzel, Conor Igoe, Elissa Wu, Mrinal Verghese, Ananya Rao, all the associates of the Darlington house, and my friends at Pigeon, Manor, Commonplace, and Allegro for making Squirrel Hill feel like home. Gokul Swamy, Pratiksha Thaker, Ananya Joshi, Sayan Chaudhry, and Arka Chaudhury for enabling my tomfoolery. Sanika Moharana was no help at all.

I would like to thank my three aunts around Pittsburgh: Ameeta, Vidya, and Preeti. They've helped in ways big and small, from picking me up at the airport when I first landed in Pittsburgh to somehow sensing when my diet is out of whack and driving over with food for me and my flatmate.

My flatmate, Tejus Gupta, gets his own line. Thanks, Tejus. It has been a pleasure to grow together, to discuss what we're reading and watching, and to take on projects together. You're always welcome to move in with me.

My biggest thanks must go to my family; to my parents, my brother, and my grandmoms, whose unconditional support, constant encouragement, everyday sacrifices, and effortless humor, made grad school possible, and enjoyable. How I got to be so lucky, I will never know!

Contents

List of Figures

List of Tables

Chapter 1

Introduction

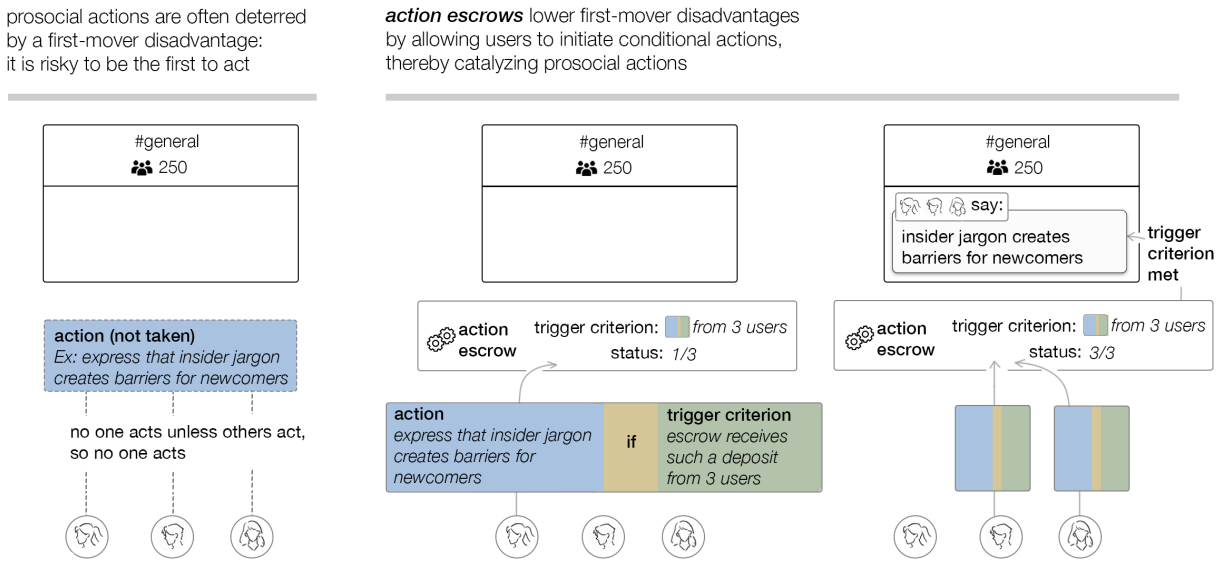

Figure 1.1: This dissertation begins with the unifying observation that many long-standing problems in the online communities literature, such as groupthink, silent majorities, bystander effects, and collective action failures, can all be traced back to first-mover disadvantages: it is risky to be the first to speak up. I then show how a general design pattern—which I call action escrows—can be applied to lower first mover disadvantages and make progress on this broad set of problems.

In an online community, it can feel risky to be the first to express interest in a particular topic, call out misbehavior, or propose acting on a concern. A broad range of prosocial actions that designers of online communities hope to support, including authentic self-presentation, bystander intervention against misbehavior, and collective action, are all deterred by *first-mover disadvantages*: isolated individuals deciding whether or not to take the first action often cannot be sure whether others will welcome it, and back them up.

First-mover disadvantages result in dilemmas where "no one acts unless others act, so no one

acts". Others' actions often provide social proof that one's own actions will be welcome, but this social proof will never exist when everyone is waiting for others to act first. This general structure underlies several classic dilemmas across HCI literature. For instance, the first-mover disadvantage of expressing a diverging perspective results in the *silent majority* effect [25], where a vocal minority set norms in a community because members of the silent majority, unsure if their opinions are shared by others, think it is risky to speak up. Similarly, the failure of a community to intervene on misbehavior is often attributed to *the bystander effect* [100]: where people desire to aid a victim but prevent themselves because they believe it would violate norms. And finally, collective action efforts run into *critical mass problems* [99]: even efforts with widespread private support may never reach the tipping point because individuals are reluctant to make public commitments without substantial support from others.

The consequences? First mover disadvantages **prevent users from initiating progressive interventions** that might actually have substantial support. They also **distort the perceived prevalence of different perspectives**. If no one counters extreme opinions with their moderate takes, then the distribution of opinions in a community can begin to seem more extreme than it actually is [91]. Similarly, if no one intervenes in response to misbehavior, it can make misbehavior seem more acceptable in a community than it actually is [56]. The distortion of norms can also cause **alienation** by giving each user the illusion that they alone are the deviant with beliefs that diverge from everyone else, that they alone are discontented with the status quo [74].

This dissertation proposes that designers of online communities can lower first-mover disadvantages through a general design pattern that I call an *action escrow*—a mechanism that allows users to commit to socially risky actions whose execution is deferred until a prespecified trigger criterion is met. For instance, an escrow for collective action might allow a user to deposit the action of making their commitment public if (and only if) the escrow receives similar deposits from at least thirty other individuals (the trigger criterion). By tuning the trigger criterion, designers can lower the first mover disadvantage. An individual can now place the first commitment into escrow with the confidence that their commitment will only be made public if accompanied by others. They need not worry about whether or not there is substantial support when making their commitment; it will remain confidential if the trigger criterion is not met by other commitments.

1.1 Thesis Statement and Contributions

In this dissertation, I argue that **action escrows—mechanisms that let users commit to socially risky actions that only execute when enough other users make similar commitments—can enable communities to overcome first-mover disadvantages, thereby encouraging users to initiate progressive interventions, revealing suppressed perspectives, and improving feelings of inclusion**. To advance this argument, I make the following contributions:

1. I **define action escrows and delineate their scope and utility**. This offers a name to an existing but loosely applied design pattern in social computing systems. In defining this previously informal design pattern, I reveal its potential to resolve numerous online community issues rooted in first-mover disadvantages. I also describe how action escrows offer advantages over existing behavioral design paradigms in HCI, such as anonymity, extrinsic incentives, and social norms marketing.

2. I **show how action escrows can be instantiated in practice, and describe studies evaluating their effectiveness.** This dissertation develops two systems, Nooks and Empathosphere, each of which advances action escrows further as an approach to addressing first-mover disadvantages. Nooks shows how action escrows can lower first-mover disadvantages in bringing up new topics in a community. Empathosphere shows how action escrows can be applied to give communities translucence into privately held opinions. Through a field deployment of Nooks and an experimental evaluation of Empathosphere, I show how action escrows do indeed encourage prosocial action, reveal suppressed perspectives, and improve inclusion.
3. I **systematize the design space of action escrows.** To inform future implementations, I additionally describe design cases of previously deployed research prototypes and publicly available systems that implicitly instantiate action escrows, and outline the design space of action escrows. Viewing these cases through the lens of action escrows also reveals conceptual bridges between previously unrelated implementations, illuminating how seemingly disparate systems are in fact variations on the same fundamental design pattern.
4. I **characterize the limitations and broader opportunities of action escrows.** I outline the limitations and risks of introducing action escrows into online communities, identify how these risks can be mitigated, and synthesize broader opportunities for escrow mechanisms to address social computing challenges beyond first-mover disadvantages.

1.2 Dissertation Overview

The remainder of this dissertation is organized as follows.

I start by providing a functional typology of situations with first-mover disadvantages in Chapter 2, revealing how first-mover disadvantages underlie several dilemmas described in HCI literature on online communities. In doing so, I map the terrain of problems that action escrows can productively address.

Next, in Chapter 3, I introduce the design pattern of an action escrow, grounding it in traditional escrows used in legal and economic processes. Here, I also describe the advantages that action escrows offer over existing behavioral design paradigms in HCI research—anonymity and extrinsic incentives and social norms marketing—that may also be deployed to mitigate first-mover disadvantages.

Chapter 4 introduces the Nooks application. Nooks supports the creation of action escrows to lower the first-mover disadvantage in bringing up new topics in a community. I describe the design and implementation of Nooks and a field deployment of the application.

Building on the discussion of Nooks, in Chapter 5, I detail a fuller design space of action escrows, outlining the different parameters that can be configured by a designer. To tease apart the underlying design space, I also draw on design cases outside of my work, of three previous social computing systems that can be seen as instantiating action escrows.

Chapter 6 introduces Empathosphere, a system that embeds action escrows in the group chat, and shows how action escrows present a novel opportunity to provide translucence into privately-held viewpoints that would otherwise remain entirely hidden from the community. I describe the

design and implementation of Empathosphere and a controlled study evaluating its effectiveness.

The concluding chapters reflect on the challenges and opportunities that this research raises. Chapter 7, discusses the limitations of escrows, strategies to mitigate potential risks of using action escrows, and highlights immediate opportunities for future work on action escrows. Chapter 8 reviews the contributions of the dissertation and articulates a vision for a future where thoughtfully designed mechanisms can re-engage dormant voices, counter false polarization, and foster inclusion, in digital spaces and ultimately in broader society.

1.3 Situating this Dissertation

My approach to this work is rooted in two disciplines: HCI and social pscyhology. My home discipline, and the discipline to which I primarily intend to contribute with this dissertation, is HCI. Although HCI and social psychology are both interested in how people interact through technology, their goals differ fundamentally. As Alan Newell observed, "psychologists want to understand the world; computer scientists want to change it" [62]. This work engages primarily with the literature and problems in the field of HCI, and my approach prioritizes changing over understanding: I am primarily interested in contributing a new computational approach (action escrows) to addressing long-standing problems. The intellectual tools I explore for studying these problems and the contributions of this dissertation are primarily intended for use by HCI researchers and practitioners.

However, my intellectual approach is also strongly influenced by work in social psychology. Per Deutsch and Krauss, social psychologists are interested in understanding "*how people affect one another*" [22]. They go on to write: "social psychologists are interested in the *conditions* that lead a person to conform to another's judgment, the *conditions* that determine a person's attitude, the *conditions* that lead to cooperative or competitive interrelations. Also, the social psychologist is interested in studying the *effects* of an individual's attitudes on his relations with others, the *consequences* of competitive or cooperative interactions" [22]. Social psychology, in short, provides the tools to represent the causal forces that shape collective behavior—forces that HCI lacks the means or vocabulary to represent and explain. HCI has excellent tools for prototyping novel technical inteventions. However, it does not have tools for representing these invisible forces, and to explain how a computational intervention changes them. It is for this reason that I have drawn on social psychology to develop a causal lens through which to see collective behavior (Figure 2.1), to explain why an intervention works (Figure 3.2), and to shed light on meaningful differences in how different interventions work (Figure 3.2).

So, in this dissertation, I write as an HCI researcher taking advantage of certain social psychology approaches to understanding collective behaviors. I don't claim social psychology expertise but draw from the field to find inspiration to address problems in HCI. This approach advances the field of HCI on its own terms: by abstracting up to psychological contructs, we can see relationships between previously disconnected problems (silent majorities, critical mass) and their technical remedies (action escrows), exposing common roots in first-mover disadvantages. This ultimately increases the generative power of our solutions: we can more clearly see other similar problems to which an approach might generalize.

1.4 Relevant Publications

This dissertation primarily presents research described in the following publications:

- **I'm In If You're In: Action Escrows as a Design Pattern to Achieve Social Change in Online Communities**. Pranav Khadpe*, Lindsay Popowski*, Kyzyl Monteiro, Lindy Le, Geoff Kaufman. *Under Review. (* denotes equal contribution).*
- **Nooks: Social Spaces to Lower Hesitations in Interacting with New People at Work**. Shreya Bali, Pranav Khadpe, Geoff Kaufman, Chinmay Kulkarni. *In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI 2023).*
- **Empathosphere: Promoting Constructive Communication in Ad-Hoc Virtual Teams through Perspective-Taking Spaces**. Pranav Khadpe, Chinmay Kulkarni, Geoff Kaufman. *In Proceedings of the ACM on Human-Computer Interaction (CSCW 2022).*

It also draws on observations described in the following publications:

- **Hug Reports: Supporting Expression of Appreciation between Users and Contributors of Open Source Software Packages**. Pranav Khadpe*, Olivia Xu*, Geoff Kaufman, Chinmay Kulkarni. *In Proceedings of the ACM on Human-Computer Interaction (CSCW 2025). (* denotes equal contribution).*
- **Explaining the Reputational Risks of AI-Mediated Communication: Messages labeled as AI-assisted are viewed as less diagnostic of the sender's moral character**. Pranav Khadpe*, Kimi Wenzel*, George Loewenstein, Geoff Kaufman. *In Proceedings of the AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society (AIES 2025). (* denotes equal contribution)*

Chapter 2

First-Mover Disadvantages

The dissertation focuses on the problem of *first-mover disadvantages*. These are situations where a substantial number of people in a community privately support a progressive intervention, whether a one-time action or a lasting norm change, but it fails to occur because no one wants to intervene first. A common example of this is the familiar classroom dynamic: even though many students may want to request clarification, no one does because they are afraid of asking a stupid or ill-formed question [75]. Or the tale of *The Emperor's New Clothes*, where adults pretend to see nonexistent clothes because they are afraid of causing a scene, or worse, attracting punishment. Being the first to intervene entails social risks: the risk of appearing uninformed, the risk of repelling others with diverging attitudes, the risk of incurring disapproval, and even the risk of attracting retaliation.

First-mover disadvantages show up frequently, offline and online, because they fall out of a common predicament: one wants to simultaneously respond to internal pressure (to take actions consistent with one's own attitude) and conform to social expectations (to take actions consistent with others' attitudes), without visibility into others' attitudes [51, 54]. People have a fundamental desire to take actions in line with their own convictions *and* a fundamental desire to take actions that others approve of. Yet even in face-to-face interactions, like the classroom situation above, there is only so much we can infer about others' private attitudes (how many others desire clarification) from their public behaviors (lowered hands) and appearances (nods that seem to convey understanding). There is always uncertainty about whether acting on our private convictions will attract disapproval. This uncertainty is heightened in online communities, where we cannot physically observe other members, and where the scale of interaction may be so large that, at best, we can try to infer modal attitudes of a sample of community members.

Interaction situations with first-mover disadvantages exhibit dilemmas where "no one acts unless others act, so no one acts." Variations of the idea of first-mover disadvantages have been invoked to explain several classic dilemmas across computer-mediated communication contexts. For instance, *groupthink* [39, 40, 43]—where a group of competent people end up making incompetent decisions—can be traced back to the first-mover disadvantage in expressing a diverging perspective. The first-mover disadvantage of expressing a diverging perspective is also used to explain the *silent majority* effect [25], where a vocal minority set norms in a community because members of the silent majority, unsure if their opinions are shared by others, think it is risky to speak up. First-mover disadvantages also help explain the *bystander effect* [56] in online com-

munities, where users who witness harassment or hate but are not targets themselves, refrain from initiating interventions or counterspeech because it is risky to be the first to do so [100]. Similarly, the *online authenticity paradox* [31]—a substantial number of people actually prefer authentic expression online, but everyone continues to filter and curate their posts thinking it will increase peer approval [48, 117]—can be traced back to the first-mover disadvantage to authentic self-expression [48, 49, 117]. Finally, collective action efforts online run into *critical mass problems*: even community reform with widespread private support may never reach the tipping point because of the first-mover disadvantage of publicly opposing the prevailing norm. Legal scholar Sunstein argues that positive social change requires a critical mass of initial "objectors" who publicly point out problems in collective behavior [99]. However, the strong disincentive to speaking up can prevent any public opposition, causing the change to fizzle out [98].

Significant evidence confirms that these dynamics actually play out in online spaces. Contemporary research studying online political expression in the US has repeatedly run into the silent majority effect: ideologically moderate individuals, despite showing up as the majority in offline polling data, often avoid countering extreme opinions online, because they think themselves to be in the minority and fear negative reactions [50, 66, 79, 91, 94, 107]. Between 60% and 70% Americans have been bystanders of misbehavior directed at others online [24, 100], yet only 30% of them report having intervened [24, 100]. Nearly 45% of social media users think people ought to show more of their "real" selves [68], yet underestimate how many others think similarly and rarely aim for authenticity themselves (only 32% report making the effort) [68].

This dissertation suggest that designers of online communities can lower first-mover disadvantages across a broad range of social situations—including those that we have just described—through the design pattern of an action escrow. In Section 2.1, we present a causal representation of situations with first-mover disadvantages, and describe the feedback mechanisms due to which they persist, or even worsen over time. We will use this representation throughout this dissertation to provide intuition on how first-mover disadvantages are addressed by action escrows as well as other behavior design paradigms. In Section 2.2, we describe the consequences of first-mover disadvantages that action escrows, if successful, should mitigate.

2.1 Causal Representation of First-Mover Disadvantages

Figure 2.1 illustrates the different components of situations with first-mover disadvantages. This directly draws on the representation of *spirals of shame* by Carbone, Dezső, Sunstein & Loewenstein [14], and generalizes it to the context of first-mover disadvantages. .

To begin, when taking action first entails social costs—the cost of appearing uninformed, repelling others, or incurring disapproval—people naturally limit how frequently they intervene or speak up (link "a"). Second, when fewer individuals take an action, that action will naturally occur less frequently in the aggregate across the community (link "b"). Third, people are likely to infer how many people privately support an action at least in part by how often they observe others taking similar actions (link "c"). Fourth and finally, the inferred distribution of private attitudes is one factor that influences judgments of how risky it is to act first—actions that appear to have widespread support are less likely to be seen as socially risky (link "d"). This creates a feedback loop: low perceived support increases social costs, which further reduces action-taking,

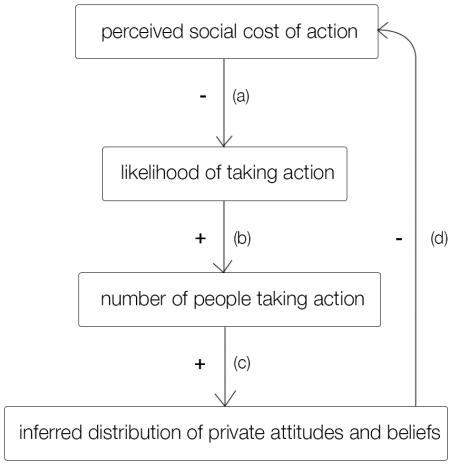

Figure 2.1: The self-reinforcing impact of the perceived social cost of an action (first-mover disadvantage) on number of people taking action and inferences about private attitudes and beliefs across the community. Adapted from [14].

which reinforces the appearance of low support.

First-mover disadvantages, in short, have self-reinforcing consequences: they get worse over time because they distort norms in a community. If no one counters extreme opinions with their moderate takes, then the distribution of privately-held beliefs can begin to seem more extreme than it actually is [91], which further raises the risks of expressing a moderate opinion. This progressive distortion of norms is central to Noelle-Neumann's concept of the *spiral of silence* [80, 116]. Such spirals are common to most situations with first-mover disadvantages. If no one attempts being authentic, then it can normalize filtered and distorted beauty standards that might actually be unrealistic [114], further discouraging authenticity [48, 49, 117]. Similarly, if no one intervenes in response to misbehavior, it can make misbehavior seem more acceptable in a community than it actually is [56]. The distortion of norms can also cause alienation by giving each user the illusion that they alone are the deviant with beliefs that diverge from everyone else, that they alone are discontented with the status quo [74].

The links in Figure 2.1 find support in empirical work across the behavioral sciences and HCI. Perceived social costs do indeed lower people's likelihood to act (link "a"). HCI research on self-censorship behaviors finds that people are less likely to post about or bring up topics in online communities for which they fear disapproval [66, 116]. For instance, in gun discourse on Reddit, many users in pro-gun subreddits refrain from raising moderate points or countering coalitional points because they fear downvotes [66]. The connection between likelihood to act and number of people taking action (link "b") is primarily mathematical: when fewer individuals take an action, that action will naturally occur less frequently in the aggregate, and observable support for it will be lower. The link between the number of people taking action and inferred distribution of private attitudes and beliefs (link "c") arises because people use a mental shortcut: they assume that what others *do* reflects what others *believe*. This is consistent with social psychology accounts of "pluralistic ignorance" [75]. When people see low participation in an action, they conclude that few others actually support it. Even if people recognize that *they*

themselves might hold back from acting due to social risks, they take others' expressions of attitudes and beliefs at face value [75, 92, 102]. So while someone might think "I support this cause but I'm not speaking up because it seems risky," they are likely to look at others' silence and conclude "those people must not actually care about this issue." The final link between the perceived private support for an action across a community and the perceived cost of taking the action (link "d") follows from the source of the perceived cost: the more we think others disagree with us, the costlier it feels to act. Research on descriptive norms shows that when we believe our perspective diverges from the community norm, we expect more disapproval if we take action [9, 10, 14].

We will return to this causal representation throughout this dissertation to provide intuition on how first-mover disadvantages are addressed by action escrows.

2.2 Consequences of First-Mover Disadvantages

From the above dynamics, we can also see that first-mover disadvantages cause three undesirable consequences, which action escrows, if effective, should mitigate:

First-mover disadvantages block progressive interventions. Progressive interventions that would actually have substantial community support never happen because no one wants to take the social risk of going first. Communities miss opportunities for positive reform simply because everyone is waiting for someone else to act.

First-mover disadvantages distort the perceived prevalence of different perspectives. When moderate voices stay silent, extreme opinions dominate the conversation, making the community seem more polarized than it really is. When no one calls out misbehavior, bad conduct appears more acceptable than it actually is. The visible community culture increasingly diverges from what most members privately believe.

First-mover disadvantages cause exclusion. Each person sees the distorted norms and thinks they're alone in wanting something different. Members can feel like deviants for holding certain views, creating widespread alienation even if a substantial number of people share similar concerns about the community's direction.

Therefore, in subsequent chapters, when evaluating action escrows, we will assess whether they: (1) encourage progressive interventions, (2) reveal suppressed perspectives, and (3) improve inclusion.

2.3 Conclusion

This chapter outlines the broad range of situations where first-mover disadvantages arise. We presented a conceptual representation showing how perceived social costs reduce action-taking, which decreases observable participation, which reinforces perceptions of low support. These dynamics result in three key problems: blocked progressive interventions, distorted community norms, and member exclusion. This chapter lays the foundation that we will use to explain *how* action escrows (and other behavior design paradigms) work and to evaluate *whether* action escrows do indeed address first-mover disadvantages.

Chapter 3

Action Escrows: An Approach to Addressing First-Mover Disadvantages in Online Communities

Action escrows unlock coordination by flipping the "no one acts unless others act" problem on its head. Rather than waiting to see who will make the first move, they allow everyone to say "I'm in if you're in" simultaneously. By doing so, they transform the paralyzing question of "will anyone back me up?" into the empowering assurance that "we'll all step forward together"—creating the conditions for prosocial actions that might otherwise never materialize. By making commitments conditional rather than immediate, action escrows bridge the gap between individual hesitation and group potential. And, through automation, computationally implemented escrows can ensure that action proceeds collectively, without the possibility of one person flaking last moment.

Consider, for example, the first-mover disadvantage in expressing a diverging perspective, which can cause the silent majority effect. An action escrow might allow a user to place a diverging comment into escrow with the instruction that it be automatically posted publicly only if the escrow system receives similar comments from twelve other users (see Figure 3.1). Now, the user can submit a comment with diminished fears of the social risks, and with confidence that the comment will only be made public to others in the community, accompanied by twelve other individuals who think similarly. This mechanism effectively lowers first-mover disadvantages for *anyone* wanting to express that perspective, creating conditions for more of them to take individual action, and giving voice to what might otherwise have remained a silent majority.

We define an *action escrow* to be any mechanism that allows a user in an online community to deposit a potentially socially risky action, which is to be automatically executed if (and only if) a prespecified trigger criterion is met. By *action* we mean a one-time event that can occur within the community and is initiated by a community member. We envision action escrows as broadening the space of actions afforded to a user to encompass *conditional* actions. The designer must identify an effective trigger criterion that can lower the user's aversions in the specific context. With an effective trigger condition, action escrows can increase the volume of actions by allowing users to initiate conditional actions where they may have been unwilling to act otherwise.

We present action escrows as a design pattern [2, 11, 33, 52, 69]: a recipe rather than a frozen

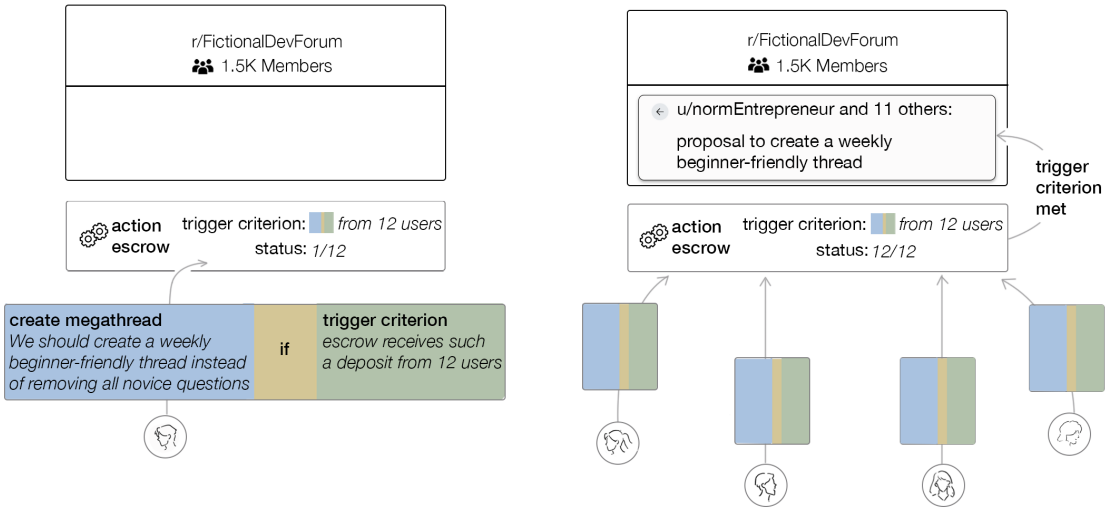

Figure 3.1: A hypothetical vignette to illustrate action escrows in, well, action: In r/FictionalDevForum, many subscribers privately questioned the subreddit's ban on beginner questions, but feared downvote brigades if they challenged the status quo established by vocal power users. Jordan used the subreddit's experimental "GroupSpeak" feature to escrow the post "We should create a weekly beginner-friendly thread instead of removing all novice questions" with a trigger requirement of twelve similar submissions. The interface showed only Jordan "1/12 support escrowed" in their personal view. Within three days, the counter reached 12/12, automatically publishing all the escrowed opinions as a single megathread. The moderators, confronted with this unexpected collective voice rather than what would have been dismissed as one user's complaint, initiated a community vote on implementing weekly beginner threads—transforming what had been silent majority frustration into tangible community governance change.

dinner. Unlike a "frozen dinner" (an existing system that can be used as-is), action escrows are an abstraction that designers can localize and implement for a specific context. Just as a recipe provides core ingredients and techniques that home cooks can adapt with personal touches, action escrows describe the operating principles that designers of online communities can customize to address particular first-mover disadvantages. To support this process, this dissertation also contributes concrete systems demonstrating how the pattern can be applied. Importantly, our definition does not prescribe a specific low-level software implementation; there are multiple ways to implement an action escrow and the specific choice often depends on interoperability with the rest of the system (including the existing API and data model).

In defining action escrows, we extend the game-theoretic "escrow mechanism" to online communities, recognizing its particular aptness for addressing first-mover disadvantages and its enhanced feasibility in digital environments. In this chapter, we describe how action escrows relate to financial and legal escrows, and describe the benefits that action escrows offer over existing HCI behavior design paradigms. In the following chapters, we will see concrete systems that instantiate action escrows.

3.1 Extending the Operating Principle of Financial and Legal Escrows

Action escrows build on the general mechanism of an escrow, which has traditionally been used in the contexts of negotiating settlements [15] and campus sexual assault reporting [4]. The core function of an escrow is to support "conditional intermediated communication" [4]. Escrows, in general, allow a user to make some kind of deposit (a piece of information, an allegation, a monetary offer, or, in our case, an action) into an escrow lockbox with instructions to the escrow agent that the deposit only be released to prespecified recipients under prespecified circumstances. For instance, in escrows used for settlement negotiation [15], buyers and sellers each privately deposit their price—what they're willing to pay or accept. The deposit is made on the condition that the escrow agent only announces a deal if the buyer's price is higher than the seller's price. If the buyer offers less than what the seller wants, no deal happens, and the deposited prices are not revealed.

Action escrows extend the general principles of escrow mechanisms to address challenges associated with first-mover disadvantages in online communities. In this, it particularly draws inspiration from the allegation escrow mechanism [4] proposed by Ayres and Unkovic, which aims to reduce the first-mover disadvantage that prevents victims of sexual assault from coming forward with allegations. In their mechanism, a victim can place a private complaint into escrow with instructions that the complaint be lodged with the proper authorities only if the escrow agent receives, for example, two additional allegation against the same individual. Our work shows that this approach can be extended and effectively leveraged to make progress on problems of interest to the HCI community.

A key difference between action escrows and traditional escrows is their coordination mechanism: action escrows are managed computationally rather than by human intermediaries. Unlike traditional escrows where a human agent manually holds deposits and evaluates trigger conditions, action escrows take advantage of a unique opportunity—they can be directly embedded into the software of the very platforms where first-mover disadvantages occur. They automatically collect conditional commitments, determine when trigger criteria are met, and execute actions accordingly.

Automation enables action escrows to scale efficiently to high-throughput actions such as posting a comment in a community, while maintaining consistent application of trigger criteria across thousands of users. Unlike human intermediaries who might become overwhelmed by volume or introduce inconsistencies in judgment, computational systems can process large numbers of conditional commitments simultaneously, evaluate trigger conditions instantly, and release coordinated actions at precisely the right moment. This makes action escrows particularly valuable in digital environments where many users might benefit from coordination but where traditional human-mediated approaches would be prohibitively expensive or otherwise impractical to implement. Computational management can also provide a layer of psychological safety: it can encourage participation from individuals who would be reluctant to disclose their conditional commitments to a human intermediary due to fears of judgment, gossip, or premature exposure of their willingness to act.

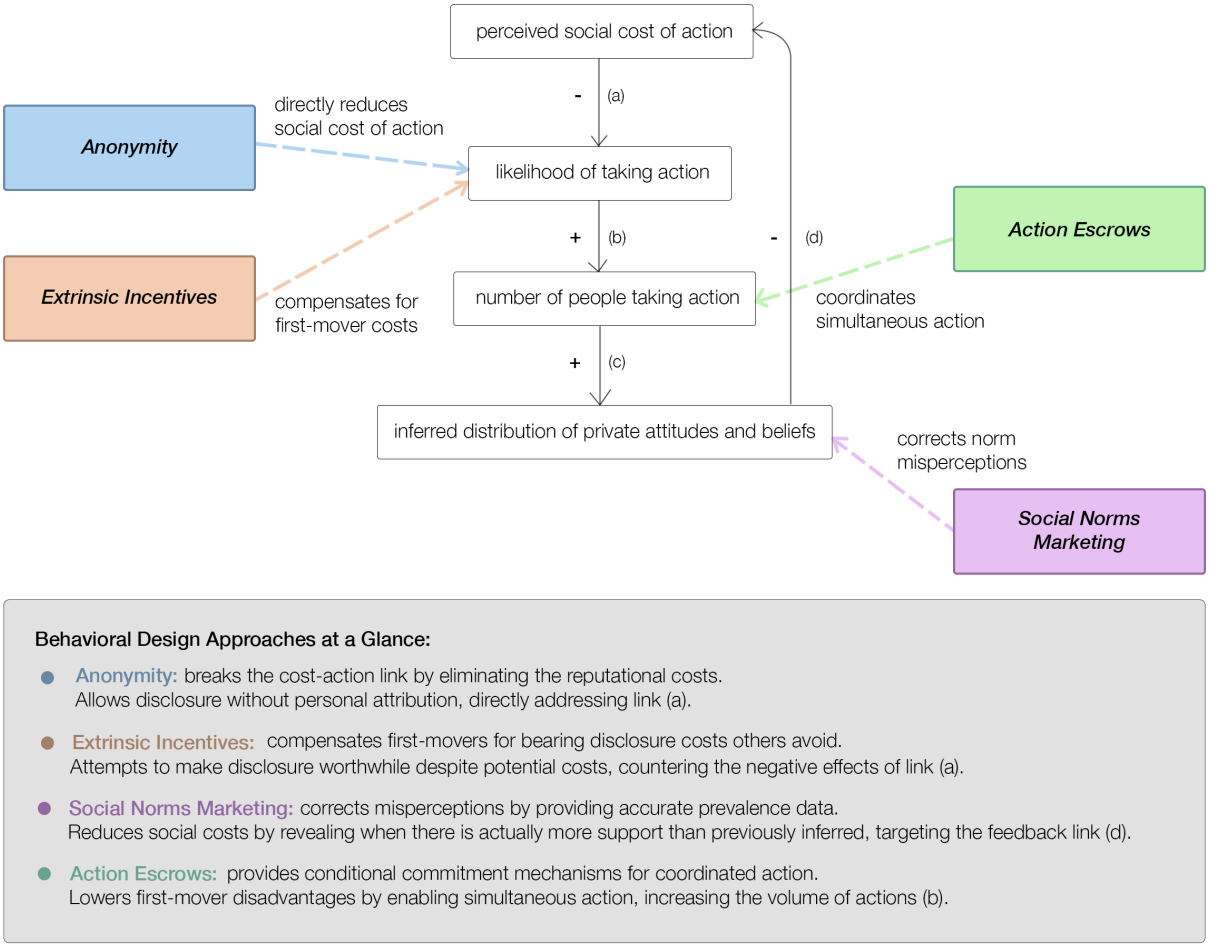

Figure 3.2: Overview of different behavioral design approaches for addressing first-mover disadvantages.

3.2 Advantages Over Existing Behavior Design Paradigms in HCI

Significant HCI research has attempted to address many of these problems that we trace back to first-mover disadvantages. Here we outline three influential behavior design approaches that have come out of this work, and the benefits that action escrows offer over each. Figure 3.2 situates these different approaches relative to each other, showing how they each address first-mover disadvantages.

3.2.1 Anonymity

One approach to reducing first-mover disadvantages is anonymity; when people are anonymous, they face fewer personal consequences for their actions which makes them more likely to take

social risks they wouldn't otherwise take. Anonymity as a paradigm is therefore used in many online social contexts, especially those where a large degree of self-disclosure or vulnerability is desired [65]. Anonymity is especially important and more often used on platforms where the discussion topics or actions are stigmatized and can help assuage embarrassment [89]. Even *perceived* anonymity can be empowering. Perceived anonymity can lessen the spiral of silence effect [112] and even the relative visibility difference of liking versus commenting can affect how much people self-silence [81].

However, even in anonymous communities, social risks do not completely disappear [65]: when user handles persist over time and accumulate reputation, users once again have social capital at stake. This effectively reintroduces first-mover disadvantages as users might fear damaging their carefully built pseudonymous reputation, being targeted for harassment, or losing standing within the community. Action escrows sidestep these concerns.

More generally, action escrows offer a strategic middle ground between full identification and complete anonymity. Unlike permanent anonymity, which hides identities but limits accountability, action escrows enable conditional identity disclosure—participants remain anonymous until specific trigger criteria are met, then identities are strategically revealed. This makes them ideal for situations requiring eventual real-world coordination (like offline gatherings), when verifying genuine support levels is crucial (such as petition signing or collective action pledges), or when communities need the ability to retroactively address harmful actions (like identifying sources of harassment or misinformation). Action escrows provide both the initial safety of anonymity and the eventual accountability of identification, precisely when each is most valuable.

3.2.2 Extrinsic Incentives

If a substantial number of members in a community are reluctant to speak up or initiate actions in line with their convictions, then at first glance, one solution might be to explicitly rebalance activity through extrinsic incentives [53]. Can rewarding participation or penalizing silence address first-mover disadvantages?

Here, one approach is extrinsic incentives that increase *minimum levels* of participation required of each member. Examples include offering badges [70], implementing point-based reward systems for regular contributions [70], or adopting systems similar to karma requirements whereby subreddits gate privileges until a member demonstrates minimum activity [30]. Alternatively, incentives can also directly target *balanced* activity across a community. Collective streak systems, for example, motivate everyone to participate lest they break the group's long-running "streak" [16, 76]. Similarly, visualizing interaction imbalances creates social disincentives that simultaneously discourage individuals from dominating or remaining silent [60, 61].

But because these systems do not directly address perceived risks, they can cause people to falsify their preferences while chasing extrinsic incentives [6]. Accumulating evidence suggests that, when subjected to extrinsic incentives, if people's private attitudes diverge from what they think to be the prevailing majority, then people sometimes publicly align with perceived majority attitudes even if they privately disagree [6, 94, 103, 114]. This can perpetuate groupthink [6], silent majorities [94], online inauthenticity [103, 114], and other dilemmas that arise from first-mover disadvantages. Additionally, if previously silent people provide lip service to a perspective or norm they don't agree with, it can further distort assessments of private attitudes, further

heighten first-mover disadvantages, and can intensify illusions of deviance [84, 85]. Action escrows, by contrast, mitigate these risks of false preference signaling.

3.2.3 Social Norms Marketing

In part, first-mover disadvantages stem from misperceived social norms. So, another approach for researchers, educators, social advocates, or community designers is to directly address these misperceptions by sharing information about the actual distribution of opinions or practices within a population. This is the foundation of social norms marketing [7].

The process typically begins when researchers discover a divergence between perceived norms and private beliefs. On college campuses, for example, students consistently overestimate peer support for heavy drinking—believing their classmates drink far more than they actually do or want to [84]. Armed with this insight, researchers craft corrective messages like "Most students prefer to drink 0–4 drinks when they party" to address the misperceived descriptive norm [21, 109]. These interventions have shown measurable success: students develop more accurate perceptions of peer alcohol use, which correlates with self-reported reductions in drinks per sitting and fewer blackouts [21, 109]. This approach has also been used to promote prosocial action by correcting misperceptions about climate change beliefs [72] (for example, revealing that people underestimate how many of their peers actually hold pro-climate views) and dominant attitudes about women working outside the home, such as male attitudes about women leaving the house for work in Saudi Arabia, which turn out to be more favorable than most men realize [13].

I've applied this social norms marketing approach in my own work with *Hug Reports* [47]. In open source communities, a common misperception exists: while users frequently find software works well for them and developers genuinely want to hear positive feedback, users mistakenly believe that developers don't want or need such feedback. Upon discovering this divergence, our research team developed a simple code editor extension that enables users to easily thank contributors, directly addressing the gap between perceived and actual norms around positive feedback in these communities. See Figure 3.3 for an overview. Beyond my own work, other HCI researchers have also applied social norms marketing to address misperceived norms in digital communities. Das, Kramer, Dabbish and Hong demonstrated how social influence techniques could increase adoption of security behaviors by giving users' an accuration perception about peer security practices [20].

The problem is that social norms marketing only works where researchers think to look: they require external observers to conduct studies and identify specific misperceptions before designing targeted interventions. Additionally, these interventions are static snapshots that address misperceptions documented at particular moments, while community norms and coordination challenges evolve continuously. Action escrows address both limitations: they surface coordination problems organically from community members themselves, and they operate dynamically to handle whatever first-mover disadvantages emerge over time.

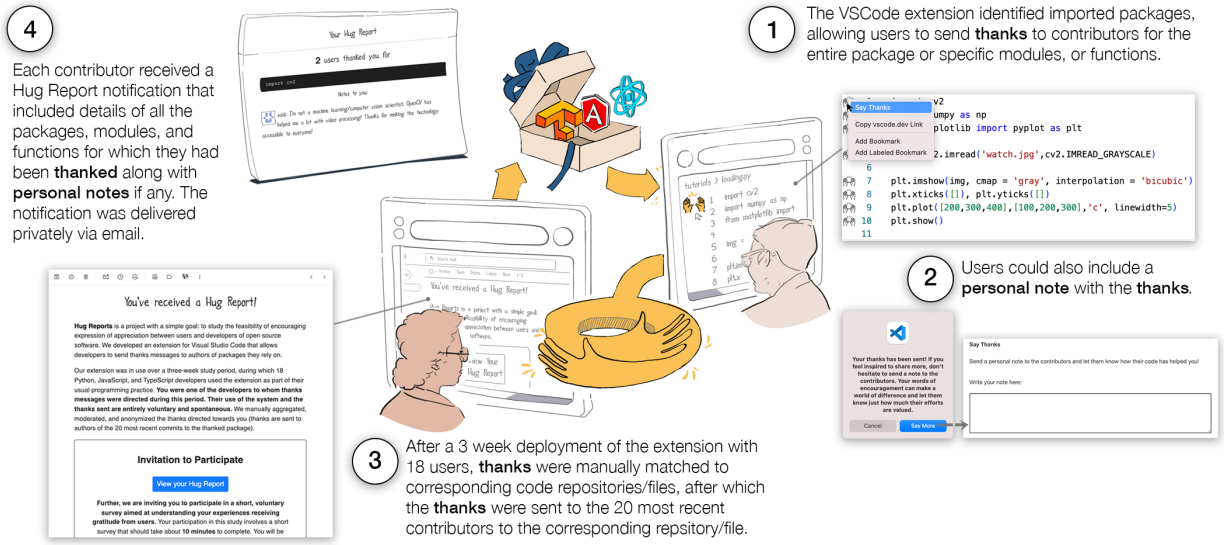

Figure 3.3: Overview of the Hug Reports project.

3.3 Conclusion

This chapter introduced action escrows as a design pattern to address first-mover disadvantages. We described how action escrows build on traditional financial and legal escrows,

Action escrows offer distinct advantages over existing behavior design paradigms. Unlike anonymity, they provide strategic identity disclosure—protecting users initially while enabling accountability later. Unlike extrinsic incentives, they avoid the risk of false preference signaling by addressing perceived risks directly rather than simply incentivizing participation. And unlike social norms marketing, they surface coordination problems organically from community members themselves rather than requiring researchers to pre-identify and study specific misperceptions.

In the following chapters, we will explore concrete implementations of action escrows.

Chapter 4

Nooks: Action Escrows for Bringing up New Topics in a Community

This dissertation suggests that action escrows can help communities overcome first-mover disadvantages. To show this, I will now focus on the often-risky action of bringing up a new topic in a community and demonstrate that action escrows can effectively encourage it.

This chapter focuses on that familiar feeling of wanting to bring up a topic but worrying "what if no one else cares?" or "what if this makes me look weird?" Most of us have experienced that moment of hesitation before posting something new or suggesting an activity in the #general channel of a Slack workspace, unsure whether others will respond with enthusiasm or awkward silence. This creates a frustrating coordination problem where conversations that many people actually want to have never happen simply because everyone is waiting for someone else to go first

Nooks [5] is a Slack application to create escrows that overcome the first-mover disadvantage in bringing up new topics in a community's workspace. It allows individuals to deposit their intention to interact on a topic into escrow, which is revealed only to others in the workspace who have also expressed an intention to interact on the same topic (the *trigger criterion*). The individual making the first deposit (*creating a nook*) can specify the topic. The application reveals the proposed topic (but not the identity of the depositor) to everyone in the workspace and waits 24 hours to receive deposits of interest from others in the workspace (this is the nook's *incubation period*; if a deposit goes unmatched, it expires). Specifically, it asks them to categorically express whether they are interested in interacting about the topic or not ('interested' vs 'not for me'). At the end of 24 hours, it creates a new Slack channel including everyone who has expressed interest in the topic, at which point their identities are revealed to each other. Figure 4.1 shows *Nooks*' action escrow in the context of the following usage scenario.

Usage Scenario: Tejus works on the global incidents response team at TechGiant, a large multinational technology company with employees spread across different time zones. As someone who works night shifts, Tejus is interested in connecting with others in the company who have unconventional work hours, to exchange tips on tackling isolation and managing health and social connections. He is connected to others through Slack and has opportunities to approach them in person, but is unsure about who might be interested and whether bringing this up would

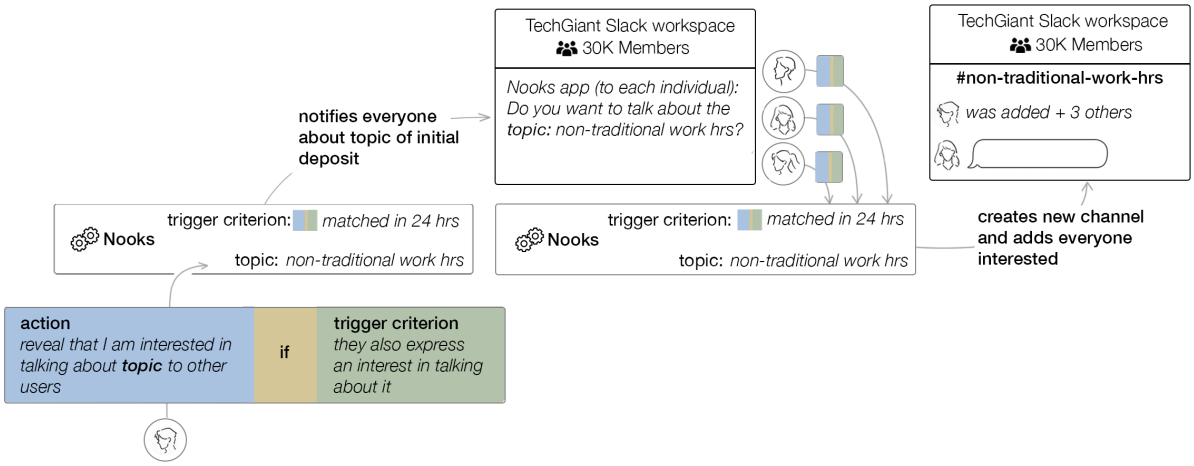

Figure 4.1: Nooks instantiates an action escrow to lower first-mover disadvantages in bringing up new topics in a community. Here, a user anonymously proposes a discussion topic about non-traditional work hours. The topic (but not the depositors identity) is shown to all workspace members. Once others express interest within 24 hours, a new channel is created that includes only interested participants, revealing their identities to each other

be appropriate.

Tejus decides to use Nooks to address this challenge. He proposes a nook on "Exchanging advice for Non-Traditional Work Hours", thus depositing his interest in interacting on the topic. The Nooks application homepage displays the proposed nook to everyone in the TechGiant Slack workspace, without revealing that Tejus initiated it. Priya, who works early mornings to coordinate with European teams, sees the proposed topic and privately indicates her interest. Similarly, Miguel, who splits his workday to accommodate both Asian and American time zones, also expresses interest in the topic. Throughout the day, employees from various departments and regions who work non-traditional hours notice the nook proposal. By the end of the 24-hour waiting period, twelve employees across four different time zones have expressed interest in discussing challenges related to unconventional work schedules. The Nooks application automatically creates a new Slack channel named "non-traditional-work-hours" and adds all twelve interested participants, including Tejus, Priya, and Miguel. Their identities are now revealed to each other, and they can begin sharing experiences and advice without any individual having to risk bringing up the potentially sensitive topic publicly. The channel quickly becomes a valuable resource for the participants, who share strategies for maintaining work-life balance, health tips for shift work, and social connection opportunities. The success of this nook leads to regular virtual meetups among the group and eventually influences company policy on support resources for employees working non-traditional hours.

In the rest of this chapter, we describe the design and implementation of Nooks. We then describe findings from a multi-month field deployment during which Nooks was used by REU students in CMU HCII's summer REU program.

4.1 Nooks: Design and Implementation

Nooks is instantiated as a Slack application. Here, we first describe the design of the application, focusing on how users can start an escrow around particular topic by making the first deposit (creating a nook), and how subsequent deposits are sought and matched to the initial deposit (incubating the nook). Then, we describe how *Nooks* is implemented and describe features that are designed to support adoption and customization.

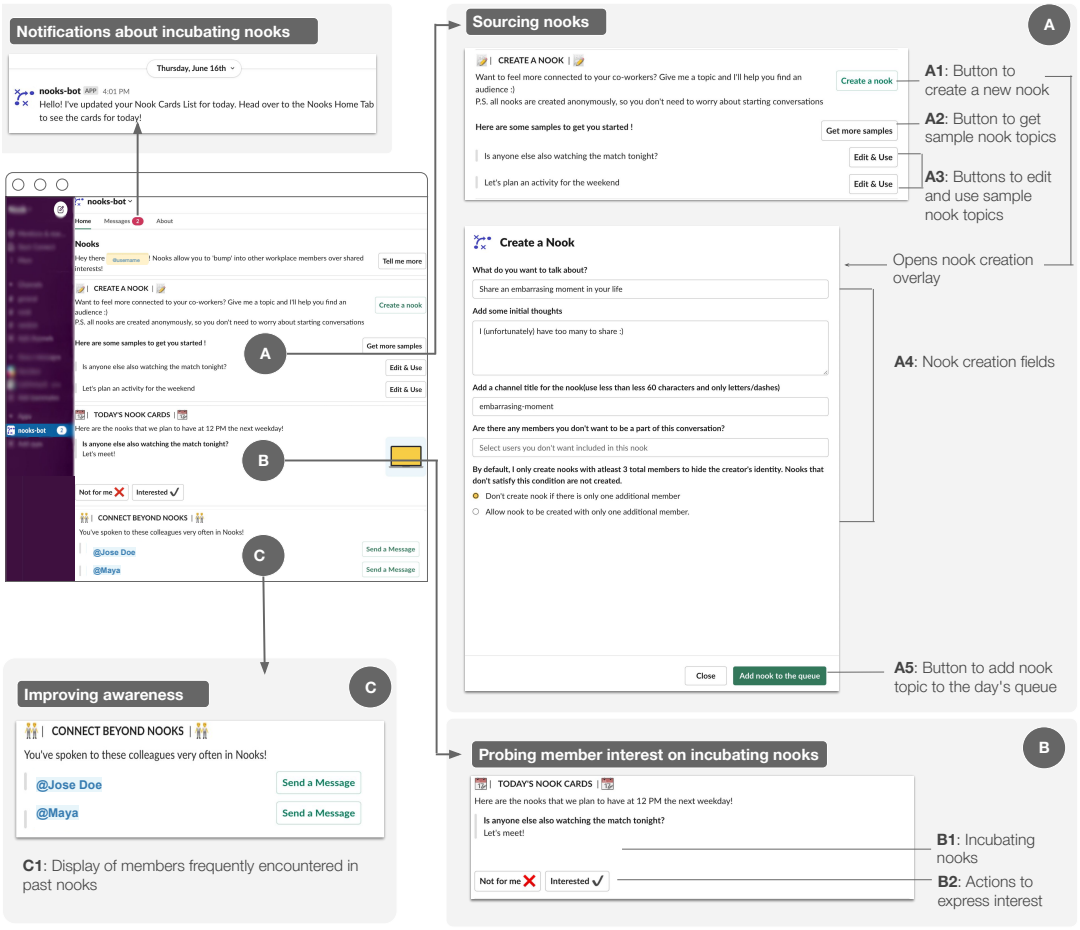

Figure 4.2: A tour of the application homepage. The 'create a nook' panel (A) allows users to anonymously create a nook (A1), as well as edit and use sample nooks (A2,A3). Attempting to create a nook, opens up the nook creation overlay (A4) that asks users to provide a title for their nook as well as their description of what they want to talk about. The homepage also shows users nooks that are currently being incubated (B), displaying the topic and description of the nook (B1) and providing them with the choice to opt-in/opt-out (B2). To alert users of available nooks, the application also sends notifications to users in the form of a Slack direct message. Finally, the homepage also shows users a list of other users they encounter most often in nooks (C1).

4.1.1 System Design

Nooks is a Slack application that manages a dynamic pool of nooks, to which anyone can anonymously contribute a nook that they would like hosted. A continually-running engine incubates nooks by routinely making incubating nooks visible to individuals across the group and probes them about their interest in each. Here, we describe the features of the system, focusing on: (1) how users can make ongoing contributions to the pool of nooks through the *Nooks* homepage (Figure 4.2); and (2) how the continually-running engine routinely probes individuals' interest and activates nooks, with the flow shown in Figure 4.4.

Creating a Nook

To create a new nook, users can proceed to the "Create a Nook" panel in the Nooks home page (Figure 4.2) within Slack. Clicking on "Create a Nook" (A1) opens an overlay form (A4) asking the user to describe the nook they want to create. Specifically, they are asked to provide a 'title' and their 'initial thoughts'. Here they can describe not just *what* they want to talk about but also what they want the norms of the conversation to be. By default, nooks are open for anyone in the group to join, however, the creator can optionally select specific users they don't want included— a control requested by many users in our pilot studies. Finally, they can create the nook by clicking the "Add Nook to the Queue" button at the bottom of the overlay(A5). This panel also shows users a variety of pre-populated sample nooks to inspire users. They can navigate through these samples by clicking on "Get more samples" (A2). Clicking the "Edit and Use" button (A3) opens the overlay form (A4).

Incubating a Nook

Using a user-contributed nook to then assemble a group of people interested in it requires a mechanism that probes people across the workspace about their interest in joining it. Within *Nooks*, each user-contributed nook is "incubated"—other users across the group are shown the topic and description of the nook and asked for their interest in joining a nook. Next, we describe how nooks are incubated and activated.

Incubation and activation: When a nook is being incubated, it is displayed on the *Nooks* homepage (Figure 4.2, panel B) for all members across the workspace, excluding those who the creator has explicitly requested to exclude. Users can click 'interested' or 'not for me'(B2) on each nook they are shown (B1). If multiple nooks are being incubated at the same time, they are shown sequentially. *Nooks* records each user's choice. When asking users about their interest in joining an incubating nook, we only display the 'title' and 'initial thoughts' but do not display any social signals such as how many others have expressed interest. Social signals can cause people to conform to the choices of others [17, 73] instead of choosing nooks based on what they personally find interesting. This can undermine *Nooks*' goal to match people to conversations based on their interest in the conversation and so, we exclude social signals to ensure that users' decisions about which nooks to join are primarily driven by their own interest in the nooks' topics and not by others' choices. At the end of the incubation period, each nook is activated as a private Slack channel (not discoverable to non-members) to which the *Nooks* bot automatically adds only those users who have a priori expressed interest. The bot greets the channel by posting

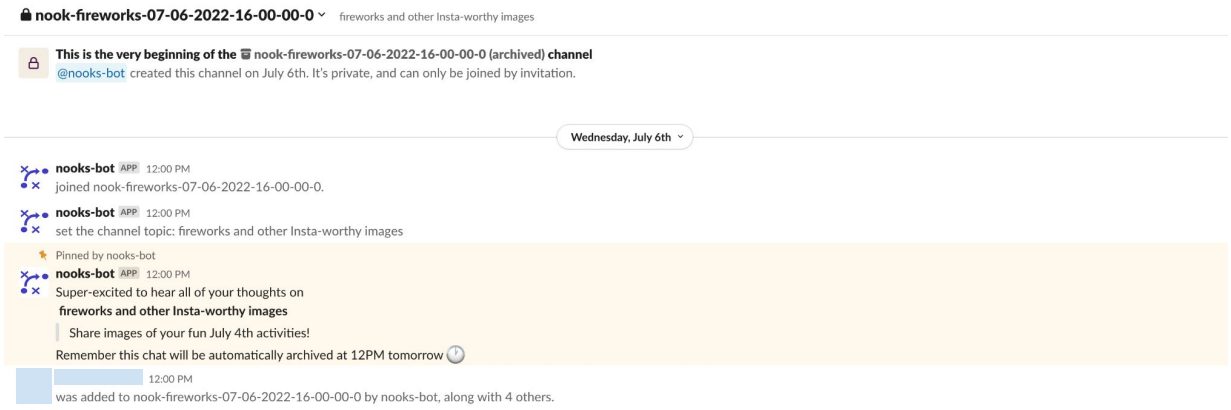

Figure 4.3: At the end of the incubation period, each nook is activated as a private Slack channel to which the *Nooks* bot automatically adds only those users who had expressed interest, including the creator. The bot greets the channel by posting the topic and the initial thoughts added by the creator but does not identify the creator.

the topic and the initial thoughts added by the creator but does not identify the creator. Figure 4.3 shows an example of a nook once activated. Launching of nooks is in no way conditioned on their popularity. Online spaces that optimize for activity attempt to minimize sparsely populated spaces in favor of more populous spaces [1]. This perspective would recommend launching only the most popular nooks. However, *Nooks* focuses not on maximizing activity, but on surfacing new topics in a community which can occur just as well, if not more effectively, in smaller conversations: smaller conversations are more effective at fostering openness and trust [64, 95].

Slack channels that host nooks are always activated at 12pm and open for 24 hours, after which they are automatically archived. Long-lived spaces, and especially persistent spaces, can lead to diffused participation and can impede conversation. If nooks persist by default, it can also create clutter in a workspace. Best practices on the design of online spaces recommend establishing 'expected active times', during which participants can expect each other to be active in the space [1] and consequently, promotes activity in a self-fulfilling way. Since users may not always be active on Slack, and diffused participation can impede conversations, limiting the lifetime of nooks can help establish expected times of activity and promote engaging conversations. For this reason, we set the lifetime of a nook to be 24 hours. Participants are notified when the conversation is archived and if participants in a conversation want to convert the nook into a more persistent space, they have the option to unarchive the channel.

Incubation and notification routine: Nooks houses communities at the scale of a workspace and so, interaction opportunities—incubating nooks—within it can be sparser than interaction opportunities, such as discussion threads, in large online communities. Prior work suggests that when interaction opportunities are sparser, users lower the frequency of their visits [1], which can result in some incubating nooks going unnoticed. To ensure that incubating nooks are viewed by a sufficient number of users, we chose to use a *push model* where users are notified when new interaction opportunities—incubating nooks—are available [1]. However, notifying individuals every time a nook is created can lead to excessive notifications. To simultaneously ensure that

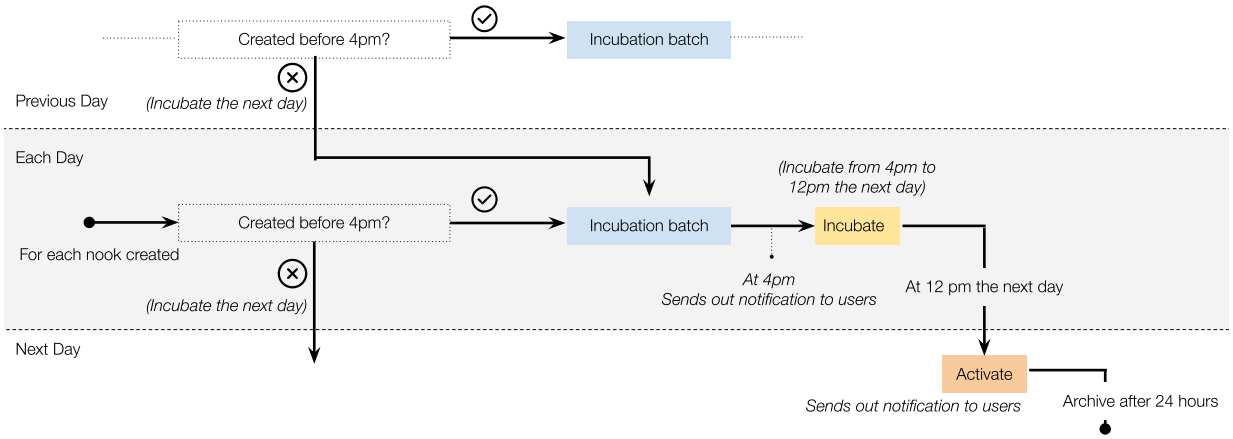

Figure 4.4: The flow of a nook through the execution engine: All nooks created before 4pm on any given day are added to that day's incubation batch and are incubated together from 4pm that day to 12pm the following day. Users are notified that a new batch of incubating nooks is available via a Slack message triggered by the application. Nooks being incubated together are displayed sequentially to users on the homepage. At 12pm the following day, incubating nooks are activated as Slack channels, including all members who have expressed interest in it by then. Nooks created after 4pm are added to the following day's incubation batch.

each nook is viewed by a sufficient number of people and that the interruption cost of notifications is low, we devised a temporal routine according to which nooks are incubated in batches. Figure 4.4 shows this routine. All nooks created before 4pm on any given day are added to that day's incubation batch and are incubated together from 4pm that day to 12pm the following day. Users are notified that a new batch of incubating nooks is available via a Slack message triggered by the application (Figure 4.2). Nooks being incubated together are displayed sequentially to users on the homepage. At 12pm the following day, an incubating nook is activated as a Slack channel and includes all members who have expressed interest in it by then. Nooks created after 4pm are added to the following day's incubation batch.

4.1.2 Implementation

Nooks is a Slack application and a companion Slack bot, implemented in Python with a Flask back-end[1]. We used the Slack Bolt API[2] to create the Slack interface and monitor events that are triggered as users interact with the bot. The back-end was served on a Digital Ocean[3] instance with a MongoDB database[4]. To maintain users' privacy, we do not log any conversations. We collect users' demographic data when they create their profile. Additionally, for every nook created, we record the title and details of the nook created, the user ID of the creator, the user IDs of those who express interest and disinterest in the nook, and finally user IDs of all members

[1]https://flask.palletsprojects.com/en/2.2.x/
[2]https://api.slack.com/tools/bolt
[3]https://www.digitalocean.com/
[4]https://www.mongodb.com/atlas/database

who are added to the nook. When a user first signs-up on the application, they are walked through the consent form within the application. The project is open source and available at: https://github.com/Sbali11/Nooks.

4.1.3 Adopting, Customizing, and Promoting Use Within a Workplace

We envision *Nooks* as a tool that workspace administrators and decision-makers—those responsible for making workspace-wide decisions—can adopt for their workspace as a way to support users in initiating new conversations. Like most social computing systems, for *Nooks* to be useful as a tool for initial interactions, individuals across the workspace need to sign-up as and actively use the application. Once installed in a Slack workspace, *Nooks* allows administrators to onboard members from a specific channel in the Slack workspace, to control who participates in *Nooks*. Onboarding members from the *#general* channel, for instance, would onboard everyone in the Slack workspace. Alternately, they can also invite specific members using their Slack usernames. *Nooks* sends invited users a Slack message, walking them through the sign-up process.

Administrators can further customize *Nooks* to their workspace by editing the sample nooks (A2 in Figure 4.2) to influence the nature of conversations. For instance, they might insert sample nooks relating to a recurring event, if they want to promote conversations about it. Finally, to promote the use of the application, administrators can create a predefined set of nooks that are inserted into the application during set-up. Allowing users to view and join these predefined nooks, provides users with interaction opportunities even before they have created their own nooks. When populated with predefined nooks, by administrators, *Nooks* acts similar to systems that deliver online icebreaking prompts to groups. However, they differ in that prompts are only *suggested*, not *enforced*: users make decisions about what conversations they join. Participating in predefined nooks also supports users in learning *Nooks*' underlying mechanism and allows them to experience conversations in a nook which can inform their decision to create their own.

4.2 Deployment Study

We conducted a nine-week deployment study to investigate how a community used *Nooks* in the field. To maximize ecological validity, we asked participants to use the application however, and as frequently as they desired, and observed patterns of use that emerged naturally. There was no monetary compensation associated with using the app. This study was approved by our university's Institutional Review Board.

Specifically, this longitudinal deployment focused on these research questions:

- RQ1—*How do participants perceive the experience of interacting within a nook?*
- RQ2—*How do participants use Nooks?* What kinds of communication does it afford and what patterns emerge?
- RQ3—*How does Nooks influence participants' sense of community?*

4.2.1 Participants and Research Setting

We worked with administrators of CMU's HCII summer REU program for this study. At the start of the program, a total of 25 students (19 female, 4 male, and 2 non-binary) in the program joined the deployment on a voluntary basis. All participants were aged 18-24. As part of the summer research program, students were employed full-time during the summer and paid a salary as research assistants. These student were primarily engaged in conducting research alongside existing research professionals at the university including faculty members, postdocs, and graduate students. Participants worked in areas relating to computing including smart classroom sensors, educational games, accessibility, and smartphone privacy tools. None of them were taking classes or engaged in other part-time positions. Further, only 20% of the students in the program were affiliated with the university prior to the start of the employment and so, 80% of the students were new to the university and the city. Participating students had desk assignments across four different buildings across the university campus but were within walking distance of each other. Students were fairly mobile, occasionally choosing to work from home. Students were all located in the same city and they interacted with each other both in-person and online.

The program design included some socialization: administrators for the summer research program had already added all the students to a common Slack workspace to create a space for announcements, while also giving them a space to interact with each other. In addition, as part of the program, students were also invited to attend a seminar session twice a week where invited speakers would present their work to the students. Beyond this, since students shared work areas, opportunities to initiate interactions would also arise opportunistically when they would bump into each other.

4.2.2 Procedure

We first worked with the administrators of the program to add the *Nooks* application to the Slack workspace used by the students. Then, we held a short demonstration session for the students during an upcoming weekly seminar. This demonstration session comprised a brief tutorial on how to sign-up on *Nooks*, how students could contribute their own nooks, express interest in contributed nooks, and related functions.

We informed students that *Nooks* was part of a research study and we wanted to learn about their experience using it. We also informed students that if they chose to sign-up, they would be enrolled as participants in our study and that doing so would allow the application to monitor their usage levels for nine weeks, but that the application did not record the content of any conversation. At the end of the demonstration, the program administrators sent all students in the workspace an invitation through Slack to sign-up on the application. To encourage naturalistic behavior, there was no monetary incentive for signing up on or using *Nooks*. Participants were free to use the application as they wanted and could discontinue their use or deactivate the application at any point.

To bootstrap usage, we populated the application with 9 predefined nooks (Figure 4.8) which were basic icebreaking prompts. These nooks were incubated on different days during the first three weeks only. In later weeks, only participant-created nooks were incubated. To capture their

usage of *Nooks*, we recorded the nooks created by participants, as well as the nooks in which they expressed interest or non-interest.

After nine weeks, we invited all participants to additionally participate in a semi-structured interview about their experience using *Nooks*. Nine of the 25 participants accepted the offer to be interviewed and were compensated $15. Interviews were semi-structured and lasted between 20 and 35 minutes. Interviews were conducted in English by one of the authors. Interviews focused on the participants' overall thoughts and perceptions about the application, their experience of participating in nooks, and how they used the application. To help participants recall their experiences, participants were encouraged to open the application and revisit their conversations within *Nooks* while responding to questions. All interviews were recorded and transcribed.

4.2.3 Analysis

Our findings are based on a triangulation of (1) usage data of the participants including the nooks participated in, nooks created, and their expressed preferences among nooks; and (2) transcripts of the recorded interviews. One of the authors performed an initial line-by-line open-coding of the transcripts, iterating over the transcripts as necessary. Codes generated in this phase were in part inductive, driven by the data, and in part guided by our original research questions– we remained open to capturing observations that emerged through the data while also looking out for observations that related to our main guiding questions. Finally, all authors collectively discussed the analysis and iteratively generated, refined, and solidified themes. Themes were generated at a semantic level, reflecting what participants explicitly said [12].

4.3 Findings

Over the 9 week study period, participants used *Nooks* to have 25 conversations. 16 of these conversations were on topics contributed by 7 participants (Figure 4.5). Each of these conversations had between 4 and 15 participants, with a median of 9 participants (Figure 4.7). The remaining 9 nooks were predefined nooks and had between 2 and 5 participants each, with a median of 3 participants (Figure 4.8). Levels of use varied across the participants and across the 9 weeks of the deployment. 22 participants participated in at least one nook (Figure 4.5) and the median number of nooks that participants joined was 6. Activity was concentrated towards the initial weeks of the deployment and decreased subsequently (Figure 4.6). Although every incubating nook was displayed on every paticipant's *Nooks* homepage, the frequency with which individuals viewed the *Nooks* homepage, and responded to incubating nooks (regardless of whether the response expressed interest or non-interest), varied across individuals and across time. As a result, the number of incubating nooks that each individual responded to varies across individuals (Figure 4.5(C)), and the number of users that viewed and responded to each incubating nook also varies across nooks (Figure 4.7 and Figure 4.8). None of the 25 participants in the study formally withdrew participation during the course of the study, however, like previous deployments of social computing systems, we did observe non-use [8, 57], especially 3 participants (P23, P24, and P25) who did not participate in even a single nook (Figure 4.5(A)).

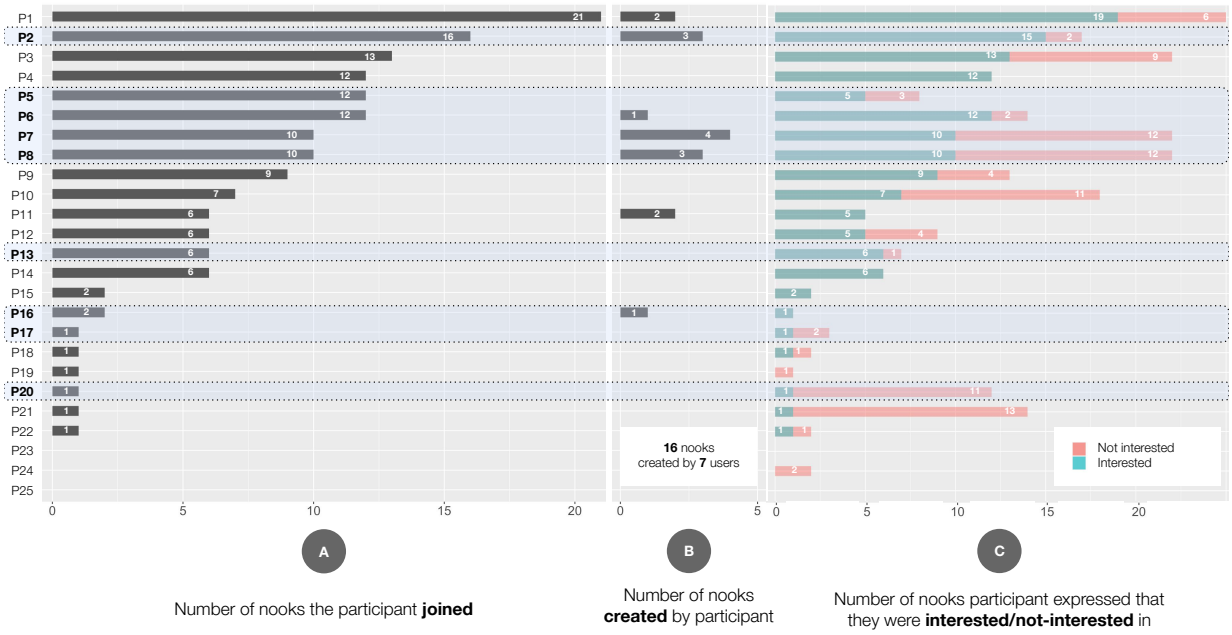

Figure 4.5: Summary of participants' engagement with *Nooks*. Participants who agreed to be interviewed are highlighted in blue. A) The 25 participants varied in how many nooks they joined. The median number of nooks joined was 6. Additionally, 22 participants participated in at least one nook. B) 7 participants contributed a total of 16 nooks C) Participants also varied in how picky they were about selecting nooks to join. In some cases, the number of nooks participants interacted in were more than the number of nooks they expressed interest in (eg. P12, P19) because they were manually added to those additional nooks by interactants in those nooks. Differences in the number of nooks that participants responded to also reveals varying levels of use and monitoring of the application. Responses of non-interest communicate that users viewed the incubating nook but did not want to join it, indicating a low propensity for those specific nook topics. On the other hand, an absence of responses indicates that users did not see the nooks homepage or did not view incubating nooks on certain days indicating a lowered desire to explore incubating nooks on some days.

We invited all 25 participants to additionally participate in interviews and 9 participants agreed to be interviewed. Those who participated in the interviews varied widely in their usage of *Nooks* (highlighted in blue in Figure 4.5) and included 8 female and 1 male participant. Participants in our interviews represent a reasonably stratified sample—5 interviewees participated in more nooks than the median user (> 6 nooks) while 4 participated in fewer than or as many nooks as the median user (≤ 6 nooks).

Next, we present the major themes that we identified in our analysis, each presented as a separate subsection. We first discuss how *Nooks* catalyzed new conversations (Section 4.3.1). These findings speak to participants' experiences interacting within nooks (RQ1). Then, we discuss how *Nooks* helped identify initiatives for which there was critical mass (Section 4.3.2). These findings describe emergent patterns of use (RQ2). Then, we discuss how *Nooks* promoted inclusivity (Section 4.3.3), and provided ambient awareness about interests across the collective

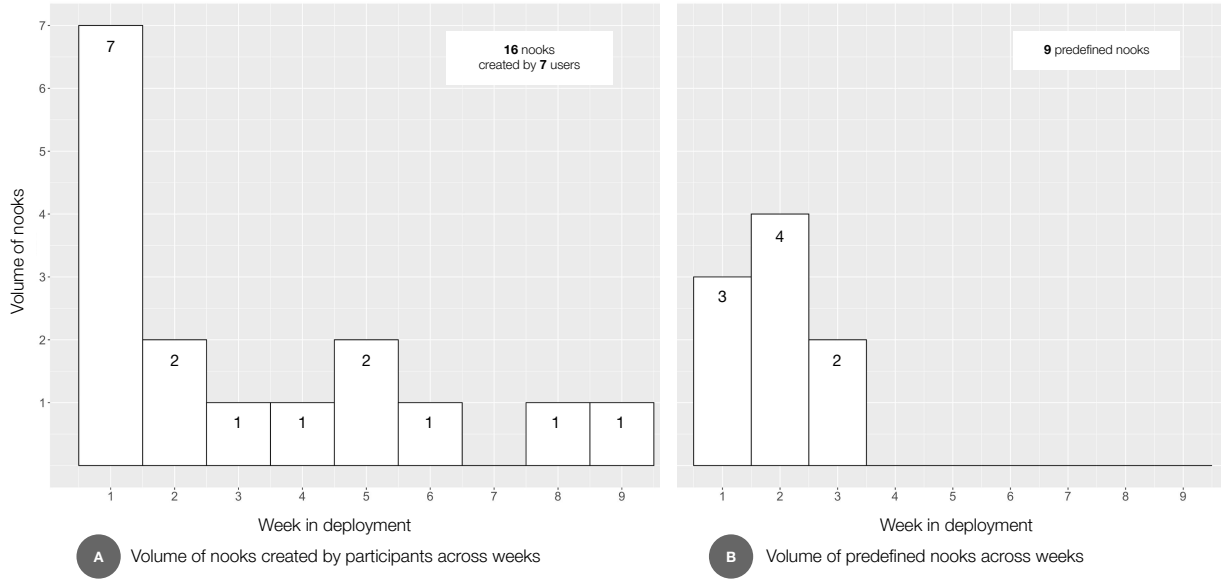

Figure 4.6: Number of nooks activated within the application across the weeks of deployment varied. A) User created nooks, though present most weeks, were concentrated towards the start. B) Nine predefined nooks (that were inserted to bootstrap use) were only activated during the first three weeks.

(Section 4.3.4), illuminating the ways in which *Nooks* might influences users' sense of community (RQ3).

4.3.1 *Nooks* Catalyzed New Conversations

Nooks catalyzed new conversations by (1) lowering the risk of initiating a conversation, by (2) creating social proof that people were willing to engage in new topics, and by (3) making conversations a shared responsibility.

Nooks Lowered the Risk of Initiating a Conversation

Participants mentioned how typical channels to engage in interactions online—such as the large broadcast channels in Slack—were not conducive to initiating interactions because they require individuals to address the group as a whole, which could be intimidating (P5, P13, P16, P17). *Nooks* lowered this barrier, for participants, by providing them with a non-threatening way to initiate and engage in such initial interactions. For instance, P13 mentioned:

> I think it makes it a lot easier to do online conversations, because I guess we don't really have anything to start a conversation with online, especially in a Slack channel where you have, everyone in there. It's kind of awkward to just start a conversation with a 60 person slack so I mean it definitely made it easier there. (P13)

Similarly, P16 mentioned how *"talking to everybody, as a collective can be a bit overwhelming"* while P17 called this a *"psychological burden"*. Because *Nooks* allowed people to find

others through anonymously-created topics, participants suggested that *Nooks* provided a non threatening approach to initiating interactions:

> I think it's less intimidating and it's also anonymous so I think people who might not be super extroverted or [those uncomfortable] just throwing an idea out there [to the whole group], would feel most comfortable [using Nooks]. Because you can actually gauge whether people are interested or not...and most of the time, at least a couple of people will be interested in your topic. I think it's less intimidating and can really help when you're new to a workspace. (P5)

By establishing that a topic was acceptable to talk about within a nook, and that everyone within the nook had opted into this norm, *Nooks* eliminated P17's hesitation about engaging in casual conversations with others. Other participants described how *Nooks* contributed an *"informal environment"*(P2, P5) and provided a venue for conversations that are less "important" than you might post in the larger channels (P2), suggesting that *Nooks* provided an alternate sociable sphere for casual conversations. Additionally, because conversations within *Nooks* usually involved groups of people, it increased participants' levels of comfort going into conversations (P7,P8). P8 noted, *"you know going into the group, you have more than one other person you're going to be interacting with and so it's not going to be this awkward one on one thing"*.

Nooks Created Social Proof That People Were Open to New Conversations

Existing activity within *Nooks* stimulated participants to engage within it, creating social proof that people might be interested in new topics. P5 mentioned how others' use of *Nooks* prompted her to participate and use it as a way to find new connections : *"I think it definitely encouraged me to see others using it and trying to find each other with Nooks"*. Participants also mentioned how predefined nooks reminded them that they could contribute their own conversation topics and stimulated further use (P5,P8).

Nooks Made Conversations a Shared Responsibility

Through our interviews, we found that *Nooks* distributed the responsibility and ownership of the conversation, across the group, unburdening creators' of the pressure of social evaluation (P2, P6, P7, P8, P16). Thus, the onus of driving the conversation forward was on everyone in a nook and not just the creator (P2, P6, P7, P8, P16). As a result of this, in some cases, conversations gained momentum even without the creator driving the conversation. P16 mentioned how by the time she joined the conversation in a nook she had created, the conversation had already taken off: *"I think I was busy or doing something in the hour that it was created and by the time I checked my messages someone had already messaged the chat and there was a conversation going on. So it indicated to me that people are actually interested in this. I'm glad I created it."*

However, because the creator was no longer solely responsible for the conversation, it occasionally led to failures through social loafing, where the conversation failed to take off as no participant drove the conversation. P8 mentioned how even though joining a nook communicated intention, there were still situations where people remained inactive within a nook. P6 and P7 suggested that nooks were more likely to succeed when the creator made the effort to steer the conversation. P7 mentioned:

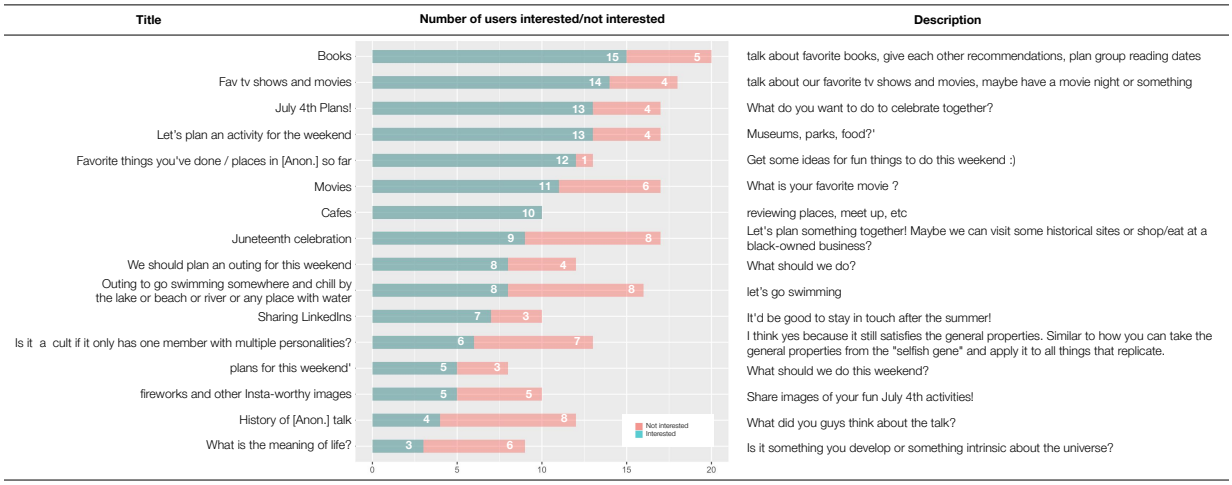

Figure 4.7: Nooks created by users. Of the 16 nooks created by the participants, 10 nooks were attempts at initiating offline activities. In 3 of these 10 nooks, the nook was created with a dual intention of facilitating interest-based interaction as well as planning a related activity. 'fav tv shows and movies' was one such nook, described as a space to 'talk about our favorite tv shows and movies, maybe have a movie night or something'. Each of these conversations had between 4 and 15 participants, with a median of 9 participants. Here, the total number of responses against each nook indicates the number of participants that viewed and responded to the nook when it was incubating. Although every incubating nook was displayed on every users' *Nooks* homepage, some nooks were viewed by fewer people (eg. 'plans for this weekend' was viewed by 8 people whereas 'Books' was viewed by 20) because the frequency with which individuals viewed the homepage varied across individuals and with time.

> I feel it would be nice for the person who created the nook to start the conversation or to have a description of what they want to say first and have that be in the nook when you join, so the conversation has already started and you don't have to wait for someone.

4.3.2 *Nooks* Helped Identify Initiatives for Which There Was Critical Mass

Participants Overwhelmingly Used *Nooks* as a Way to Initiate Offline Activities

Of the 16 nooks created by the participants, 10 nooks were attempts at initiating offline activities (Figure 4.7). In 3 of these 10 nooks (Figure 4.7), the nook was created with a dual intention of facilitating interest-based interaction as well as planning a related activity. For instance, one nook titled 'fav tv shows and movies' was described as a space to 'talk about our favorite tv shows and movies, maybe have a movie night or something'. Similarly, a nook titled 'books' encouraged participants to 'talk about favorite books, give each other recommendations, plan group reading dates'. In the remaining 7 nooks, the intention was primarily to plan an activity. As an example, one nook was titled 'Let's plan an activity for the weekend' that invited participants who were interested in visiting museums, park, or grabbing food. Nooks that focused on planning activities

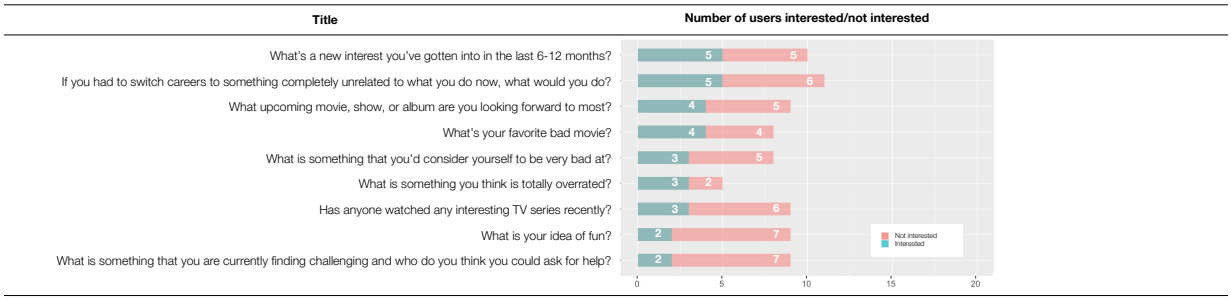

Figure 4.8: Predefined nooks were basic icebreaker prompts intended to catalyze initial use. These nooks did not have a description attached to them—only a topic. 9 predefined nooks were activated during the study period and had between 2 and 5 participants each, with a median of 3 participants. Here, the total number of responses against each nook indicates the number of participants that viewed and responded to the nook when it was incubating.

also generated more interest. The 10 nooks that were about planning activities had an average of 11 participants. In contrast, nooks that did not focus on planning activities had 7 participants on average.

Multiple participants described how they were primarily interested in connecting with others offline (P2, P5, P7, P8, P17) and that they used *Nooks* to find new opportunities to do so. P8 mentioned how she was enthusiastic about *"making things leave the group chat"* and so, was *"more interested in nooks that could lead to something outside of the online context."* Other participants mentioned similar motivations for creating and participating in nooks.

> Since there was like no one in my office—it was just me—I didn't get to interact with that many students in-person. So, I used [nooks] to make plans with them so that I could hang out with them later on. I thought it helped me a lot in talking more with others and hanging out with them which I know wouldn't have been possible if I relied on in-person [interactions], because, we're all kind of scattered across [campus].

And Things Actually "Left the Group Chat"

Activities that were proposed in some nooks did end up happening. For instance, P8 mentioned:

> One nook that worked out well was a cafe nook and we actually did go out and get food together...a few people in the nook as well. Actually, the interest was started in the nook and we translated it into our food channel that's always open and just shared a broader like wider invitation for anyone else to come.

Because of this, *Nooks* became a way for people to deepen their relationships offline. P17 mentioned how she was interested in hanging out with people and how *Nooks* helped her meet 'new people': *"[Nooks helped] with meeting new people in some ways. We both knew each other, because we have the weekly seminar in our program but we hadn't talked to each other, before."* Similarly, P2 mentioned how a desire to be included in offline activities prompted her to participate in most *Nooks*: *"I am worried that people might plan stuff and then I'll miss out on it. I just join all the nooks to see what's happening"*.

In Fact Participants Found *Nooks* Especially Useful for Planning Offline Activities

Participants repeatedly described how the collective practices that emerged around *Nooks*, made it a uniquely effective means to *"generate momentum"*(P16) for activities and events, and provided a *"jumping off point to engage socially outside of slack"*(P6). P7 described how expressed interest within *Nooks* increased chances of plans materializing: *"[if I try to initiate plans in #general], there's a chance I just get like an emoji reaction on [it] and I didn't want that, so I just used a nook because I knew that I would actually get a response from those [who] were actually interested"*. A nook provided a centralized venue to coordinate and make a plan. For instance, P2 mentioned:

> It (my invitation) might get lost if I put it in general, so I was hoping that having a [activity-related] nook would be the kind of place where people would be centralized.

The intentions established on joining a nook and the short duration of the conversations, accelerated scheduling: *"because there was like that short time period, but also the nature of the conversation was about cafes to go to, the scheduling was something people had to figure out and did kind of make those plans"* (P16).

Nooks were also effective as a way to "prototype" activities and assess whether there was support. P17 noted how in one nook, *"the person who created the nook said that she was going to some places with the student she works with and she was wondering whether other people were interested in joining them."* Similarly P5 noted how a conversation with another participant led to the creation of a nook: *"my project partner and I were thinking of playing tennis and then we were just like okay let's collectively create a tennis nook to see if people are interested."*

Nooks also allowed people who share an interest to find each other and *collectively* propose an idea to others: *"with the fourth of July plan, we shared the chat inside general so we can get [more] people"* (P2).

4.3.3 *Nooks* Promoted Inclusivity

Nooks Made It Easier to Include Others in a Conversation

P16 mentioned how *Nooks* contributed feelings of inclusivity by naturally including everyone interested in a conversation:

> Creating a channel and adding a bunch of people intentionally is bound to create some conflict, if you forget a person or something so I thought [Nooks] was a low effort way to allow other people to opt in.

In typical chat conversations, everyone is excluded by default: the creator is required to invite people they want to the conversation. In contrast, *Nooks* includes everyone by default, while providing creators with controls to exclude members. In doing so, *Nooks* naturally prioritizes inclusivity. While this is the very mechanism that makes them useful for initiating interactions, they can often be ill-suited for interactions where the initiator wants to include only specific individuals in the conversation. For instance, P2 recognized how nooks takes away some control over who participates in a conversation: *"[nooks] can only be sent to the entire group and you can't decide who goes in"*.

***Nooks* Also Made Participants Feel Included in the Program and Helped Them 'Find Their People'**

By bringing people in contact with others over shared interests and making the shared interest the focus of conversation, *Nooks* promoted inclusivity by making pre-existing ties less pronounced within conversations. P17, who joined the program late, mentioned how she had fewer opportunities to initiate interactions with others because they had already formed tight-knit groups:*"I still remember in our first seminar I just came and stayed in the last row. I didn't know who to talk to, and even people who were sitting next to me seemed like they didn't really want to talk to me and were just talking to people they already knew."* P17 mentioned how *Nooks* gave her a chance to get to know others better, reflecting on how otherwise she might have just known what others in the group *"look like and probably their name."*

Activity within *Nooks* decreased after initial weeks. Participants attributed this to the fact that as they had a better understanding of each others' interests, and once they had formed closer relationships with some people, they didn't feel a need to use *Nooks* (P5, P7, P8, P13). P5 mentioned:

> I think for me personally, it was a really good initiation tool. Once we got to actually know some people I think I just didn't use [*Nooks*] as much because I felt like I already knew who is going to respond to what. Later on, I could tell [who was interested in what].

P8 noted, *"halfway through, a lot of people kind of found their little groups and have mostly stuck to them, myself included"*. When participants had already formed closer relationships, *Nooks*' approach to facilitating conversations became constraining: *"once you know your friends, you don't need to be able to text them between a 24 hour period and only about certain topics"*(P13).

4.3.4 *Nooks* Provided Ambient Awareness About Others' Interests and Their Desire to Connect

Conversations in *Nooks* improved participants' understanding of each others' interests and often led to conversations beyond *Nooks*. P13 mentioned how a conversation within *Nooks* led to an extended conversation offline: *"We were both in a nook about TV shows, and then we were just talking about it, the next day, and I was like oh, I saw your response in that as well, and so we were able to build off the conversation a little bit more"*. Similarly, P5 noted *"I've even had conversations about [the nook topic], offline like not in a nook."*

In many cases, participants were successful in initiating offline activities through conversations in *Nooks* which led to new connections. P6 noted how a plan that emerged within a nook led to him meeting two new people. Similarly, P2 mentioned how she was able to get a group of people together to visit a park and watch fireworks. Participants mentioned how these experiences helped them get closer to *"familiar strangers"* (P5, P7, P17). For instance, P5 noted, *"these are people I had already met, but in a very loose sense, you know. It's like I've seen them around but it's definitely not like we've been in really friendly situations before"*.

Beyond the awareness that emerged through participating in conversations, P5 mentioned how observing activity in *Nooks* itself contributed an improved awareness of others' interests and their desire to socialize. P5 notes,*"you don't even need to interact with other people to*

learn what they are interested in. Because, you, swipe through the nooks and you see oh there's a group of people interested in like board games, or something. I think it's even useful just to gauge what the interests of your colleagues are" (P5). None of the participants used the option to 'send a message' to commonly encountered individuals within nooks. Instead, as our findings indicate, they used alternate pathways to continue interactions such as by scheduling offline activities through the nook or by approaching each other opportunistically offline. However, the signals from this feature still contributed to an improved awareness about colleagues' propensity for socialization. By observing who was active in nooks—aided by the list of individuals they encountered most commonly—participants were able to identify people that were interested in socializing: *"It allowed me to see [who was] willing to connect with other people and would be down to maybe have a chat"* (P5).

4.3.5 Deployment Limitations

Our deployment revealed that participants were able to use *Nooks* as a way to initiate online and offline interactions. That they personalized their use of the application to meet their needs is further evidence that they found it useful. However, our field deployment was not a field *experiment*, because it did not include an appropriate control group. The lack of a control group prevents us from finding statistical benefits to using *Nooks* (for example through a pre- and post-study comparison). As with other studies that trade off control for ecological validity, finding a control group in our setting is challenging, as social environments are complex and their evolution is path-dependent [93], such that dynamics among group members and their social relationships can yield widely varying states.

Like other prior work in sociotechnical systems [8], we also found it challenging to study non-usage. Although our interviewees had varying levels of usage (Figure 4.5) and included participants that engaged minimally within *Nooks* (P17, P20), participants that didn't join a nook at all (P23, P24, P25) did not participate in our interviews. As a result, though we were able to characterize some factors that led to reduced usage, we were unable to investigate why users might *entirely* avoid interaction after signing-up on *Nooks*.

Some aspects of our deployment context also complicate the interpretation of our findings for understanding how action escrows might reveal suppressed perspectives in online communities. Our participant pool was drawn from a small, bounded community of summer research students, which differs significantly from many online communities where perspective suppression is most problematic. First, the community was relatively homogeneous, with participants sharing similar demographics (all college students aged 18-24), educational backgrounds, and professional contexts. This homogeneity may have limited the diversity of potentially suppressed perspectives that could emerge through Nooks. In contrast, many online communities grapple with perspective suppression precisely because they contain diverse viewpoints that create social risks for minority opinion holders.

Second, the temporary nature of the community (nine weeks) and the absence of significant power differentials may not reflect the dynamics that lead to perspective suppression in many online communities. Persistent communities often develop entrenched norms and hierarchies that make certain viewpoints feel risky to express, while communities with formal or informal power structures may suppress perspectives that challenge authority or dominant groups.

However, the core mechanism that Nooks provides—allowing anonymous testing of whether others share similar perspectives before revealing one's identity—addresses fundamental barriers to authentic expression that exist across diverse online communities. While our deployment primarily revealed shared interests rather than suppressed opinions, the underlying action escrow pattern could be adapted to surface minority viewpoints, challenge false consensus, and encourage authentic expression in communities where such perspectives face greater social risks. Further investigation in communities with more diverse membership, controversial topics, and established power dynamics would help establish how effectively action escrows can reveal genuinely suppressed perspectives.

4.4 Conclusion

This chapter introduced *Nooks* and showed how action escrows can be applied to lower first-mover disadvantages involved in bringing up new topics in a community. Our field deployment revealed how *Nooks* catalyzed new conversations, it helped identify initiatives with critical mass, and it promoted inclusivity in the community.

When it is unclear whether a particular affinity or norm is welcome in a community, Nooks encourages users to "test the waters" rather than remain silent. By allowing anonymous proposals for private discussion spaces, it creates a low-risk way to gauge interest without social exposure. This mechanism can help uncover the existence of silent majorities—groups of people who share affinities or concerns but haven't voiced them due to perceived social risks. The mechanism can be especially useful for spawning *counterspaces* [86], where individuals can experiment with norms and affinities that are untested in the community's public forums. For example, proposing a space for authentic sharing could reveal widespread desire for vulnerability, directly addressing the online authenticity paradox.

An interesting emergent effect was that participants gained ambient awareness of collective interests simply by observing proposed nooks *even before they were triggered*. We will take advantage of this observation later on, in Chapter 6.

Before that, in the next chapter, we develop a design space of action escrows, describing the parameters that designers can configure.

Chapter 5

Design Space of Action Escrows

In this chapter, we build on our discussion of Nooks to characterize the design space of action escrows. We first introduce the two key parameters that designers must configure when creating an action escrow: the trigger criterion and the interim disclosures. Then, we work through three additional illustrative design cases of existing action escrow systems. These cases enlist action escrows to lower first-mover disadvantages in three additional contexts: (1) planning a collective action effort; (2) forwarding content into public forums; and (3) admitting romantic interest. Through a comparative analysis across Nooks and these three cases, we can tease apart the action escrow design space by revealing how different parameter configurations enable the pattern to work across varied social contexts. Through this comparative analysis, we can move beyond seeing action escrows as a monolithic pattern to understanding the rich space of design possibilities and the principled ways that designers can configure them for different community needs.

5.1 Key Parameters of an Action Escrow: Trigger Criterion, and Interim Disclosures

Designers of online communities can use the design pattern of an action escrow to encourage a broad class of actions for which there is a first-mover disadvantage. Such actions can include publicly signaling interest in a collective action effort (e.g. commenting "I'm in!"), or as in the case of Nooks, starting a conversation about a topic that hasn't yet surfaced in a community. Once a designer has identified the action they hope to support, to set up an action escrow, they must make decisions about two key parameters: the trigger criterion, and the interim disclosures.

5.1.1 Trigger Criterion

Action escrows lower the first-mover disadvantage by allowing users to initiate a conditional action, where the action's execution is contingent on a prespecified *trigger criterion*. For instance, an escrow system can offer to keep a user's signaled interest private until the system has received a prespecified number of complementary signals from other individuals (e.g. comment "I'm in!" if 40 people are in).

The design cases we describe here use two primary types of trigger criteria: *activation thresholds* and *reciprocal deposits*. The above example—where a public signal of interest is withheld until it can be accompanied by complementary signals—employs an activation threshold. Activation thresholds lower first-mover disadvantages by creating *ambiguity* about who the first-mover is, thus promising to distribute the consequences, if any. On the other hand, action escrows triggered by reciprocal deposits employ a different psychological mechanism. For instance, in the case of Nooks, a user deposits their interest into escrow, and the system connects them only with others who indicate matching interest. This creates *social assurance* by ensuring interactions occur only among community members who have explicitly expressed prior interest in the discussion.

5.1.2 Interim Disclosures

A designer also needs to decide how, if at all, members of the community are notified of the escrow deposits that are waiting for their trigger criterion to be met: through *interim disclosures*. For example, it is possible to make members of the community aware of the aggregate number of individuals who have currently submitted a signal of interest in a collective action effort, or how many individuals have expressed interest in talking about a particular topic, without revealing individual's identities (disclosing *progress towards trigger*). Revealing the level of support can catalyze follow-on deposits by reducing uncertainty about the viability of the proposed action. But disclosing the level of support is not always desirable. For potentially viable efforts that are just slow to get off the ground, it can convey lack of momentum and prematurely kill effort that might have succeeded. It can also enable targeted opposition before sufficient support has developed. In contexts where these concerns matter, designers can choose to reveal less—simply notifying the community that interest in a certain topic or collective effort exists, without disclosing the initiator's identity or the number of sequent deposits (disclosing *only receipt of first deposit*). With Nooks, we opted for this latter approach. The cases we describe next disclose either progress towards trigger or receipt of first deposit.

5.2 Design Cases

The cases we present include research prototypes and publicly available systems. In selecting cases, our goal was to move beyond Nooks to demonstrate the broad potential of action escrows and display some possible configurations for the key parameters.

While all these systems follow our definition of action escrows and address first-mover disadvantages, neither the research prototypes nor the publicly available systems describe themselves using this terminology. One of this dissertation's contributions is highlighting the conceptual similarities across these diverse systems and domains to introduce a shared vocabulary with which to discuss them. Through this, we hope to shed light on a common design pattern that has received limited attention thus far.

Due to the absence of a shared vocabulary (and therefore common keywords), our selection process was naturally limited to systems we were familiar with; we couldn't, for instance, exhaustively aggregate papers using a keyword-based search. However, we believe this initial

collection provides a strong foundation for understanding the design space, while also offering designers concrete jumping off points to begin adapting and implementing action escrows for their own context.

5.2.1 Catalyst: Lowering First-Mover Disadvantages in Committing to Collective Action Efforts

Catalyst [17] supports the creation of escrows that overcome first-mover disadvantages in publicly commiting to collective action efforts. It is a web platform that also integrates email messaging. It allows individuals to deposit their commitment into escrow, which is only called in if the number of deposits reaches the prespecified activation threshold (*trigger criterion*). The individual making the first deposit can specify the cause and the activation threshold at which commitments are made public. sequently, others can submit their commitment to the cause (a categorical 'join up' or not) into escrow if they think the activation threshold is above their personal threshold: where the benefits of public commitment outweigh the drawbacks. Until the trigger criterion is reached, Catalyst reveals the cause of the escrow, the activation threshold, and the aggregate number of deposits received so far, so that previous and potential depositors can see the current status of the cause (*interim disclosures*). Thus, Catalyst uses an *activation threshold* as its trigger criterion and makes interim disclosures about the *progress towards trigger*. Figure 5.1 shows Catalyst's action escrow in the context of the usage scenario described next.

Usage Scenario: Riley is a member of a large creative content platform where thousands of artists share their work. The platform has recently announced controversial new terms of service that would claim partial ownership of all user-created content. Many creators are upset but hesitant to speak out individually due to fear of being targeted, shadow-banned, or losing their audience.

Riley creates a Catalyst escrow called "Creator Rights Protection Coalition" with an activation threshold of 500 verified creators. The system is configured so that no individual names will be publicly revealed until the threshold is reached, at which point the platform would receive a collective statement via email that the enlisted creators are prepared to simultaneously leave the platform on a specific date if the policy isn't reversed.

Riley shares the secure link through trusted Discord channels and private creator groups. Jordan, who has built a modest following of 10,000 fans over three years but depends on platform income, sees that 275 creators have already committed. They join the coalition.

Over the next week, word spreads carefully through creator networks. Taylor, a highly influential creator with over a million followers who has previously been given special treatment by the platform, has been hesitant to take a public stance despite private concerns. After seeing that 499 other creators have committed, Taylor becomes the 500th participant, pushing the escrow over its threshold.

Once the threshold is reached, Catalyst automatically emails the collective statement on behalf of the coalition announcing their unified stance, and notifies all depositors. The platform now faces the prospect of 500 creators simultaneously announcing their departure unless the terms are

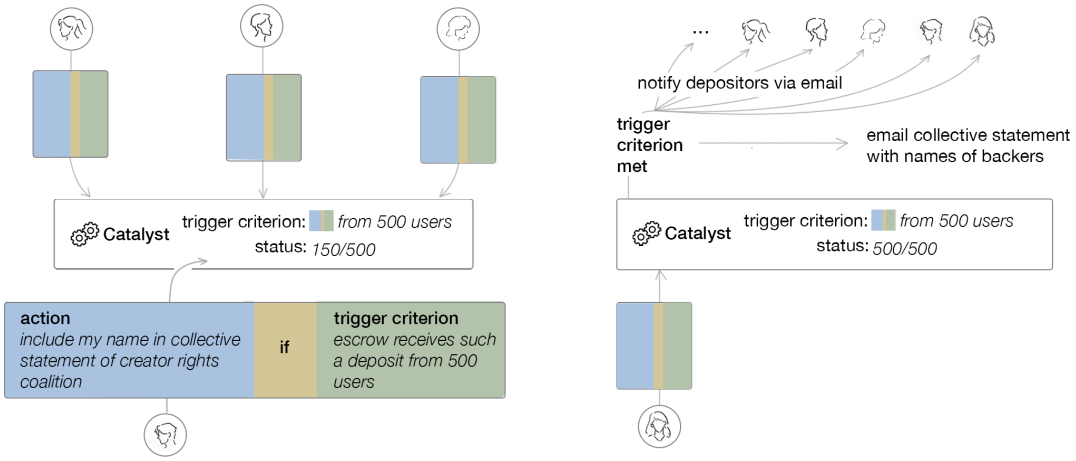

Figure 5.1: Catalyst instantiates an action escrow to lower first-mover disadvantages in collective action efforts. Here for instance, creators add their name to a collective statement to protect their rights, with their names only revealed when 500 depositors commit, enabling unified action with reduced individual vulnerability.

revised, creating stantial public pressure while protecting individual creators from being singled out for retaliation

Discussion: Catalyst demonstrates how action escrows can overcome critical mass dilemmas: if critical mass exists, it ensures that individuals can act collectively without being held back by first-mover disadvantages [17]. This risk-reduction approach parallels mechanisms used in crowdfunding platforms like Kickstarter and GoFundMe, where supporters' money is held in escrow until either the funding threshold is met (releasing funds to project creators) or the campaign fails (returning funds to supporters).

5.2.2 Burst: Lowering First-Mover Disadvantages in Forwarding Content into Public Forums

Burst[1] is a micro-blogging social media platform where interaction is organized into different channels (from large public spaces to small private ones), but with an added feature: action escrows that overcome the first-mover disadvantage in forwarding content from group to group (e.g., a team-specific channel to #general). People who voice interesting opinions in small groups may want to keep them there, worried about being poorly received. For example, a researcher may share an incisive critique of the field only to their local colleagues, worried about whether it will be well-received or understood by the broader community. While the message is a legitimate and thoughtful consideration of an issue plaguing the broader field, the individual researcher is afraid to share it widely and be thought of as taking shots at peers.

[1] https://testflight.apple.com/join/tdiSYv1H

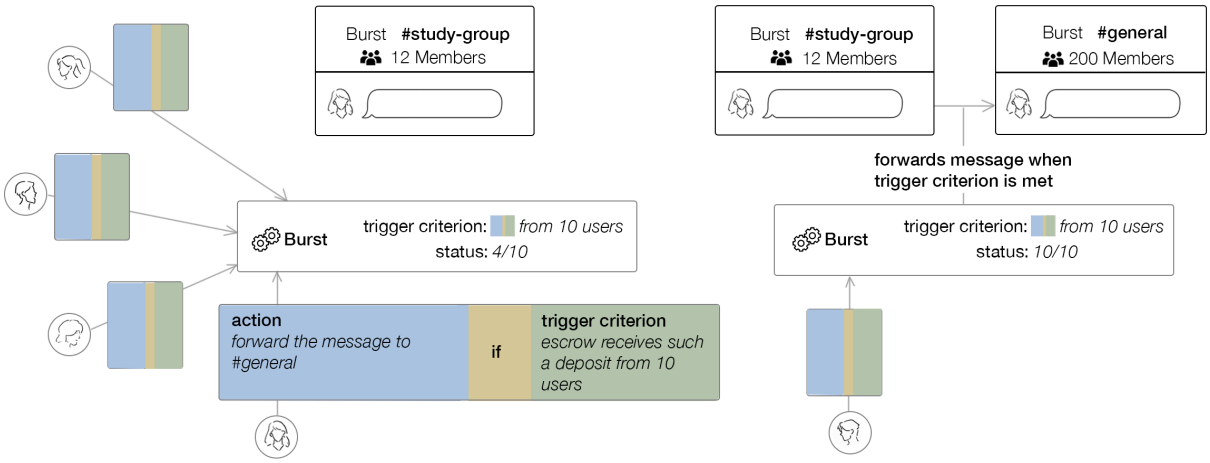

Figure 5.2: Burst instantiates an action escrow to lower first-mover disadvantages in forwarding content from private channels to public forums. Here, a student proposes forwarding a message from a small study group to the #general channel, requiring support from 10 group members. The system shows progress toward the threshold (left: 4/10 bursts), and once 10 members commit their support (right: 10/10), the message is automatically forwarded to the larger channel with indication of collective backing.

Rather than facing this sharing dilemma alone, Burst allows "forwarding together" by asking users to deposit their intention to forward the post into an escrow system. Posts are first shown to a small group of users trusted by the poster. The message is only shared to a new group when the activation threshold is crossed (*trigger criterion*): when enough people from this trusted group agree to burst it to a new group, thereby depositing their intention to support forwarding that message. The original author implicitly makes the first deposit by posting, indicating their desire to share their message with the broader audience, conditioned on further approval. The platform requires a specific number of deposits (bursts) before the post and the number of backers are shared to the selected audience. As these bursts accumulate, Burst reveals the current count of deposits, allowing participants to see the progress toward the activation threshold (*interim disclosures*) for forwarding it to the public. When the activation threshold is met and the message is "burst" into the new channel, it arrives with backing—each burst represents someone publicly vouching for the message's importance. This collective backing significantly reduces the vulnerability of the original author, distributing the risk that would otherwise fall solely on them. It is also a guarantee that members of the community already receive the content favorably; each burster is simultaneously a representative of the audience it is going to reach. Burst's approach to action escrows is exemplified in Figure 5.2, which presents the following usage scenario.

Usage Scenario Alex is a conscientious student in an advanced database course. After struggling with an ambiguous assignment rubric, Alex drafts a polite message requesting clarification on specific grading criteria that have confused many classmates. Though the message is respectful and constructive, Alex hesitates to post it directly in the course's #general channel where the professor would see it, fearing it might seem confrontational coming from just one student.

Instead, Alex shares the message in a private study group channel where fifteen other students have expressed similar concerns. Using Burst, Alex proposes forwarding the message to #general, where the platform has a pre-set activation threshold of ten supporters for course-related content. The study group members review the carefully worded request and begin to deposit their ”bursts” of support. When the tenth student adds their burst support, the message is automatically forwarded to the #general channel, appearing with an indicator showing it has backing from nine classmates. The professor responds appreciatively to the collectively endorsed feedback, clarifying the rubric points and thanking the students for their constructive approach. The clarification helps the entire class understand expectations better, and Alex's reputation remains intact; feedback from her peers assured that she wasn't forwarding a poorly-thought-out whinge but a reasoned and appropriate critique.

Discussion: While Burst can help overcome individual reluctance to post, the net impact of this system attacks dilemmas like the silent majority effect or the bystander effect, where views are never expressed publicly because individuals are afraid to express them without existing signs of approval within the community. The Burst architecture allows users to solicit some feedback from a friendly audience to determine if something is appropriate to post publicly, rather than relying the signal of what has already been posted, which may be subject to the same self-censoring inclination that user is experiencing. While Nooks enables the formation of private spaces around shared interests, Burst facilitates the transition of ideas from these private spaces back to public forums, and Catalyst empowers communities to act collectively. Together, these mechanisms demonstrate how action escrows can lower first-mover disadvantages throughout the entire process of enacting change. Now, we turn to a more familiar example of action escrows to highlight their broad applicability across different domains of social interaction.

5.2.3 Secret Crush: Lowering First-Mover Disadvantages to Admitting Romantic Interest

Secret Crush[2] is a Facebook Dating feature that creates escrows that overcome the first-mover disadvantage in admitting romantic interest to friends: even if two people like each other they may each be reluctant to confess first. Secret Crush allows individuals to deposit their romantic interest in a friend into escrow, which is only revealed if the friend also expresses romantic interest in them (*trigger criterion*). The individual making the deposit can select up to nine friends they are interested in. The application notifies the selected friend that someone has a romantic interest in them (*interim disclosure*) without revealing the identity of the depositor. If the selected friend also adds the original depositor to their own Secret Crush list, both users receive a notification that they have matched. Secret Crush uses *reciprocal deposits* as its trigger criterion and in its interim disclosures reveals *only receipt of the first deposit*.

Usage Scenario Maya and Eli have orbited each other for months in their friend group, sharing quiet conversations and genuine laughter during weekend gatherings, but neither knowing if

[2]https://www.facebook.com/help/347243103977573

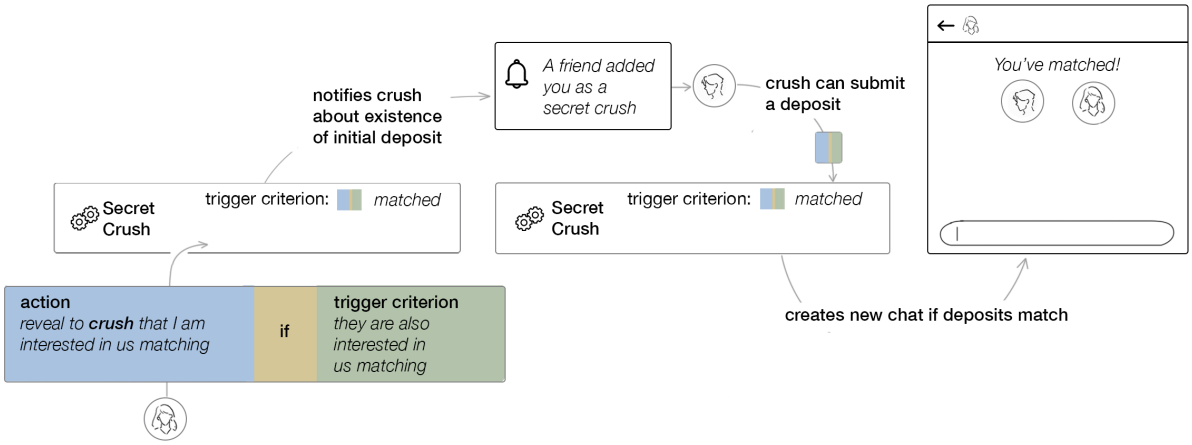

Figure 5.3: Secret Crush instantiates an action escrow to lower first-mover disadvantages in admitting romantic interest, similar to familiar dating app matching algorithms but specifically for existing Facebook friends. It uses reciprocal interest as its trigger criterion. Here, when a user adds someone to their Secret Crush list, the other person is notified they have a secret admirer without revealing who. Only if both users add each other to their lists do they "match," creating a chat where both can communicate with the knowledge of mutual interest, protecting either from rejection if interest isn't reciprocated

their feelings went both ways. After discovering Secret Crush, Maya adds Eli's name to her list. Eli receives a notification that someone has added him to their Secret Crush list, sparking his curiosity but giving no hint about who it might be. A few days later, while remembering their conversation at last weekend's barbecue, Eli adds Maya to his own list. Their phones simultaneously buzz with matching notifications, and they exchange texts to eventually meet at their usual coffee spot—where they finally talk about their mutual feelings that they'd been too insecure to voice.

Notes: Secret Crush illustrates that the mechanism powering dating apps (including Tinder, Bumble) is, also, an action escrow. In Tinder's case, where users can only contact each other after matching, escrows don't just reduce first-mover disadvantages, they also enhance safety by prohibiting unintermediated contact. (Secret Crush can't *forbid* direct contact since it operates among Facebook friends who already have messaging access to each other).

5.3 Design Space of Action Escrows

By presenting four distinct contexts—Nooks, Catalyst, Burst, Secret Crush—where action escrows can mitigate first-mover disadvantages, we have aimed to provide concrete examples of action escrows, while inviting you to consider additional domains where the design pattern can be beneficially applied. We have also shown the potential choices that can be made in configuring the two key parameters: the trigger criterion and the interim disclosures. In Figure 6.2, we summarize a fuller design space, including auxiliary parameters, that can merit explicit consid-

Design parameter	*Possible choices*		
Trigger criterion (On what is the deposit's release conditioned?)	activation thresholds *(Catalyst, Burst)*	reciprocal deposits *(Nooks, Secret Crush)*	
Interim Disclosures (What, if anything, is disclosed before trigger criteria is met?)	none	only receipt of first deposit *(Nooks, Secret Crush)*	progress towards trigger *(Catalyst, Burst)*
Trigger evaluation algorithm (How are follow-on deposits matched to initial deposits and counted towards trigger?)	exact (requires follow-on deposits to be categorical) *(Catalyst, Nooks, Burst, Secret Crush)*	fuzzy	
Forbidden acceleration (Is the user forbidden from acting alone before trigger is met?)	no *(Catalyst, Nooks, Secret Crush)*	yes *(forwarding into some channels in Burst)*	
Withdrawal (Can the user back out retroactively?)	no *(Nooks)*	yes *(Catalyst, Burst, Secret Crush)*	
Expiration (Do deposits expire if not released in a certain time?)	no *(Burst, Secret Crush)*	yes *(Nooks, Catalyst)*	

Figure 5.4: The design space of action escrows.

eration when implementing action escrows. Here we discuss these auxiliary parameters and the potential choices for each.

5.3.1 Trigger Evaluation Algorithm

Trigger evaluation algorithm refers to the method by which follow-on deposits are matched to initial deposits and counted towards meeting the trigger criterion. This can be implemented in two distinct ways: *exact* or *fuzzy* matching. Exact matching requires follow-on deposits to be categorical responses from predetermined options, such as "interested/not for me" in Nooks or "join up/not" in Catalyst. This is because categorical responses allow us to exactly link follow-on deposits to initial deposits and count them accurately toward the trigger threshold. With exact matching, we know precisely which topic the sequent user is expressing interest in or which specific collective action effort they are committing to join. In contrast, fuzzy matching accommodates open-ended deposits and allows for imprecise inputs. Consider an alternate version of Nooks that might match someone who wrote they're "looking for creativity workshops" with someone who specified "interested in collaborative brainstorming sessions." While we are not aware of systems that have explored fuzzy matching for action escrows, we regard this exploration as ripe for future work, enabled by both established approximate string matching algorithms [78] and recent advances in large language models [55, 101]. For example, in public counterspeech applications [77], fuzzy matching could trigger the release of drafted responses only when a threshold is met—users who wrote "The study actually found vaccination reduces infection rates by 70%" and "Research shows vaccines cut transmission by more than

two-thirds" would have their comments publicly posted only after five similar corrections were escrowed, despite their different specific wording.

5.3.2 Forbidden Acceleration

Is the user required to wait for the trigger criterion to be met, or can they accelerate action independently? Offering this acceleration option is particularly valuable when a user's commitment level can change, either due to urgency, new information, shifting priorities, or growing confidence—situations where they may become willing to accept the first-mover disadvantage. Catalyst, Nooks, and Secret Crush don't forbid acceleration: users can always express public commitment, message public forums directly, or contact the friend they're crushing on through Facebook if they choose not to wait. However, some Burst communities require approval (in the form of bursts) before posts are allowed in to maintain quality standards and norms, and some systems like Tinder explicitly prevent users from making independent contact for safety reasons, requiring them to wait until the matching condition is satisfied.

5.3.3 Withdrawal

Can a user back out after having made a deposit? Allowing withdrawal provides greater control to users who may change their minds, enabling them to retract their commitment without consequence. However, this flexibility comes with drawbacks: participants may question whether others will remain committed when the trigger condition is met. As with any trade-off, the "right" choice depends on the specific context, the stakes involved, and how much certainty is required for the escrow system to effectively serve its purpose.

5.3.4 Expiration

Should deposits expire if they remain unreleased after a certain period? Implementing expiration dates for action escrows creates a natural time boundary for commitment, preventing indefinite limbo states and allowing users to move on when sufficient interest fails to materialize. This temporal constraint can increase urgency and encourage more decisive participation, while also keeping the system free of stale, abandoned deposits. However, setting appropriate timeframes requires balancing enough time for critical mass to form against the risk of waning user interest and relevance. Designers need to consider whether the specific action context benefits from time pressure or whether some commitments should remain valid indefinitely until matched. Among our design cases, Nooks and Catalyst implement expiration periods—Nooks uses a fixed 24-hour window while Catalyst allows the initial depositor to define the expiration timeframe.

5.4 Conclusion

This chapter has attempted to expand our understanding of action escrows beyond the specific case of Nooks to reveal a rich design space for addressing first-mover disadvantages. Through our analysis of Catalyst, Burst, and Secret Crush, we have demonstrated that action escrows can

be effectively configured to support collective action efforts, content forwarding, and romantic disclosure—each addressing different first-mover disadvantages.

While all action escrows share the core mechanism of conditional commitment, choices in trigger criteria, interim disclosures, and auxiliary considerations like expiration periods and withdrawal options all shape how these systems function in practice.

By establishing this design space, we have attempted to provide designers with a principled framework for implementing action escrows tailored to their specific contexts. We have hoped to show the broad applicability beyond the cases presented here, and open up opportunities for future work to explore new matching algorithms, and novel applications across different domains of social interaction.

Chapter 6

Empathosphere: Action Escrows for Translucence Into Privately-Held Opinions

So far in this dissertation, I have shown how action escrows can address first-mover disadvantages by allowing people to take coordinated action. In this chapter, I show how action escrows can also help communities accurately assess private attitude distributions without triggering any actions.

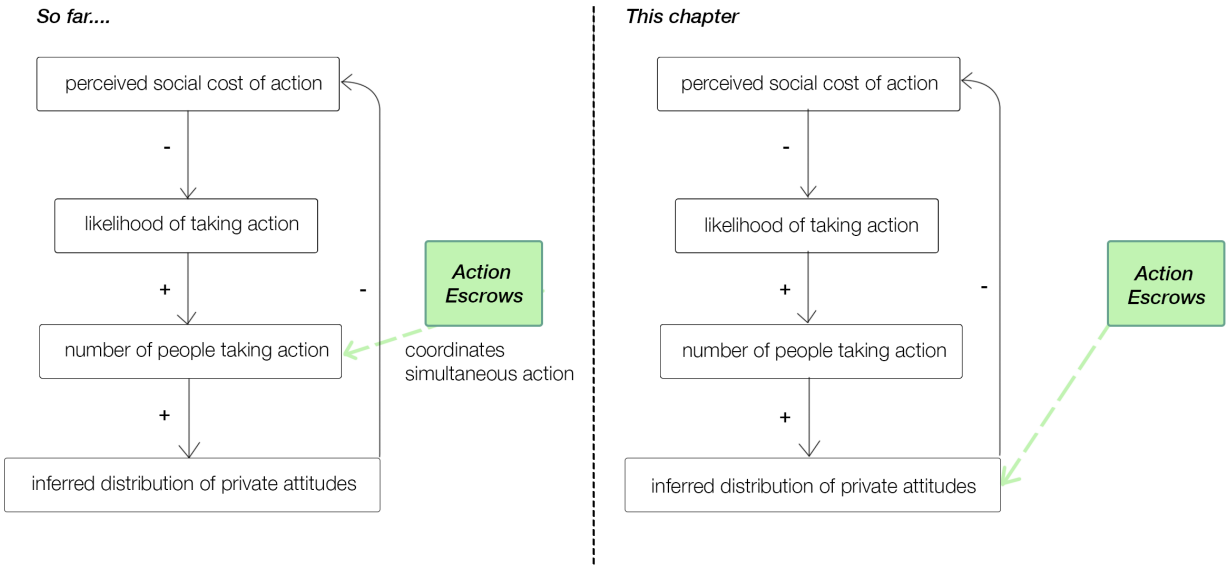

Figure 6.1: While traditional action escrows (left) coordinate simultaneous action, this chapter explores using action escrows purely for attitude assessment (right), revealing the distribution of community perspectives without enforcing any public actions.

The motivation for this stems from a limitation in how we have applied action escrows so far: when they trigger and execute deposited actions—like posting a diverging position—people's positions become public and binding. This creates psychological pressure to defend those positions, preventing people from sharing evolving positions and preventing genuine reconsideration once

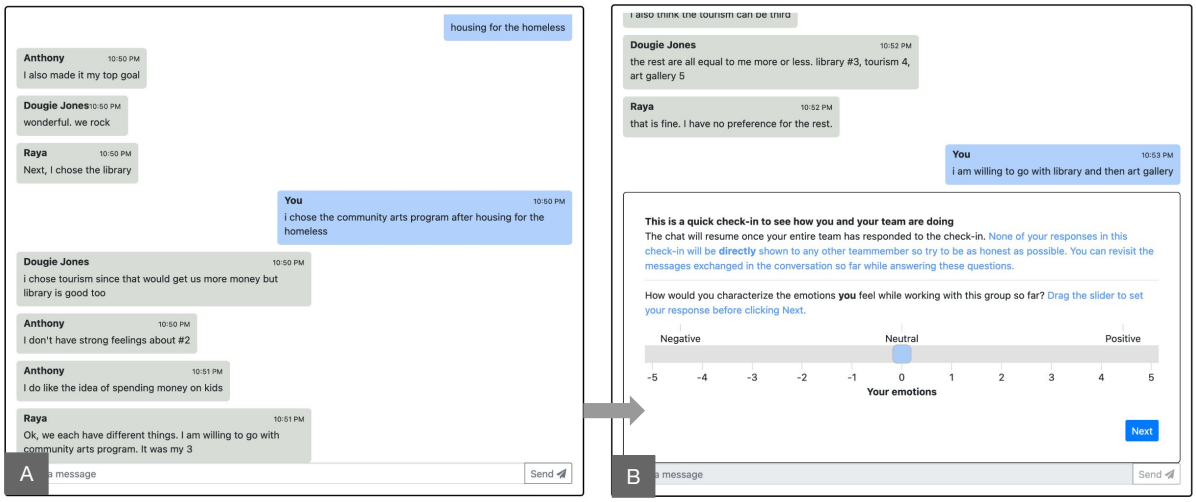

Figure 6.2: When triggered, *Empathosphere* appears as a widget in the chat interface and disables the chat while groups follow the action escrow workflow. The system prompts members to deposit their own feelings about the ongoing conversation and their guesses of how others might be feeling into escrows that never trigger. Instead, the escrow provides interim disclosures showing aggregated group sentiment and feedback on individual perceptiveness. This approach increases groups' desire to continue working together and encourages more open communication within teams.

positions have been shared. By revealing attitude distributions without executing actions, we preserve people's ability to evolve their thinking as new information emerges while still providing valuable community insights. This can be especially useful for rapidly changing situations, such as an workplace community navigating a potential remote work policy where employees' views might shift as they learn more about implementation details, or civic engagement forums where participants' positions on local issues evolve as they encounter new information and engage in online deliberation, but need space to reconsider without being locked into early public stances that become part of their digital identity.

The key insight of this chapter is leveraging action escrows for their *interim disclosures* rather than action execution. This builds on observations from Nooks, where participants gained ambient awareness of collective interests simply by observing proposed topics before any triggers were met.

Empathosphere [45] embeds action escrows into group chats to help groups assess the distribution of private attitudes. When triggered, it prompts individuals to deposit their private feelings about an ongoing conversation into escrows that are explicitly designed never to trigger—meaning individual responses remain private forever. Instead, the system reveals only aggregated sentiment across the group once enough deposits are received (the *interim disclosure*), providing translucence into collective attitudes without exposing anyone's specific position. To further improve social perception accuracy, Empathosphere also asks participants to guess how others in the group are feeling and provides personalized feedback on their accuracy. Figure 6.2 shows *Empathosphere*'s interface during an active group chat session.

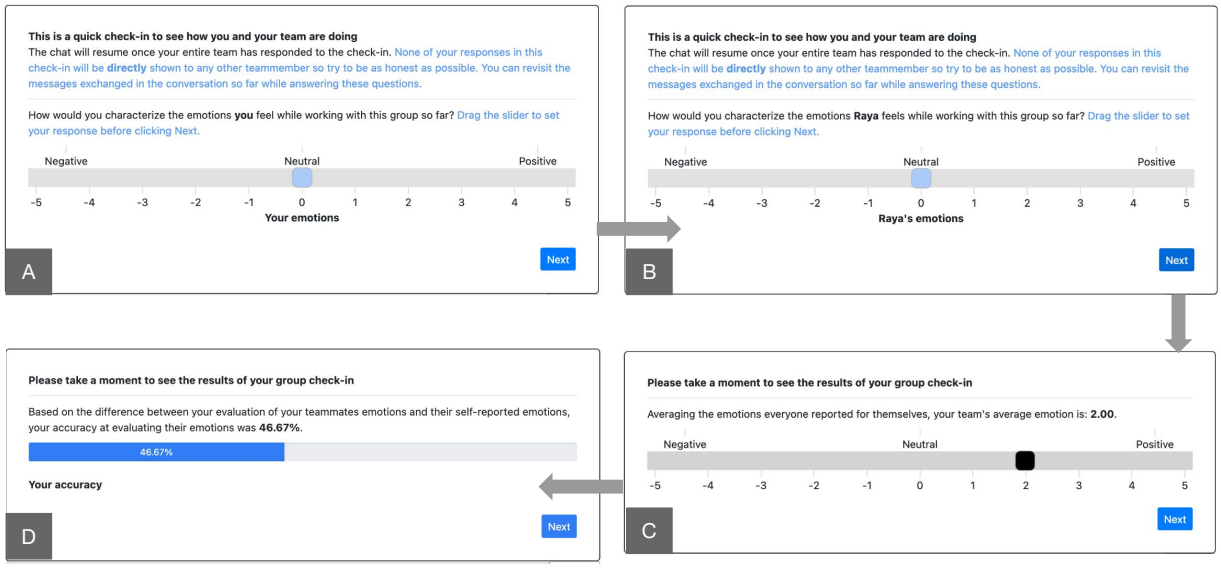

Figure 6.3: *Empathosphere's* action escrow workflow. Members of a group A) share their own socio-emotional states, B) think about the socio-emotional states of their teammates and finally, *Empathosphere* presents them with C) a measure of the group's cumulative socio-emotional state and D) a measure of how accurate they were at evaluating their teammates' true socio-emotional states.

In the rest of this chapter, we describe the design and implementation of *Empathosphere*, and describe a controlled study that evaluates its effectiveness.

6.1 Empathosphere

Empathosphere is a system that brings action escrows to the group chat. To explore different possible designs within a completely configurable chat environment, we built a custom chat-based collaboration environment (Figure 6.2). This was a client-server web application built using Meteor.js. When triggered, *Empathosphere* appears as a widget in the chat interface and disables the chat while groups follow the action escrow workflow (Figure 6.3).

6.1.1 Empathosphere's Action Escrow Workflow

First, *Empathosphere* privately elicits how team members actually feel about working with their group (Figure 6.3A) on a scale ranging from -5 to 5; -5 being the most negative and 5 the most positive. Next, it asks each member, in private, to guess how each of the other members in the team might be feeling on the same scale, to nudge them to direct their attention towards others in the team (Figure 6.3B). Finally, *Empathosphere* calculates the mean of responses from the first stage to present each participant with feedback about the aggregate group climate, i.e. how positive/negative the group as a whole is feeling, without revealing individual responses (Figure 6.3C). (While emotions i.e., how people feel, can be multi-faceted, we focus on a simple

positive/negative model in our intervention to ease reflection and evaluation for participants.) Forcing reflection on others' emotions and presenting the average group climate are the first way *Empathosphere* draws members' attention to the possibility that some team members might not be having a positive team experience. To further sharpen collective perceptions, *Empathosphere* also presents every member with feedback on how accurate they were at guessing others' emotions (Figure 6.3D) in their responses in the second stage above. Specifically, we use a representation of mean absolute error that facilitates easier interpretation. For a team member i, the accuracy $\mathcal{A}_i$ is calculated as below, where G_{ji} is i's guess about j's emotional state, and S_j is j's self-reported emotional state:

$$\mathcal{A}_i = max(0, 1 - \frac{1}{5}\sum_{i,i\neq j}^{n} \frac{|G_{ji} - S_j|}{n-1})$$

In an earlier version of *Empathosphere*, we presented the raw mean absolute error, however, participants in our pilot studies found it hard to interpret this number. In response, we chose to transform the mean absolute error to an accuracy metric such that the accuracy is at 100% if a member guessed all other members' emotions perfectly, and at 0% (in expectation) for random guesses. This measure can be negative (when disagreements exceed what can be expected by chance) but this is not useful for our reflective purposes, so we show participants 0% accuracy at minimum. $\mathcal{A}_i$ is displayed to team member i as a percentage for easier interpretation. By controlling for agreements by random chance, this measure emphasizes discrepancies between an individual's evaluation of others' emotions and their actual emotions, with the goal of drawing attention to the valence and arousal levels of others' emotions.

6.2 Evaluation Study

To investigate the impacts of *Empathosphere* on teams' expression and handling of conflicting opinions and diverse perspectives, we conducted a between-subjects study with teams of crowd workers on Amazon Mechanical Turk. This study was approved by our university's Institutional Review Board.

6.2.1 Participants

Participants were recruited from Amazon Mechanical Turk (AMT). We restricted participation to workers located in the United States who had completed at least 100 tasks and had an approval rate greater than 95%. All participants were compensated at a rate of $15/hr with a bonus that was adjusted based on the time they spent on the experiment. Prior work on team viability has studied teams ranging in size from four to eight [111] members so we aimed to form teams of similar size in our study. We wanted to ensure that teams studied had at least four members so we split participants into groups of six to account for subsequent reduction in group size due to disconnections and dropouts. We only analyzed data from teams that had at least four members till the end of the task. If at any point the number of team members dropped below four, the task was terminated and participants were compensated for their time. In total, 38 participants were

unable to complete the full study either because they dropped out themselves or because their teammates dropped out and the team strength dropped below four.

6.2.2 Experimental Setup

We extended the Meteor.js collaboration application that houses *Empathosphere*, with the TurkServer [67] framework. This allowed us to recruit participants on AMT and draw them into a lobby where they waited for other participants to arrive. Once there were a sufficient number of participants, the application automatically assigned teams to the different experimental conditions and created multiple parallel worlds where different teams worked simultaneously. Before they were added to chatrooms with their teams, participants were asked to set a pseudonym for themselves to preserve privacy, allowing them to control what aspects of their identity they wanted to reveal. Teams were randomly assigned to either the *Empathosphere* or control condition:

Empathosphere Condition: For teams in this condition, *Empathosphere* was triggered at the midpoint of the task and they were prompted to carry out the perspective-taking exercise.

Control Condition: Teams in the control condition were asked to take a two-minute pause and reflect on their teamwork experience individually. The specific prompt we used was: *"The experiment will proceed after a brief two-minute pause. Use this time to revisit the messages exchanged in the conversation so far and reflect on how the experience of working with this group has been."*

We chose this silent, individual reflection activity to design a conservative control condition. Prior work suggests that a break in work can have some benefits for team members' ability to navigate conflict, irrespective of the activity they engage in during the break [19, 59, 104]. This research suggests that time delays can be effective at de-escalating negative emotions and encourage mindful decision making and so our control condition consists of a break with a silent individual activity. This control thus seeks to isolate the effects of processes engaged by specific activities in a counter-normative space, rather than outcomes that could be attributed to the presence of a break.

Through our pilots, we observed that the perspective-taking exercise took around two minutes on average and so, the control group's break duration was set to this time; so that the average time spent on the task would be similar for both the control and the *Empathosphere* group. Preserving the task time between conditions seeks to further eliminate differences in familiarity between team members as a potential confound.

6.2.3 Task

Similar to previous studies of group work [110, 111] and group decision making [32], we asked teams in both conditions to work on a negotiation task. The task required teams to decide how to allocate funds across competing project proposals. This task was an instantiation of a "cognitive conflict" task in the McGrath's Task Circumplex Model [71] and as such required teams to work interdependently and resolve conflicting opinions and perspectives to arrive at a solution.

We use the well-established "foundation task" [32, 108] to create our specific task: groups were asked to allocate $500,000 across five competing project proposals, each in need of $500,000.

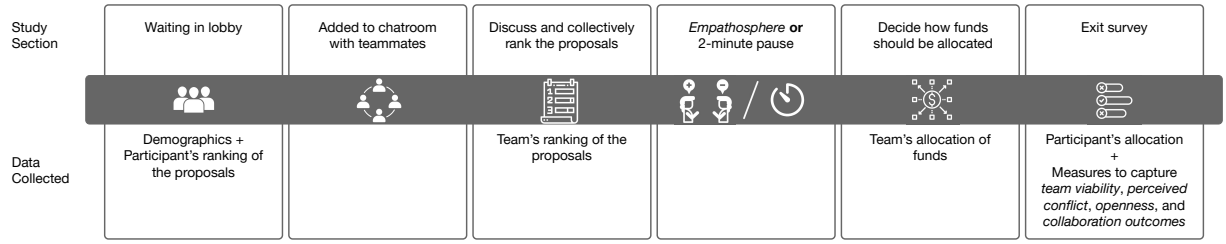

Figure 6.4: The study workflow and data collected in the different stages of the study.

The specific proposals were: 1) To establish a community arts program featuring art, music, and dance programs for children and adults 2) To create a tourist bureau to develop advertising and other methods of attracting tourism into the community 3) To purchase additional volumes for the community's library system 4) To establish an additional shelter for the homeless in the community 5) To purchase art for display in the community's art gallery. While the original "foundation task" [108] had a single phase where participants decided allocations, our experiment adapts this task to use two phases (discuss and decide) as below, to better measure the effects of perspective-taking.

6.2.4 Procedure

Figure 6.4 illustrates the study procedure. The workflow of the study was as follows:

Waiting in the Lobby

On joining the study, participants were added to a waiting room till there were a sufficient number of participants for the study. Before they were assigned to groups and the task began, each participant was presented the proposals and asked to rank them in terms of their relative merit. They were asked to evaluate the projects based on their own beliefs and values. During this time, we also collected demographic information of the participants.

Data collected: Demographic Information, each participant's ranking of the proposals.

Added to Chatroom With Teammates

As soon as there were a sufficient number of participants, they were split into teams of six participants each and were added to a chatroom with their team members. Participants were asked to pick a pseudonym before they began to interact through the chat-based interface. The task was split into two phases: *discuss* and *decide*.

Discuss and Collectively Rank Proposals

In the *discuss* phase, the teams were asked to weigh the pros and cons of each proposal and rank them in terms of their relative merit collaboratively. Participants were asked to advocate for projects that aligned with their personal values. At the end of this phase, we asked teams if they

were able to agree on a ranking of the proposals and if so, we ask them to submit their collective ranking. Each team had nine minutes for this phase.

Data collected: Each team's ranking of the proposals, chatlogs.

Empathosphere or 2-Minute Reflective Pause

Following this, teams in the *Empathosphere* condition carried out the perspective-taking exercise while teams in the control condition were asked to pause for two minutes and reflect on how their experience of working in the team had been thus far.

Decide How Funds Should Be Allocated

After this, the teams moved on to the *decide* phase. Teams were informed that they had to allocate $500,000 across the different projects. Each project was in need of $500,000, and the more money a project got, the more likely it was to succeed. The rankings decided in the *discuss* phase were meant to guide the *decide* phase, but teams were told that their final allocations need not reflect their rankings from the *discuss* phase. At the end of the task, we asked each team to enter their decision on how the $500,000 should be allocated following which the participants were directed to an exit survey. Each team had nine minutes for this phase.

Data collected: Each team's allocation of funds, chatlogs.

Exit Survey

The exit survey included questions corresponding to the measures described in 6.2.5. Finally, the exit survey also asked each participant how they would have allocated the $500,000 themselves.

Data collected: Each participant's allocation of the funds, responses to likert-type and open-ended questions.

6.2.5 Measures

To capture different aspects of teams, their collaboration experiences, and their collaboration outcomes, we included measures for (1) baseline disagreement, (2) team viability, (3) perceived conflict, (4) openness, (5) collaboration outcomes, and (6) conversational behavior. Forms of data captured included proposal rankings provided by participants while in the lobby, chatlogs, teams' allocation of funds, and open-ended and Likert-based questions in the exit survey. Items for quantitative measures included in the exit-survey were scored on a 5-point Likert scale.

Baseline Disagreement

We computed the amount of disagreement between members in a team using each participant's initial ranking of proposals as entered by them while in the lobby. For every team, we computed the Spearman footrule distance [97] between the rank vectors of team members in a pairwise fashion. For every pair, the Spearman footrule distance provides a proxy for disagreement between the two members. Averaging this pairwise disagreement across all possible pairs in a team, we obtained a measure of team-level disagreement.

Team Viability

To measure participants' desire to continue collaborating with their teammates, we measured team viability using a three-item scale ($\alpha = 0.76$): *"Most of the members of this team would welcome the opportunity to work as a group again in the future" "As a team this work group shows signs of falling apart." "The members of this team could work together for a long time."* The items were selected from an item pool developed to measure viability [18] and have been used in other studies to measure team viability [110, 111]. To elicit honest responses, we told participants that we might use their responses to decide whether to team them up with the same or different people if we ran subsequent experiments.

Measures to Capture Perceived Conflict

To understand difference in perceptions of conflict across conditions, we measured perceived conflict on task-related issues- *task conflict* as well as perceived interpersonal tension- *relationship conflict*.

- *Task conflict* was measured using a two-item scale [41] ($\alpha = 0.79$) where items were: *"there was a lot of conflict of ideas in our group"*, and *"my team had frequent disagreements relating to the task we were assigned."*
- *Relationship conflict* was also a two-item scale [41] ($\alpha = 0.86$). Items included: *"people in my team often got angry while working together."*, and *"there was a lot of relationship tension in my group."*

Measures to Capture Openness

Openness is characterized as frank communication of issues and feelings about both task-related and personal matters [58]. To further understand differences in openness in teams across the two conditions, we measure team members' desire to give each other feedback and their receptiveness to feedback. We also include an open-ended question to probe for openness.

Quantitative measures:

- To measure participants' *willingness to give feedback*, we ask them if they would be willing to give feedback to other members of this group on their teamwork practices. To elicit more honest responses, we added that giving feedback was optional, and that if they were willing, we might follow up later to collect their feedback.
- To measure participants' *willingness to receive feedback*, we ask if they would be willing to receive feedback from other members of their team on their teamwork practices.

Open-ended question:

- We included an open-ended question asking participants if they would characterize the conversation in their group as open or guarded and asked them to explain their characterization.

Measures to Capture Collaboration Outcomes

As there is no clear performance measure for cognitive conflict tasks [110], we measure collaborative outcomes in terms of *satisfaction with solution* and the degree of *compromise*.

- *satisfaction with solution* captures the extent to which individual team members are satisfied with the team's final allocation of funds (survey item- *"I am satisfied with my team's final solution"*).
- To measure *compromise* in a team, we measured the differences between participants' individual allocation of funds as provided by them in the exit survey, and the allocations that were decided by the team. For each team, we calculated the compromise measure as the mean divergence of individuals' allocations from the group's allocation. (This is calculated as the arithmetic mean of root-mean-square-of-differences between each members' allocation vector and the team's allocation vector.) This measure captures the extent to which the group's final decision aligns with individual members opinions[1]. A higher average divergence suggests that members did not agree with the final consensus of the group— the final decision did not involve a fair compromise.

Conversational Behavior

To capture changes in the conversational behavior induced by *Empathosphere*, we compared shifts in LIWC indicators across the two conditions. We also included two open-ended questions in the exit survey.

Quantitative measures:

- *Changes in LIWC indicators in the two conditions* To analyze differences in teams' chatlog in the two conditions, we use the popular linguistic dictionary Linguistic Inquiry and Word Count, known as LIWC [82]. LIWC contains words bucketed into 125 psychometric categories including categories such as 'first-person pronouns', 'second-person pronouns', 'positive emotion', 'negative emotion', 'informality' etc. This allows us to analyze a given text along these 125 dimensions. For every message, we compute normalized frequency counts for the LIWC categories, i.e. the number of times words from the category were present in the message divided by the total number of words in the message. To obtain psychometric measures that are independent of the length of the chatlog, we then average these normalized frequency counts across all messages in a chatlog. Therefore, the mean LIWC indicator for *positive emotion* for a chatlog measures *on average* how positive messages in the conversation were.

Open-ended question:

- We included an open-ended question to capture perceived changes in conversational behavior: "How did you engage with the group in the second stage?"

[1]We also tried other measures such as the mean absolute differences between individual and group allocations. We use our current measure because it is less susceptible to outlier inputs. Results are consistent with other measures we tried.

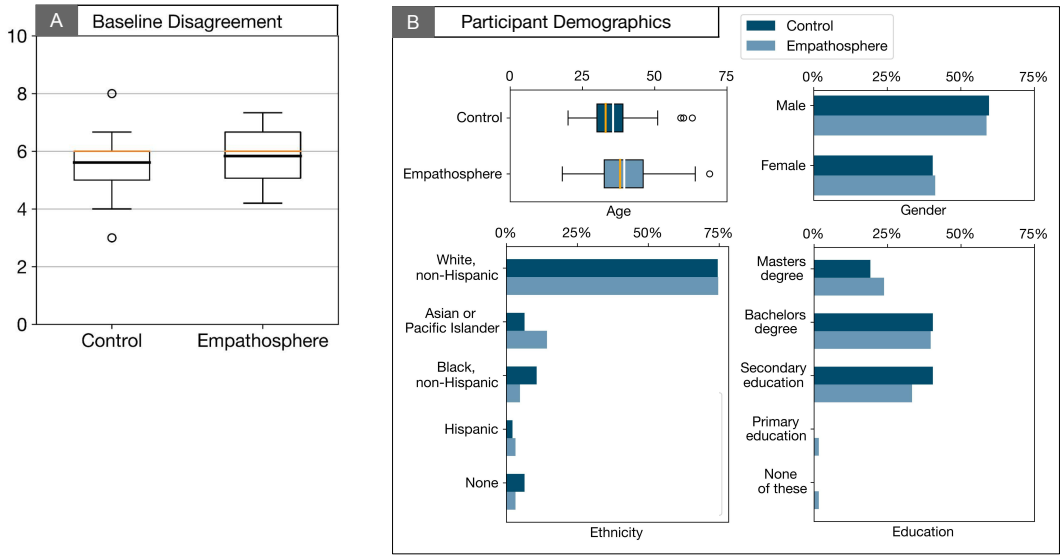

Figure 6.5: A) The baseline disagreement in teams in the two conditions where disagreement in a team is measured as the average of the Spearman footrule distance between the ranking of proposals across all possible pairs of members in that team. Median disagreement in both conditions is indicated by the orange marker and the mean in black. B) Demographic information of participants in our study. Median age is indicated by the orange marker and mean in white.

6.3 Findings

A total of $N = 110$ participants completed the experiment across 24 teams with 4-6 members each. We conducted our analyses on these 24 teams which included 11 teams in the control condition and 13 teams in the intervention condition. 59% of the participants were male and the average age of participants was 38 years ($\sigma = 10.4$). 74.5% participants identified as White, non-Hispanic, 10.9% identified as Asian or Pacific Islander, 7.3% as Black, and 2.7% as Hispanic. 21.8% reported having a masters degree, 40% a bachelors degree, and 36.4% a secondary education. Figure 6.5B shows a breakdown of demographics by experiment condition.

6.3.1 Baseline Disagreement

To check whether team composition varied significantly across the two conditions, we compared the baseline disagreement in teams in the two conditions (Figure 6.5A). We found no significant difference between disagreement in the control condition teams ($\mu = 5.6, \sigma = 1.27$) and intervention condition groups ($\mu = 5.83, \sigma = 1.01$) via a t-test: $t(109) = -0.45, p = 0.65$. Hedges' $g = 0.197$.

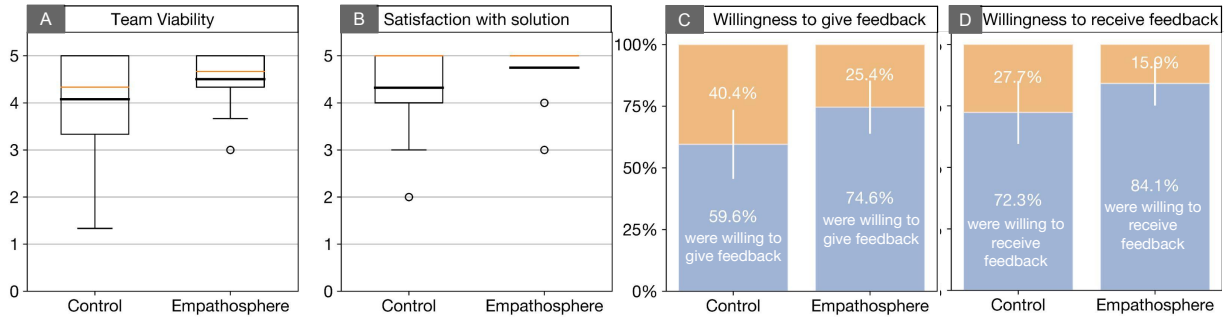

Figure 6.6: A) Compared to the control condition, *Empathosphere* led to significantly higher team viability. Median score is indicated by the orange marker and the mean by the black marker. B) Participants in the *Empathosphere* condition expressed significantly higher satisfaction with their teams' solution than participants in the control condition. Median score is indicated by the orange marker and the mean by the black marker. C) Participants in the *Empathosphere* condition were more likely to give their teammates feedback. The chart shows the proportion of participants that were willing and unwilling to give feedback to other team members, with 95% CI at the boundary. D) Participants in the *Empathosphere* condition were also more open to receiving feedback from their teammates. The chart shows the proportion of participants that were willing and unwilling to receive feedback from other team members, with 95% CI at the boundary.

6.3.2 Did *Empathosphere* Improve Inclusion in a Group?

Empathosphere Improved Team Viability.

Participants in the *Empathosphere* condition scored team viability to be higher ($\mu = 4.50, \sigma = 0.63$) than in the control condition ($\mu = 4.08, \sigma = 0.98$) (Figure 6.6A). We fit a mixed effects linear regression model with the experiment condition and team disagreement as fixed effects, their interaction, team grouping as a random effect, and team viability scores as the outcome variable. Results in Table 6.1. We observe a significant effect of both the condition ($\beta = 0.49$; 95% CI = 0.06,0.93; $p < 0.05$) and disagreement ($\beta = -0.35$; 95% CI $= -0.63, -0.07$; $p < 0.05$) on team viability, with no significant interaction effects. Teams with lower initial disagreement had higher viability, suggesting that teams with lower opinion diversity, and therefore inherently lower potential for conflict, tend to exhibit higher viability. Meanwhile, the effect of *Empathosphere* on team viability shows that prompting perspective-taking is a promising approach to improving viability of all teams, regardless of the degree of diversity in team members opinions.

Empathosphere Improved Participants' Satisfaction With Their Teams' Solutions

Participants in the *Empathosphere* condition reported higher satisfaction with their teams' solutions ($\mu = 4.74, \sigma = 0.50$) than participants in the control condition ($\mu = 4.31, \sigma = 0.94$) (Figure 6.6B). We fit a mixed effects linear regression model with the experiment condition and team disagreement as fixed effects, their interaction, team grouping as a random effect, and satisfaction with solution as the outcome variable. Results in Table 6.2. We observed a significant effect

	Outcome Variables							
	Team Viability				Satisfaction with solution			
Fixed Effects	Coeff.	SE	z	p	Coeff.	SE	z	p
(Intercept)	4.02***	0.17	23.43	<0.001	4.29***	0.14	30.48	<0.001
Condition	0.49*	0.23	2.13	0.033	0.45*	0.19	2.35	0.018
Disagreement	-0.35*	0.15	-2.34	0.019	-0.11	0.13	-0.89	0.372
ConditionXDisagreement	0.29	0.23	1.31	0.191	0.08	0.58	0.41	0.684
Random Effects	Var.	SE			Var.	SE		
Team	0.21	0.17			0.09	0.10		

Note: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 6.1: Results of mixed effect linear regression analyzing the impact of experiment condition and disagreement within the group on measures of *team viability* and *satisfaction with solution*. The condition had a significant effect on both measures with participants in the *Empathosphere* condition expressing higher viability and satisfaction with solution.

of the experiment condition on satisfaction with solution ($\beta = 0.45;\ 95\%$ CI $= 0.09$,$0.80;\ p < 0.05$) while there was no significant effect of disagreement. There was also no interaction effect.

We Did Not Observe Significant Differences in Compromise Across Teams in the Two Conditions

The difference between *compromise* in teams in the control condition ($\mu = 0.075, \sigma = 0.022$) and teams in the *Empathosphere* condition ($\mu = 0.070, \sigma = 0.036$) was not statistically significant, possibly due to small sample size (*compromise* was measured at the group level, and so had fewer observations than some of the individual level measures above).

6.3.3 How Did *Empathosphere* Impact Communication in the Group?

Teams in the *Empathosphere* Condition Used More Second-Person Pronouns and Exchanged More Informal Messages

We compare the changes in LIWC indicators from the *discuss* to *decide* phase in the *Empathosphere* and control conditions. We find that *Empathosphere* was followed by an increase in use of second-person pronouns (you, you've y'all, u). Teams used 89% more second-person pronouns in the *decide* phase, after *Empathosphere* ($p < 0.05$) but there was no significant difference in the usage of second person pronouns between the *discuss* and *decide* phase in the control condition. This indicates that *Empathosphere* potentially shifted the conversation to be more other-focused, engaging with what other team members are saying and drawing them into the conversation.

The intervention was also followed by an increase in informality (okay, yaas) and netspeak. Usage of informal words was 27% higher in the *decide* phase than the *discuss* phase in the intervention condition ($p < 0.05$). Similarly, usage of netspeak was 281% higher in the *decide* phase than the *discuss* phase for the intervention condition ($p < 0.001$). Meanwhile, in the control condition, the use of informal language decreased in the *decide* phase compared to the *discuss* phase by 24%, however, this difference was only marginally significant ($p = 0.09$).

Taken together, this also suggests that *Empathosphere* has the potential to improve social bonds in teams [115].

Empathosphere Helped Foster Higher Comfort Levels and Led to Team Members Respectfully Engaging With Each Other

We analyzed participants' responses to the open-ended question asking how they engaged with the group in the second stage. Participants in the intervention condition noted higher comfort levels in their team: 'It felt like coming back to a group of coworkers that I know well." (P91) They mentioned being able to voice their opinions with their team: "People threw out ideas while others were more intent on keeping the group focused on coming up with an actual answer to the task. I chimed in when I wanted something addressed or wanted to broach a specific idea that mattered to me" (P85), "I gave my thoughts and everyone listened and gave me constructive feedback" (P107), and "I suggested an alternative allocation of funds at one point and the group reached an amicable decision taking in everyone's vote" (P67).

They also took efforts to accommodate each others' perspectives: "I tried to take everyone's ideas and formulate them into a single plan. I took the numbers and tried to make them work, so as to make the plan easier to visualize. I think this helped the group to come to an agreement more quickly, because they could see the plan." (P81) They tried to make their teammates feel heard: "I knew Williams' personality and he would have a suggestion and would want to be heard." (P71) Several participants mentioned how they attempted to strike a balance between voicing their perspectives and listening to others' opinions: "I tried to take in their perspectives but wanted to make sure they understood me too" (P94), "I tried to summarize everyone's ideas as well as contribute my own" (P96), and "I made suggestions but also was open to what they thought as well." (P93)

Teams in the Control Condition Had More Polarized Experiences, With Some Reporting either Too Little or Too Much Conflict

Analyzing responses to the open-ended question by participants in the control condition, we found that participants hinted at loafing: "I engaged with caution, trying to let some of the other members bounce ideas off of one another, but no one was really into it." (P11) Some tried to avoid conflict altogether: "I gave my proposal, but it seemed like the group wanted to finish quickly so I didn't push my ideas that much." (P36) On the other hand, some participants in the control condition noted how their opinions were disregarded: "I expressed my thoughts and ideas about how to distribute money but was getting into some arguments about the merits of some programs versus others. I was getting frustrated because it felt like my group was ignoring my suggestions" (P24). Such dismissal of ideas also triggered negative emotions: "I tried to keep the group focused. However, one person was not respectful of my ideas and made snide remarks about me being insecure. That definitely hampered our progress." (P25)

	Outcome Variables							
	Willingness to give feedback				Willingness to receive feedback			
Fixed Effects	Coeff.	SE	z	p	Coeff.	SE	z	p
(Intercept)	0.27	0.31	0.90	0.366	0.93*	0.38	2.45	0.014
Condition	0.78	0.43	1.83	0.067	0.95	0.56	1.68	0.092
Disagreement	-0.33	0.29	-1.15	0.252	-0.62	0.37	-1.65	0.098
Condition×Disagreement	0.55	0.44	1.26	0.209	0.15	0.58	0.26	0.793
Random Effects	Var.	SE			Var.	SE		
Team	0	0			0.13	0.37		

Note: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 6.2: Results of mixed effect logistic regression analyzing the impact of experiment condition and disagreement on participants' *willingness to give feedback* to teammates and their *willingness to receive feedback* from teammates. The condition had a marginally significant effect on both measures with participants in the *Empathosphere* condition expressing higher willingness to give and receive feedback.

Empathosphere Improved Willingness to Give and Receive Feedback

Since *willingness to give feedback* and *willingness to receive feedback* had binary responses, for each of them as outcome variables, we fit a mixed effects logistic regression model with the experiment condition and team disagreement as fixed effects, their interaction, team grouping as a random effect (Table 6.2). We found a marginally significant effect of condition on *willingness to give feedback* ($\beta = 0.78;\ 95\%$ CI $= -0.05$,$1.76;\ p = 0.067$). Of the participants in the *Empathosphere* condition, 74.6% were willing to give their teammates feedback while only 59.6% of the participants in the control condition were willing to do the same (Figure 6.6C). Similarly, we found a marginally significant effect of the experiment condition on *willingness to receive feedback* ($\beta = 0.95;\ 95\%$ CI $= -0.14$,$2.24;\ p = 0.092$). 84.1% of the participants in the *Empathosphere* condition were willing to receive feedback from their teammates while 72.3% of the participants in the control condition were willing to do the same (Figure 6.6D). We also saw a marginally significant effect of disagreement ($\beta = -0.62;\ 95\%$ CI $= -1.47$,$0.09;\ p = 0.098$), such that participants in teams with higher disagreement were less keen on receiving feedback from their teammates.

6.3.4 How Did *Empathosphere* Affect Perceptions in the Group?

Empathosphere Made Participants More Perceptive to Other Team Members' Behaviors

We compared participants' responses to the open-ended question on whether they thought the conversation in their group as open or guarded. Even though we did not explicitly probe for it, participants in the *Empathosphere* condition also made more specific observations about their teammates suggesting that they were perceptive to how their teammates were behaving. One participant noted: "I knew I should suggest amounts quickly because William would have a proposal and would be more particular, I think. He seems like a leader or someone who wants to be in charge and doesn't ask anyone else what they want probably. However he did compromise." (P71) Another participant mentioned: "Kate seemed to be the one that had the most ideas that

	Outcome Variables							
	Task Conflict				Relationship Conflict			
Fixed Effects	Coeff.	SE	z	p	Coeff.	SE	z	p
(Intercept)	2.19***	0.21	10.64	<0.001	1.55***	0.17	9.28	<0.001
Condition	-0.36	0.27	-1.28	0.199	-0.15	0.23	-0.66	0.512
Disagreement	0.31	0.18	1.73	0.084	0.11	0.15	0.77	0.438
ConditionXDisagreement	-0.04	0.27	-0.13	0.896	0.10	0.22	0.46	0.644
Random Effects	Var.	SE			Var.	SE		
Team	0.29	0.19			0.19	0.15		

Note: $^{*}p < 0.05$; $^{**}p < 0.01$; $^{***}p < 0.001$

Table 6.3: Results of mixed effect linear regression analyzing the impact of experiment condition and disagreement on *perceived task conflict*, and *perceived relationship conflict* showing the absence of a significant relationship between the condition and either variables.

differed from the group. The other 2 people seemed to be the most in line with me." (P96)

Participants in the *Empathosphere* Condition Felt That Conversations Were Less Guarded

Participants in the control condition mentioned how some team members chose to stay silent: "Some people had nothing to say at all while two others were very open." (P5) Some other participants mentioned how there was very little opposition or open disagreement: "It was not really as engaging as I hoped. I had to get the ball rolling and didn't really get any conflicting opinions" (P11), and "It did not appear that anyone wanted to "dominate" the conversation/debate and therefore potentially yielded quicker than they would in person or make real decisions." (P41)

In contrast to this, participants in the intervention group reported higher levels of open disagreement: "We never strongly attacked an idea and views were able to change. People were allowed to suggest their ideal solution and did not seem bothered by challenges." (P94), "everyone brought something to the table and it was a great group" (P108), "I felt like everyone could voice their opinions, and no one was shot down unfairly." (P106) One participant noted a distinct shift in their behavior after the *Empathosphere* exercise: "I was guarded on the first stage and wanted to recommend tourism as the first priority but when others said homeless I agreed. I knew I would have to speak up quickly and I was less guard[ed] on round two because they all seemed like nice people would [who] wouldn't be rude." (P71)

Despite This, *Empathosphere* Did Not Raise the Perceived Level of Conflict

We did not observe differences in perceived task or relationship conflict across the conditions. Following the same analysis strategy as before, we fit mixed effects linear regression models with perceived task conflict and perceived relationship conflict as the outcome variables and found no significant effects of condition, disagreement, and their interaction on either outcome variable (Table 6.3). There was a marginally significant effect of initial disagreement on the perceived task conflict ($\beta = 0.31;\ 95\%\ \text{CI} = -0.02{,}0.66;\ p = 0.08$) with higher initial disagreement leading to higher perceptions of task conflict, which is in line with expectations. However, the condition did not yield significant differences in perceived task or relationship conflict suggesting that

Empathosphere might not change perceptions of conflict or trigger negative emotional responses to conflict.

6.4 Conclusion

In this chapter, we've shown how action escrows present a novel opportunity to provide translucence [26, 29] into privately-held viewpoints that would otherwise remain entirely hidden from the community. Here, the escrow agent withholds *opinions* that it privately elicits from users, which users feel comfortable sharing precisely because their personal expressions remain protected from direct scrutiny.

These confidential contributions are only released in aggregate form once a sufficient quantity of opinions across the community has been collected, ensuring no opinion can be traced back to its contributor. This privacy-preserving mechanism can enable communities to discover the true distribution of perspectives among their members without exposing individuals to social risk. This could help dispel groupthink by revealing when consensus views are actually less universal than perceived. *Empathosphere* exemplifies this approach by collecting anonymous viewpoints that individuals in a group may be reluctant to express publicly, revealing collective sentiment that might otherwise remain obscured by self-censorship and fear of judgment.

Chapter 7

Discussion

So far, I have introduced the design pattern of action escrows and described the broad range of problems they can address: those with first-mover disadvantages. To show that action escrows are effective, I have introduced two systems, Nooks and Empathosphere, each of which advances action escrows further as an approach to addressing first-mover disadvantages. To inform future applications of the pattern, I have provided concrete cases of existing systems that apply the pattern, and have teased out an underlying design space. Throughout, I have also tried to reveal the relationships between previously disconnected problems (*silent majorities*, *groupthink*) and their technical remedies (Nooks, Empathosphere), exposing common roots in first-mover disadvantages. In this section, I first reflect on action escrows' limitations in achieving coordinated action. Then, I discuss potential risks of introducing action escrows in communities, while identifying design approaches to mitigate these risks. Finally, I broaden our focus beyond the action escrows described in this dissertation and outline opportunities for future implementations, and explore the gap between their theoretical utility and practical adoption.

7.1 Limitations of Action Escrows

Although we've shown the possibility for action escrows to catalyze coordinated action, they are not a panacea. In this section, we reflect on some of the limitations of action escrows.

First, the potential for social change through action escrows is fundamentally constrained by users' trust in the entity managing the action escrow—whether an individual designer or an organization. With Catalyst, creators joining the "Creator Rights Protection Coalition" must trust that the platform won't leak their identities to the company they're organizing against before reaching the 500-person threshold. If the Nooks application is managed by Tejus' employer—and they can access the underlying database—then he might be unwilling to propose topics that radically oppose management practices. In each case, the effectiveness of the action escrow depends on users believing that the system will faithfully execute its promised function without premature disclosure. Trust in the escrow manager becomes a prerequisite for the social coordination benefits these systems aim to provide.

Second, action escrows don't create motivation; they merely coordinate it. They function best when individuals are already motivated to act but hesitate solely due to first-mover disadvantages.

For action escrows to succeed, individual action must be highly likely once the participation threshold is met. Action escrows can in fact be counterproductive in situations where publicly visible action from a first mover is needed to generate motivation, as they deliberately conceal these initial contributions until the threshold is reached. Consider the case where many people are signing a birthday card for a colleague: seeing other signatures may produce the social pressure to write a more in-depth message or give a pondering signatory ideas on what to mention, while an escrowed version of the card-signing process would leave the less confident signatories to minimize social risk and write lowest-common-denominator messages of the "Happy Birthday! [Signature]" variety.

Third, action escrows fragment community activity. By design, action escrows lead to activity that is distributed across public community forums, private community subspaces (as with Nooks), and concealed in the escrows of the community. By fragmenting activity across these locations, action escrows can make it hard to keep track of both the locations and volume of activity in a community. This fragmentation can make it hard for newcomers to the community to catch up on the activity in a community, and to join in on existing efforts [90].

7.2 Risks of Antisocial Behavior Through Action Escrows and Suggested Mitigation

Action escrows can also introduce new risks of anti-social behavior in a community. We believe designers attempting to implement an action escrow mechanism can (and *should*) mitigate these risks through careful choices in how to implement the mechanism, and perhaps, even whether to implement the mechanism. Here we outline two key risks and the mitigation we envision for each.

First, action escrows can enable extreme ideologies to fly under the radar of community moderators and members by enabling filter bubbles. This could allow groups spreading discriminatory rhetoric, hate, or misinformation to organize discreetly. Consider cases like incels coordinating hate campaigns through applications like Nooks, shielded from community oversight due to the privacy-preserving nature of the system. As a potential mitigation, we suggest that implementing interim disclosures that reveal the topics proposed for discussion (but not the discussants) could at least help community moderators and members monitor the landscape of emerging filter bubbles without compromising individual privacy, allowing for appropriate intervention before harmful coordination reaches critical mass.

Another key risk is that action escrows can be weaponized by infiltrators who join solely to unmask and target participants in sensitive contexts. Malicious actors may join an escrow with the sole purpose of discovering the identities of other participants once the threshold is reached, particularly in vulnerability-sharing spaces within online communities. For example, a malicious member might join an escrow intended as a safe space for marginalized members and gather sensitive disclosures they could later use to harass participants. This vulnerability creates a significant trust problem—users cannot distinguish genuine allies from infiltrators until it's too late. At one level the "opt-in" nature of action escrows can itself mitigate this risk. Because action escrows require an explicit commitment of interest from participants, bad actors

would need to actively misrepresent their intentions rather than passively observing, creating both psychological and social accountability barriers to infiltration. In communities where offline reputations and relationships exist, this requirement for active deception serves as a meaningful deterrent, if individuals face real social consequences for discovered betrayals. As a second level of mitigation, we suggest providing users with controls to explicitly exclude certain individuals or audiences when creating escrow deposits. In the design cases, we described, users could block specific individuals when proposing nooks, and apriori prevent their message from bursting into certain channels. As a third level of mitigation, designers could implement progressive identity revelation (where participants' identities are disclosed gradually as trust builds) [96], pseudonymity options that persist even after threshold activation, or social signals [37] that help participants gauge the trustworthiness of other escrow members before full identity disclosure occurs.

Ultimately, as with any algorithmic intervention introduced in a community, we believe designers should work closely with community members to anticipate risks and determine whether those risks can be reasonably managed through the escrow's configuration, before deciding to deploy it.

7.3 Horizons on Action Escrows

7.3.1 Mixed-Initiative Action Escrows

Systems like Nooks and Empathosphere help challenge dominant norms in digital environments. When norms are unclear, such as whether a conversation or perspective is welcome, these systems encourage users to "test the waters" instead of staying silent. However, they are useful only if someone thinks to use them. Could future systems trigger interventions automatically by inferring when someone is likely feeling excluded? There is an opportunity to investigate how we can model individuals' likely internal states (e.g., beliefs) in conversation so that digital environments can intervene adaptively, predicting not just opportune moments to intervene but also the content of those interventions (e.g., a version of Empathosphere that targets specific contested points instead of overall emotional states).

7.3.2 More Expressive Deposits and Matching Algorithms to Handle Them

The action escrow systems described here rely on simple binary deposits and exact matching, where users must select from predetermined categorical options like "interested/not for me" or "join up/not." This approach ensures precise deposit matching but significantly limits the expressiveness of commitments that can be matched.

Future work could explore two complementary directions for more nuanced matching. First, semantic matching could accommodate open-ended user inputs while enabling meaningful deposit aggregation. Consider an enhanced version of Nooks where someone depositing "looking for creativity workshops" could be matched with another user interested in "collaborative brainstorming sessions"—both expressing similar underlying interests despite different phrasing. Recent advances in large language models and established approximate string matching algorithms

make such semantic matching increasingly feasible.

Second, more expressive deposit structures could better capture user preferences rather than simple binary commitments. Instead of just "I'm interested," users could specify conditional participation thresholds like "I'll participate if at least 15 others join" or "I'm willing to lead if 8 people commit but only participate if 20+ join." The system could then solve for optimal participation sets by finding the largest subset where everyone's minimum requirements are satisfied, essentially computing fixed points where each participant's threshold conditions are met by the actual group size. There is an opportunity for future work to explore more diverse models of user preference.

7.3.3 Beyond Action Escrows: Escrow Mechanisms for other HCI challenges

Throughout this dissertation, we have explored how *action* escrows address first-mover disadvantages by withholding the execution of socially risky actions until others make similar commitments, thereby meeting a predetermined trigger criterion. We now broaden our focus to demonstrate how the fundamental escrow concept—withholding something valuable and releasing it under specific conditions—can be adapted (and indeed has been adapted) to address a wider range of HCI challenges beyond first-mover disadvantages. These alternative escrow mechanisms differ fundamentally from action escrows in what they withhold (not necessarily actions). In this section, we present several illustrative examples of these alternative escrow mechanisms. Figure 7.1 presents a summary. Again, by reinterpreting existing systems through the lens of escrows, we hope to reveal how these technical solutions to different problems leverage a common operating principle. In each of the following sections, we identify a core HCI challenge and explain how escrow mechanisms can be formulated to address it.

Escrows for Reinforcing Community Participation Standards

Escrow mechanisms can also address the pervasive problems of lurking and social loafing in online communities, where many users consume content without contributing. Here, the escrow agent withholds *access* to community content (e.g. conversations, posts, and interactions from other members). Access remains escrowed until a specific release condition is met: the individual user explicitly commits to participating according to community norms.

The system grants access progressively to each user who makes this commitment. This creates a participation gate where viewing others' contributions requires a pledge to contribute oneself, establishing reciprocity as a foundational norm. Commit [83] exemplifies this approach by periodically withholding access to group discussions until users pledge to contribute meaningfully. In a controlled study, Commit more than doubled participation rates compared to simple nudges, helping communities overcome the imbalance between content consumers and content creators.

Escrow mechanisms to address HCI challenges and opportunities

What is the challenge or opportunity addressed?	Lowering first-mover disadvantages (action escrows)	Reinforcing community standards	Supporting safe interactions	Supporting data-driven collective action	...
What is withheld by escrow?	execution of socially risky action	access to a community	permission to establish contact between users	permission to forward a user's data donation	???
Released under the condition that	others commit to same action	user commits to norm	all parties independently initiate contact	data is tranformed to be non-identifying	???
Examples	Catalyst, Nooks, Burst, Facebook's "Secret Crush"	Commit, BeReal	Tinder	Gig2Gether	???

Figure 7.1: An overview of escrow mechanisms applied to address HCI challenges. This is not an exhaustive list; additional escrow applications beyond those explicitly documented here may exist or be potential directions for exploration.

Escrows for Supporting Safe Interactions

Escrows can also be employed to facilitate safe interactions online by explicitly establishing mutual consent prior to interactions. Platforms like Tinder exemplify this approach, where the messaging functionality remains locked until both parties express interest by "swiping right". Here, the escrow specifically withholds the *permission* to contact each other until mutual interest is confirmed, shielding users from unwanted advances. Only when both parties have independently indicated interest does the platform unlock the messaging feature. This conditional mechanism respects interpersonal boundaries while enabling connections wanted by all participants, providing a potential design approach for realizing affirmative consent online [38, 87, 88].

Escrows for Supporting Data-Driven Collective Action

Escrow mechanisms can also facilitate data-driven collective action by addressing privacy concerns related to personal data donation [27, 34, 35]. Here, the escrow agent withholds the *permission* to forward a user's data donation until the data is transformed to be non-identifying [113]. Gig2Gether [36] implements this approach by enabling gig workers across multiple platforms to contribute their work data, which is then aggregated to create collective insights. This aggregation mechanism—by converting individual, potentially vulnerable data points into a powerful collective resource–simultaneously protects worker privacy while shifting power dynamics away from platforms and toward the workers whose labor sustains them. Escrows can thus provide the technological means for mutual aid by helping build, shift, and employ power [23, 106].

7.4 If Escrows Are Broadly Applicable, Why Haven't We Seen More of Them?

This dissertation has attempted to show that escrows *are* actually prevalent in social computing systems. At least more so than we might initially recognize—they simply haven't been conceptualized as such. Part of our goal has been to provide the analytical framework needed to identify these mechanisms in existing systems, allowing us to see that escrows have already emerged organically in various contexts. From dating apps revealing mutual interest only when both parties express it, to crowdfunding campaigns conditioning financial commitments on reaching a target, the action escrow pattern exists in numerous domains. If we haven't *seen* escrow mechanisms, it may not be because of their absence, but rather our lack of unified terminology to recognize, analyze, and deliberately improve these coordination mechanisms. By making the concept explicit, we can now identify, refine, and intentionally implement these systems where they can provide significant social value. Beyond this conceptual invisibility, we suggest that two additional factors help explain why action escrows don't seem pervasive.

7.4.1 Moral Reactance to Intermediated Communication

Some readers will probably experience a visceral aversion to the Secret Crush example, yet might have felt no such aversion to Catalyst, Nooks, or Burst. This differential reaction illustrates the first challenge. Many people feel that using systems like Secret Crush represents an uncomfortable delegation of social courage to an algorithm[1]. This reaction stems partly from deeply embedded social norms that reward displays of confidence and vulnerability, and partly from concerns about technology inserting itself into intimate social processes [28, 63]. (For more on how mediated communication affects both how messages are perceived and how senders are judged, see [46].) This moral reactance to technological intermediation is one factor that prevents the uptake of escrows. Even if escrow mechanisms reduce risk and potentially increase positive outcomes, users may resist them because they feel like a form of emotional outsourcing that undermines agency or authenticity. We suggest that escrows are most likely to find adoption in domains where their coordination benefits clearly outweigh concerns about technological mediation, rather than in domains where direct human communication remains culturally valued.

7.4.2 Difficulty of Ensuring Just-Enough Complexity

For escrows to work in social computing systems, they must strike a delicate balance between being sophisticated enough to solve the problem and simple enough for users to understand. Designing mechanisms that are simultaneously effective and intuitive is hard. To illustrate the challenge, consider the second-price (Vickrey) auction: bidders submit sealed bids, the highest bidder wins, but pays only the second-highest bid amount. This elegant design *theoretically* solves a fundamental market problem by making it optimal for each bidder to simply state their honest valuation of the item—no strategic underbidding or overbidding required [105]. Despite its mathematical elegance, the mechanism's optimal strategy remains invisible to users without

[1] https://mashable.com/article/facebook-secret-crush-bad

specialized knowledge: there's nothing in the auction description itself that guides users toward truthful bidding or makes the benefits of honesty apparent [42]. In practice, studies consistently show that participants frequently overbid or underbid, failing to recognize or trust that revealing their true values is in their best interest [42]. Back to the case of action escrows—if users don't grasp that their conditional commitments remain private until the trigger criterion is reached, they may still experience the same hesitation and social risk that the escrow was designed to mitigate. Theoretical properties only materialize when participants comprehend the system enough to follow its intended strategies. For escrows to succeed in social computing systems, they must be explained clearly and embody a level of simplicity that makes their protective properties intuitively apparent. Complex escrow designs with multiple contingencies or unclear triggers may technically solve coordination problems, but if users cannot easily grasp how their interests are being protected, they will fail in practice and will ultimately be abandoned.

Chapter 8

Conclusion

This dissertation formalizes action escrows as a design pattern to mitigate first-mover disadvantages in online communities. By shielding individual risk through conditional actions, action escrows offer a powerful mechanism to address long-standing HCI challenges like silent majorities and collective action failures. While action escrows are not without limitations, understanding their design space can enable thoughtful implementations that balance their coordination benefits with potential risks.

To conclude, this chapter will review the main contributions of this dissertation and consider important avenues for future work.

8.1 Summary of Contributions

There's growing concern that online speech misrepresents public opinion. For instance, in American political discourse, most people hold moderate views, yet online discussion appears highly polarized. If social platforms have democratized political speech, why don't moderates dominate the conversation?

The problem is that even without technological restrictions, speech faces extensive *social* regulation. Researchers have repeatedly found that people won't express moderate views first because they fear being in the minority. Such first-mover disadvantages drive silent majorities, bystander effects, and collective action failures across online communities.

Existing design solutions are inadequate: anonymity encourages speech but eliminates accountability, incentives don't prevent preference falsification, and educating people about actual opinion distributions remains unwieldy.

This dissertation presents a new approach—action escrows—that can be applied to lower first-mover disadvantages and make progress on this broad set of problems.

Nooks (Chapter 4) showed how action escrows can be applied to lower first-mover disadvantages involved in bringing up new topics in a community. Our field deployment revealed how *Nooks* catalyzed new conversations, it helped identify initiatives with critical mass, and it promoted inclusivity in the community.

Empathosphere (Chapter 6) showed how action escrows present a novel opportunity to provide translucence into privately-held viewpoints that would otherwise remain entirely hidden

from the community.

To inform future implementations, this dissertation has outlined a broader design space of action escrows (Chapter 5), detailed the limitations and risks of introducing action escrows into online communities, identified how these risks can be mitigated, and synthesize broader opportunities for escrow mechanisms to address long-standing challenges in HCI (Chapter 7).

Together, the ideas presented in this dissertation lay the foundation for designing social computing systems that re-engage dormant voices, counter false polarization, and foster inclusion, in digital spaces and ultimately in broader society.

8.2 Future Work

As online communities increasingly connect people across different identities, affinities, and ideologies, a familiar pattern seems to emerge: vocal minorities stir discontent and spark outrage, shaping the perceptions of the majority who passively lurk. I have hoped to demonstrate that rather than thinking of the presence of digital spaces as the problem that has derailed civil and constructive discourse, we can treat it as an opportunity: by thoughtfully embedding new mechanisms into these digital spaces we can give voice to those typically silent, online and even offline. The following directions aim to redesign our digital spaces toward these goals, moving beyond the assumption that polarization and toxicity are inevitable features of online interaction.

8.2.1 Participation Mechanisms for Complex Community Structures

Researchers studying gun discourse on Reddit [66] have found that members of both liberal anti-gun and conservative pro-gun subreddits want cross-ideological dialogue but avoid posting in opposing subreddits due to downvote fears. This persists even when many individuals across the ideological divide share moderate views—like support for policy changes that would allow suicidal people to temporarily leave their guns with friends—that could find broad agreement.

The same dynamic prevents people from expressing moderate opinions even within their own subreddits, despite these views often having widespread support. Pro-gun members might hesitate to suggest higher minimum purchase ages, fearing backlash from their own community.

How might we design mechanisms that bridge across these divides? The systems explored in this dissertation work for small, simple communities like group chats or Slack workspaces. Scaling to complex community structures, like Reddit, demands different approaches where patterns like action escrows become building blocks for larger systems. This scaling also introduces new challenges: preventing abuse within escrowed spaces and also avoiding excessive fragmentation that could scatter activity and undermine community cohesion.

8.2.2 Social Computing Systems to Strengthen Our Civic Muscles

The systems in this dissertation focus on inclusive participation in online communities, but mounting evidence shows that online interactions shape offline attitudes and behaviors. A recent study demonstrates that active participation in city subreddits predicts increased interest in local community engagement [3].

This suggests a powerful opportunity: can we strengthen people's civic capacities through mechanisms that let them practice these skills in online communities?

Online communities can function as gyms for our civic muscles. Just as physical gyms provide structured environments for building strength, thoughtfully designed online spaces can become places where we regularly exercise essential democratic skills—speaking up when we disagree, listening to opposing viewpoints, respecting differences, and organizing collective action.

With intentional design, these digital spaces have the potential to help people build the confidence and competence needed for meaningful civic participation in their offline communities. The mechanisms explored in this dissertation—from action escrows to preference revelation systems—represent early equipment for this kind of civic fitness center.

8.2.3 Placing the Design and Evaluation of Algorithmic Interventions on Firm Social Scientific Foundations

Action escrows are examples of algorithmic interventions in social systems. Yet we currently lack systematic frameworks for predicting and evaluating how collectives will respond to such algorithmic interventions.

This gap becomes critical especially when trying to deploy interventions in more complex, large-scale settings. When we introduce new voting mechanisms, content moderation systems, participation tools, social AI agents into existing online communities, we need robust ways to anticipate both intended and unintended consequences. How do perceptions of algorithmic mediation affect trust between community members? What causes some AI agents to spark antisocial troll interactions while others don't?

We can pursue two complementary approaches to address this challenge, both of which tightly bind the design and evaluation of algorithmic interventions to social scientific knowledge. First, we can develop frameworks that systematically connect specific dimensions of system design to measurable changes in public attitudes and behaviors. For instance, in previous work I've shown how metaphors associated with social AI agents drive collective reactions [44]. In a similar spirit, we can develop theoretical models that can predict how different design choices will influence specific collective outcomes.

Second, we can develop simulation sandboxes that allow us to anticipate effects before deploying interventions at scale. The key question becomes: at what fidelity should we embed the "laws" of collective behavior into these simulations, and which "laws" are most critical to capture? If successful, such simulations could help us understand how communities might adapt to, resist, or be transformed by new interventions before deployment.

Bibliography

[1] Starting New Online Communities. In *Building Successful Online Communities: Evidence-Based Social Design*. The MIT Press, 03 2012. ISBN 9780262298315. doi: 10.7551/mitpress/8472.003.0007. URL https://doi.org/10.7551/mitpress/8472.003.0007. 4.1.1, 4.1.1, 4.1.1

[2] Mattias Arvola. Interaction design patterns for computers in sociable use. *International journal of computer applications in technology*, 25(2-3):128–139, 2006. 3

[3] Marianne Aubin Le Quéré and Sanjay R Kairam. Welcome to the neighborhood: Assessing localized social media use and pro-community attitudes in a multi-national survey. *Social Media+ Society*, 11(2):20563051251333490, 2025. 8.2.2

[4] Ian Ayres and Cait Unkovic. Information escrows. *Mich. L. Rev.*, 111:145, 2012. 3.1

[5] Shreya Bali, Pranav Khadpe, Geoff Kaufman, and Chinmay Kulkarni. Nooks: Social spaces to lower hesitations in interacting with new people at work. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–18, 2023. 4

[6] Sepideh Bazazi, Jorina von Zimmermann, Bahador Bahrami, and Daniel Richardson. Self-serving incentives impair collective decisions by increasing conformity. *PloS one*, 14(11): e0224725, 2019. 3.2.2

[7] Alan D Berkowitz. An overview of the social norms approach. *Changing the culture of college drinking: A socially situated health communication campaign*, 1:193–214, 2005. 3.2.3

[8] Michael S Bernstein, Adam Marcus, David R Karger, and Robert C Miller. Enhancing directed content sharing on the web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 971–980, 2010. 4.3, 4.3.5

[9] Cristina Bicchieri. Norms, conventions, and the power of expectations. *Philosophy of social science: A new introduction*, 208, 2014. 2.1

[10] Cristina Bicchieri and Alex Chavez. Behaving as expected: Public information and fairness norms. *Journal of Behavioral Decision Making*, 23(2):161–178, 2010. 2.1

[11] Jan O Borchers. A pattern approach to interaction design. In *Proceedings of the 3rd conference on Designing interactive systems: processes, practices, methods, and techniques*, pages 369–378, 2000. 3

[12] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77, 2006. 4.2.3

[13] Leonardo Bursztyn, Alessandra L González, and David Yanagizawa-Drott. Misperceived social norms: Women working outside the home in saudi arabia. *American economic review*, 110(10):2997–3029, 2020. 3.2.3

[14] Erin Carbone, George Loewenstein, Cass R Sunstein, and Linda Dezső. Spirals of shame: The bi-directional relationship between shame and disclosure. *Available at SSRN*, 2025. (document), 2.1, 2.1

[15] Kalyan Chatterjee and William Samuelson. Bargaining under incomplete information. *Operations research*, 31(5):835–851, 1983. 3.1

[16] Yu Chen and Pearl Pu. Healthytogether: exploring social incentives for mobile fitness applications. In *Proceedings of the second international symposium of chinese chi*, pages 25–34, 2014. 3.2.2

[17] Justin Cheng and Michael Bernstein. Catalyst: Triggering collective action with thresholds. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '14, page 1211–1221, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450325400. doi: 10.1145/2531602.2531635. URL https://doi.org/10.1145/2531602.2531635. 4.1.1, 5.2.1, 5.2.1

[18] Jessica Nicole Cooperstein. *Initial Development of a Team Viability Measure*. PhD thesis, DePaul University, 2017. 6.2.5

[19] Anna L Cox, Sandy JJ Gould, Marta E Cecchinato, Ioanna Iacovides, and Ian Renfree. Design frictions for mindful interactions: The case for microboundaries. In *Proceedings of the 2016 CHI Conf. Extended Abstracts on Human Factors in Computing Systems*, pages 1389–1397, 2016. 6.2.2

[20] Sauvik Das, Adam DI Kramer, Laura A Dabbish, and Jason I Hong. Increasing security sensitivity with social proof: A large-scale experimental confirmation. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pages 739–749, 2014. 3.2.3

[21] William DeJong, Shari Kessel Schneider, Laura Gomberg Towvim, Melissa J Murphy, Emily E Doerr, Neal R Simonsen, Karen E Mason, and Richard A Scribner. A multisite randomized trial of social norms marketing campaigns to reduce college student drinking. *Journal of studies on alcohol*, 67(6):868–879, 2006. 3.2.3

[22] Morton Deutsch and Robert M. Krauss. *Theories in Social Psychology*. Basic Books, 1965. 1.3

[23] Alicia DeVrio, Motahhare Eslami, and Kenneth Holstein. Building, shifting, & employing power: A taxonomy of responses from below to algorithmic harm. In *Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 1093–1106, 2024. 7.3.3

[24] Maeve Duggan. Online harassment 2017. 2017. 2

[25] Linn Van Dyne, Soon Ang, and Isabel C Botero. Conceptualizing employee silence and employee voice as multidimensional constructs. *Journal of management studies*, 40(6):

1359–1392, 2003. 1, 2

[26] Thomas Erickson and Wendy A. Kellogg. Social translucence: An approach to designing systems that support social processes. *ACM Trans. Comput.-Hum. Interact.*, 7(1):59–83, mar 2000. ISSN 1073-0516. doi: 10.1145/344949.345004. URL https://doi.org/10.1145/344949.345004. 6.4

[27] Daniel Franzen, Claudia Müller-Birn, and Odette Wegwarth. Communicating the privacy-utility trade-off: Supporting informed data donation with privacy decision interfaces for differential privacy. *Proceedings of the ACM on Human-Computer Interaction*, 8 (CSCW1):1–56, 2024. 7.3.3

[28] Yue Fu, Sami Foell, Xuhai Xu, and Alexis Hiniker. From text to self: Users' perception of aimc tools on interpersonal communication and self. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–17, 2024. 7.4.1

[29] Eric Gilbert. Designing social translucence over social networks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 2731–2740, 2012. 6.4

[30] Thomas E Glass. Through the looking glass. In *Selecting, preparing and developing the school district superintendent*, pages 20–36. Routledge, 2013. 3.2.2

[31] Oliver L Haimson, Tianxiao Liu, Ben Zefeng Zhang, and Shanley Corvite. The online authenticity paradox: What being" authentic" on social media means, and barriers to achieving it. *Proceedings of the ACM on Human-computer Interaction*, 5(CSCW2):1–18, 2021. 2

[32] Helen Ai He, Naomi Yamashita, Chat Wacharamanotham, Andrea B Horn, Jenny Schmid, and Elaine M Huang. Two sides to every story: Mitigating intercultural conflict through automated feedback and shared Self-Reflections in global virtual teams. *Proc. ACM Hum.-Comput. Interact.*, 1(CSCW):1–21, December 2017. 6.2.3

[33] Thomas Herrmann, Marcel Hoffmann, Isa Jahnke, Andrea Kienle, Gabriele Kunau, Kai-Uwe Loser, and Natalja Menold. Concepts for usable patterns of groupware applications. In *Proceedings of the 2003 ACM International Conference on Supporting Group Work*, pages 349–358, 2003. 3

[34] Jane Hsieh, Angie Zhang, Seyun Kim, Varun Nagaraj Rao, Samantha Dalal, Alexandra Mateescu, Rafael Do Nascimento Grohmann, Motahhare Eslami, and Haiyi Zhu. Worker data collectives as a means to improve accountability, combat surveillance and reduce inequalities. In *Companion Publication of the 2024 Conference on Computer-Supported Cooperative Work and Social Computing*, pages 697–700, 2024. 7.3.3

[35] Jane Hsieh, Angie Zhang, Mialy Rasetarinera, Erik Chou, Daniel Ngo, Karen Lightman, Min Kyung Lee, and Haiyi Zhu. Supporting gig worker needs and advancing policy through worker-centered data-sharing. *arXiv preprint arXiv:2412.02973*, 2024. 7.3.3

[36] Jane Hsieh, Angie Zhang, Sajel Surati, Sijia Xie, Yeshua Ayala, Nithila Sathiya, Tzu-Sheng Kuo, Min Kyung Lee, and Haiyi Zhu. Gig2gether: Data-sharing to empower, unify and demystify gig work. *arXiv preprint arXiv:2502.04482*, 2025. 7.3.3

[37] Jane Im, Sonali Tandon, Eshwar Chandrasekharan, Taylor Denby, and Eric Gilbert. Synthesized social signals: Computationally-derived social signals from account histories. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, pages 1–12, 2020. 7.2

[38] Jane Im, Jill Dimond, Melody Berton, Una Lee, Katherine Mustelier, Mark S Ackerman, and Eric Gilbert. Yes: Affirmative consent as a theoretical framework for understanding and imagining social platforms. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–18, 2021. 7.3.3

[39] Nassim JafariNaimi and Eric M Meyers. Collective intelligence or group think? engaging participation patterns in world without oil. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 1872–1881, 2015. 2

[40] Irving L Janis. Victims of groupthink: A psychological study of foreign-policy decisions and fiascoes. 1972. 2

[41] Karen A Jehn and Elizabeth A Mannix. The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of management journal*, 44 (2):238–251, 2001. 6.2.5

[42] Klaus Peter Kaas and Heidrun Ruprecht. Are the vickrey auction and the bdm mechanism really incentive compatible?—empirical results and optimal bidding strategies in cases of uncertain willingness-to-pay. *Schmalenbach Business Review*, 58(1):37–55, 2006. 7.4.2

[43] Nabil N Kamel and Robert M Davison. Applying cscw technology to overcome traditional barriers in group interactions. *Information & Management*, 34(4):209–219, 1998. 2

[44] Pranav Khadpe, Ranjay Krishna, Li Fei-Fei, Jeffrey T Hancock, and Michael S Bernstein. Conceptual metaphors impact perceptions of human-ai collaboration. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2):1–26, 2020. 8.2.3

[45] Pranav Khadpe, Chinmay Kulkarni, and Geoff Kaufman. Empathosphere: Promoting constructive communication in ad-hoc virtual teams through perspective-taking spaces. *Proceedings of the ACM on Human-computer Interaction*, 6(CSCW1):1–26, 2022. 6

[46] Pranav Khadpe, Kimi Wenzel, George Loewenstein, and Geoff Kaufman. Explaining the reputational risks of ai-mediated communication: Messages labeled as ai-assisted are viewed as less diagnostic of the sender's moral character. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, 2025. 7.4.1

[47] Pranav Khadpe, Olivia Xu, Geoff Kaufman, and Chinmay Kulkarni. Hug reports: Supporting expression of appreciation between users and contributors of open source software packages. *Proceedings of the ACM on Human-Computer Interaction*, 9(2):1–32, 2025. 3.2.3

[48] JaeWon Kim, Soobin Cho, Robert Wolfe, Jishnu Hari Nair, and Alexis Hiniker. Privacy as social norm: Systematically reducing dysfunctional privacy concerns on social media. *Proceedings of the ACM on Human-Computer Interaction*, 9(2):1–39, 2025. 2, 2.1

[49] JaeWon Kim, Robert Wolfe, Ramya Bhagirathi Subramanian, Mei-Hsuan Lee, Jessica

Colnago, and Alexis Hiniker. Trust-enabled privacy: Social media designs to support adolescent user boundary regulation. *arXiv preprint arXiv:2502.19082*, 2025. 2, 2.1

[50] Mihee Kim. Facebook's spiral of silence and participation: The role of political expression on facebook and partisan strength in political participation. *Cyberpsychology, Behavior, and Social Networking*, 19(12):696–702, 2016. 2

[51] Bert Klandermans. Mobilization and participation: Social-psychological expansisons of resource mobilization theory. *American sociological review*, pages 583–600, 1984. 2

[52] Bran Knowles, Mark Rouncefield, Mike Harding, Nigel Davies, Lynne Blair, James Hannon, John Walden, and Ding Wang. Models and patterns of trust. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, pages 328–338, 2015. 3

[53] Robert E Kraut and Paul Resnick. Encouraging contribution to online communities. *Building successful online communities: Evidence-based social design*, pages 21–76, 2011. 3.2.2

[54] Timur Kuran. *Private truths, public lies: The social consequences of preference falsification*. Harvard University Press, 1997. 2

[55] Michelle S Lam, Janice Teoh, James A Landay, Jeffrey Heer, and Michael S Bernstein. Concept induction: Analyzing unstructured text with high-level concepts using lloom. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–28, 2024. 5.3.1

[56] Bibb Latané and John M Darley. The unresponsive bystander: Why doesn't he help? *(No Title)*, 1970. 1, 2, 2.1

[57] Sven Laumer, N Sadat Shami, Michael Muller, and Werner Geyer. The challenge of enterprise social networking (non-) use at work: a case study of how to positively influence employees' enterprise social networking acceptanc. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing*, pages 978–994, 2017. 4.3

[58] Michael Y Lee, Melissa Mazmanian, and Leslie Perlow. Fostering positive relational dynamics: The power of spaces and interaction scripts. *Academy of Management Journal*, 63(1):96–123, 2020. 6.2.5

[59] Jennifer S Lerner, Ye Li, Piercarlo Valdesolo, and Karim S Kassam. Emotion and decision making. *Annual review of psychology*, 66, 2015. 6.2.2

[60] Gilly Leshed, Jeffrey T Hancock, Dan Cosley, Poppy L McLeod, and Geri Gay. Feedback for guiding reflection on teamwork practices. In *Proceedings of the 2007 ACM International Conference on Supporting Group Work*, pages 217–220, 2007. 3.2.2

[61] Gilly Leshed, Diego Perez, Jeffrey T Hancock, Dan Cosley, Jeremy Birnholtz, Soyoung Lee, Poppy L McLeod, and Geri Gay. Visualizing real-time language-based feedback on teamwork behavior in computer-mediated groups. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 537–546, 2009. 3.2.2

[62] Ann Light. Hci as heterodoxy: Technologies of identity and the queering of interaction

with computers. *Interacting with computers*, 23(5):430–438, 2011. 1.3

[63] Yihe Liu, Anushk Mittal, Diyi Yang, and Amy Bruckman. Will ai console me when i lose my pet? understanding perceptions of ai-mediated email writing. In *Proceedings of the 2022 CHI conference on human factors in computing systems*, pages 1–13, 2022. 7.4.1

[64] Paul Benjamin Lowry, Tom L Roberts, Nicholas C Romano Jr, Paul D Cheney, and Ross T Hightower. The impact of group size and social presence on small-group communication: Does computer-mediated communication make a difference? *Small Group Research*, 37 (6):631–661, 2006. 4.1.1

[65] Xiao Ma, Jeff Hancock, and Mor Naaman. Anonymity, intimacy and self-disclosure in social media. In *Proceedings of the 2016 CHI conference on human factors in computing systems*, pages 3857–3869, 2016. 3.2.1

[66] Rijul Magu, Nivedhitha Mathan Kumar, Yihe Liu, Xander Koo, Diyi Yang, and Amy Bruckman. Understanding online discussion across difference: Insights from gun discourse on reddit. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2): 1–28, 2024. 2, 2.1, 8.2.1

[67] Andrew Mao, Yiling Chen, Krzysztof Z Gajos, David C Parkes, Ariel D Procaccia, and Haoqi Zhang. Turkserver: Enabling synchronous and longitudinal online experiments. In *Workshops at the Twenty-Sixth AAAI Conf. on Artificial Intelligence*, 2012. 6.2.2

[68] MarketingCharts. Consumer trend: Gen z seeks more authenticity in social media, 2022. URL `https://www.marketingcharts.com/digital/social-media-119371`. Accessed: 2025-05-05. 2

[69] David Martin, Mark Rouncefield, and Ian Sommerville. Applying patterns of cooperative interaction to work (re) design: e-government and planning. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 235–242, 2002. 3

[70] Elaine Massung, David Coyle, Kirsten F Cater, Marc Jay, and Chris Preist. Using crowdsourcing to support pro-environmental community activism. In *Proceedings of the SIGCHI Conference on human factors in Computing systems*, pages 371–380, 2013. 3.2.2

[71] Joseph Edward McGrath. *Groups: Interaction and performance*, volume 14. Prentice-Hall Englewood Cliffs, NJ, 1984. 6.2.3

[72] Matto Mildenberger and Dustin Tingley. Beliefs about climate beliefs: the importance of second-order opinions for climate politics. *British Journal of Political Science*, 49(4): 1279–1307, 2019. 3.2.3

[73] Stanley Milgram, Leonard Bickman, and Lawrence Berkowitz. Note on the drawing power of crowds of different size. *Journal of personality and social psychology*, 13(2):79, 1969. 4.1.1

[74] Dale T Miller. A century of pluralistic ignorance: what we have learned about its origins, forms, and consequences. *Frontiers in Social Psychology*, 1:1260896, 2023. 1, 2.1

[75] Dale T Miller and Cathy McFarland. Pluralistic ignorance: When similarity is interpreted as dissimilarity. *Journal of Personality and social Psychology*, 53(2):298, 1987. 2, 2.1

[76] Benedikt Morschheuser, Alexander Maedche, and Dominic Walter. Designing cooperative

gamification: Conceptualization and prototypical implementation. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 2410–2421, 2017. 3.2.2

[77] Jimin Mun, Cathy Buerger, Jenny T Liang, Joshua Garland, and Maarten Sap. Counterspeakers' perspectives: Unveiling barriers and ai needs in the fight against online hate. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–22, 2024. 5.3.1

[78] Gonzalo Navarro. A guided tour to approximate string matching. *ACM computing surveys (CSUR)*, 33(1):31–88, 2001. 5.3.1

[79] German Neubaum and Nicole C Krämer. What do we fear? expected sanctions for expressing minority opinions in offline and online communication. *Communication Research*, 45(2):139–164, 2018. 2

[80] Elisabeth Noelle-Neumann. The spiral of silence a theory of public opinion. *Journal of Communication*, 24(2):43–51, 1974. 2.1

[81] Natalie Pang, Shirley S Ho, Alex MR Zhang, Jeremy SW Ko, WX Low, and Kay SY Tan. Can spiral of silence and civility predict click speech on facebook? *Computers in Human Behavior*, 64:898–905, 2016. 3.2.1

[82] James W Pennebaker, Martha E Francis, and Roger J Booth. Linguistic inquiry and word count: Liwc 2001. *Mahway: Lawrence Erlbaum Associates*, 71(2001):2001, 2001. 6.2.5

[83] Lindsay Popowski, Yutong Zhang, and Michael S. Bernstein. Commit: Online groups with participation commitments. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW2), November 2024. doi: 10.1145/3687027. URL `https://doi.org/10.1145/3687027`. 7.3.3

[84] Deborah A Prentice and Dale T Miller. Pluralistic ignorance and alcohol use on campus: some consequences of misperceiving the social norm. *Journal of personality and social psychology*, 64(2):243, 1993. 3.2.2, 3.2.3

[85] Deborah A Prentice and Dale T Miller. The emergence of homegrown stereotypes. *American Psychologist*, 57(5):352, 2002. 3.2.2

[86] Erica Principe Principe Cruz, Nalyn Sriwattanakomen, Jessica Hammer, and Geoff Kaufman. Counterspace games for biwoc stem students. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*, pages 1–6, 2021. 4.4

[87] Li Qiwei, Francesca Lameiro, Shefali Patel, Cristi Isaula-Reyes, Eytan Adar, Eric Gilbert, and Sarita Schoenebeck. Feminist interaction techniques: Social consent signals to deter ncim screenshots. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*, pages 1–14, 2024. 7.3.3

[88] Li Qiwei, Allison McDonald, Oliver L Haimson, Sarita Schoenebeck, and Eric Gilbert. The sociotechnical stack: Opportunities for social computing research in non-consensual intimate media. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2): 1–21, 2024. 7.3.3

[89] Stephen A Rains. The implications of stigma and anonymity for self-disclosure in health blogs. *Health communication*, 29(1):23–31, 2014. 3.2.1

[90] Paul Resnick, Joseph Konstan, Yan Chen, and Robert E Kraut. Starting new online communities. *Building successful online communities: Evidence-based social design*, 231, 2012. 7.1

[91] Claire E Robertson, Kareena S Del Rosario, and Jay J Van Bavel. Inside the funhouse mirror factory: How social media distorts perceptions of norms. *Current Opinion in Psychology*, 60:101918, 2024. 1, 2, 2.1

[92] John Sabini, Kathy Cosmas, Michael Siepmann, and Julia Stein. Underestimates and truly false consensus effects in estimates of embarrassment and other emotions. *Basic and Applied Social Psychology*, 21(3):223–241, 1999. 2.1

[93] Matthew J Salganik, Peter Sheridan Dodds, and Duncan J Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *science*, 311(5762):854–856, 2006. 4.3.5

[94] William Small Schulz. *Warped Words How Online Speech Misrepresents Opinion*. PhD thesis, Princeton University, 2024. 2, 3.2.2

[95] Shane Drew Soboroff. *Group size and the trust, cohesion, and commitment of group members*. PhD thesis, The University of Iowa, 2012. 4.1.1

[96] Nouran Soliman, Hyeonsu B Kang, Matthew Latzke, Jonathan Bragg, Joseph Chee Chang, Amy Xian Zhang, and David R Karger. Mitigating barriers to public social interaction with meronymous communication. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, pages 1–26, 2024. 7.2

[97] Charles Spearman. Footrule for measuring correlation. *British Journal of Psychology*, 2 (1):89, 1906. 6.2.5

[98] Cass R Sunstein. Unleashed. *Social Research: An International Quarterly*, 85(1):73–92, 2018. 2

[99] Cass R Sunstein. *How change happens*. Mit Press, 2019. 1, 2

[100] Samuel Hardman Taylor, Dominic DiFranzo, Yoon Hyung Choi, Shruti Sannon, and Natalya N Bazarova. Accountability and empathy by design: Encouraging bystander intervention to cyberbullying on social media. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–26, 2019. 1, 2

[101] Michael Henry Tessler, Michiel A Bakker, Daniel Jarrett, Hannah Sheahan, Martin J Chadwick, Raphael Koster, Georgina Evans, Lucy Campbell-Gillingham, Tantum Collins, David C Parkes, et al. Ai can help humans find common ground in democratic deliberation. *Science*, 386(6719):eadq2852, 2024. 5.3.1

[102] Leaf Van Boven, George Loewenstein, and David Dunning. The illusion of courage in social predictions: Underestimating the impact of fear of embarrassment on other people. *Organizational Behavior and Human Decision Processes*, 96(2):130–141, 2005. 2.1

[103] José Van Dijck. 'you have one identity': performing the self on facebook and linkedin. *Media, culture & society*, 35(2):199–215, 2013. 3.2.2

[104] Kathleen Van Royen, Karolien Poels, Heidi Vandebosch, and Philippe Adam. "thinking before posting?" reducing cyber harassment on social networking sites through a reflective

message. *Computers in Human Behavior*, 66:345–352, 2017. 6.2.2

[105] William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance*, 16(1):8–37, 1961. 7.4.2

[106] Nicholas Vincent, Hanlin Li, Nicole Tilly, Stevie Chancellor, and Brent Hecht. Data leverage: A framework for empowering the public in its relationship with technology companies. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 215–227, 2021. 7.3.3

[107] Emily K Vraga, Kjerstin Thorson, Neta Kligler-Vilenchik, and Emily Gee. How individual sensitivities to disagreement shape youth political expression on facebook. *Computers in Human Behavior*, 45:281–289, 2015. 2

[108] Richard T Watson, Gerardine DeSanctis, and Marshall Scott Poole. Using a gdss to facilitate group consensus: Some intended and unintended consequences. *Mis Quarterly*, pages 463–478, 1988. 6.2.3

[109] Henry Wechsler, Toben E Nelson, Jae Eun Lee, Mark Seibring, Catherine Lewis, and Richard P Keeling. Perception and reality: a national evaluation of social norms marketing interventions to reduce college students' heavy alcohol use. *Journal of studies on alcohol*, 64(4):484–494, 2003. 3.2.3

[110] Mark E Whiting, Allie Blaising, Chloe Barreau, Laura Fiuza, Nik Marda, Melissa Valentine, and Michael S Bernstein. Did it have to end this way? understanding the consistency of team fracture. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW):1–23, November 2019. 6.2.3, 6.2.5, 6.2.5

[111] Mark E Whiting, Irena Gao, Michelle Xing, N'godjigui Junior Diarrassouba, Tonya Nguyen, and Michael S Bernstein. Parallel worlds: Repeated initializations of the same team to improve team viability. *proceedings of the ACM on Human-Computer Interaction*, 4(CSCW1):1–22, 2020. 6.2.1, 6.2.3, 6.2.5

[112] Tai-Yee Wu and David J Atkin. To comment or not to comment: Examining the influences of anonymity and social support on one's willingness to express in online news discussions. *New Media & Society*, 20(12):4512–4532, 2018. 3.2.1

[113] Siyuan Xia, Zhiru Zhu, Chris Zhu, Jinjin Zhao, Kyle Chard, Aaron J Elmore, Ian Foster, Michael Franklin, Sanjay Krishnan, and Raul Castro Fernandez. Data station: delegated, trustworthy, and auditable computation to enable data-sharing consortia with a data escrow. *arXiv preprint arXiv:2305.03842*, 2023. 7.3.3

[114] Joanna C Yau and Stephanie M Reich. "it's just a lot of work": Adolescents' self-presentation norms and practices on facebook and instagram. *Journal of research on adolescence*, 29(1):196–209, 2019. 2.1, 3.2.2

[115] Chien Wen Yuan, Leslie D Setlock, Dan Cosley, and Susan R Fussell. Understanding informal communication in multilingual contexts. In *Proceedings of the 2013 Conf. on Computer supported cooperative work*, pages 909–922, 2013. 6.3.3

[116] Dora Zhao, Diyi Yang, and Michael S Bernstein. Mapping the spiral of silence: Surveying unspoken opinions in online communities. *arXiv preprint arXiv:2502.00952*, 2025. 2.1

[117] Dorothy Zhao, Mikako Inaba, and Andrés Monroy-Hernández. Understanding teenage perceptions and configurations of privacy on instagram. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2):1–28, 2022. 2, 2.1