I'm In If You're In: Action Escrows as a Design Pattern to Achieve Social Change in Online Communities

PRANAV KHADPE*, Carnegie Mellon University, USA
LINDSAY POPOWSKI*, Stanford University, USA
KYZYL MONTEIRO†, Carnegie Mellon University, USA
LINDY LE†, University of Michigan, USA
GEOFF KAUFMAN, Carnegie Mellon University, USA

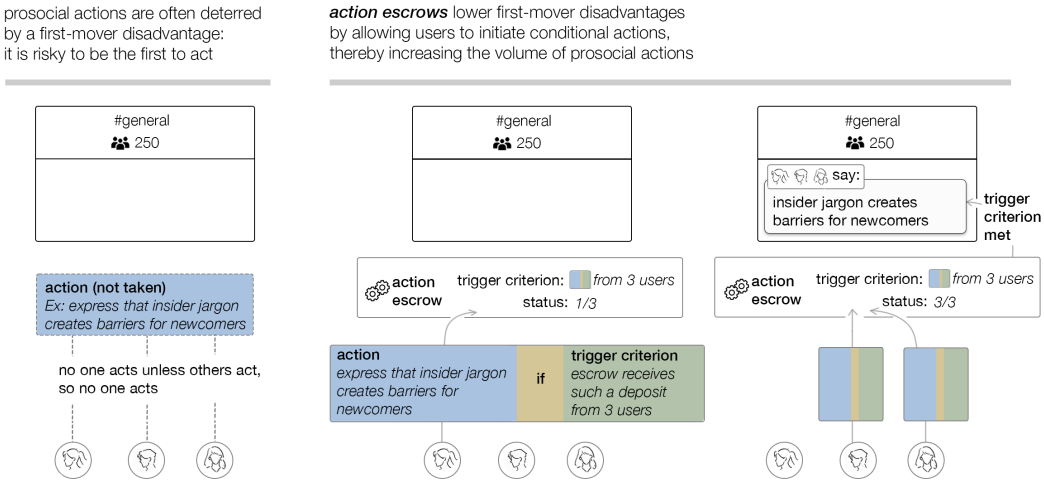

Fig. 1. In this paper, we make the unifying observation that a broad range of prosocial actions in online communities are deterred by first-mover disadvantages. We then show how a general design pattern—which we call action escrows—can be applied to lower first mover disadvantages, across a range of prosocial actions.

In an online community, prosocial actions ranging from sharing authentic opinions to intervening against misbehavior to contributing to collective action are often deterred by a first mover disadvantage: isolated

*Both authors contributed equally to this research.
†Both authors contributed equally to this research.

Authors' Contact Information: Pranav Khadpe, pkhadpe@cs.cmu.edu, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; Lindsay Popowski, popowski@stanford.edu, Stanford University, Stanford, California, USA; Kyzyl Monteiro, , Carnegie Mellon University, Pittsburgh, Pennsylvania, USA; Lindy Le, , University of Michigan, , Michigan, USA; Geoff Kaufman, gfk@cs.cmu.edu, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.

Conference acronym 'XX, Woodstock, NY

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/XXXXXXX.XXXXXXX

individuals deciding whether or not to take the first action often cannot be sure whether others would welcome it, and respond in support. 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. In this paper, we formalize how designers of online communities can lower these first-mover disadvantages through a design pattern that we 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—we formalize their scope, and utility for addressing persistent challenges in online communities. We explain the general design pattern, present design cases of implementations that apply the pattern to specific problems, describe the broader design space for action escrows, and outline opportunities for the application of escrows more generally, to address CSCW challenges.

CCS Concepts: • **Human-centered computing → Collaborative and social computing theory, concepts and paradigms**.

Additional Key Words and Phrases: online communities, design pattern, escrow mechanisms, norm misperception, critical mass

ACM Reference Format:
Pranav Khadpe, Lindsay Popowski, Kyzyl Monteiro, Lindy Le, and Geoff Kaufman. 2018. I'm In If You're In: Action Escrows as a Design Pattern to Achieve Social Change in Online Communities. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation email (Conference acronym 'XX)*. ACM, New York, NY, USA, 25 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

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 CSCW literature. For instance, the *online authenticity paradox* [19] describes how most people privately desire online authenticity, yet refrain from being authentic out of uncertainty about whether others will welcome their authentic self. Similarly, the failure of a community to intervene on misbehavior is often attributed to *the bystander effect* [70]: 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* [69]: 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. First mover disadvantages don't just impede one-time prosocial actions; they also act as a brake on positive norm change. In the situations outlined above, for instance, they allow norms of self-censorship and inaction to persist in a community, even when a substantial number of community members privately desire the opposite.

In this paper, we show how designers of online communities can lower first-mover disadvantages through a general design pattern that we call an *action escrow*—a mechanism that allows users

to deposit 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.

Action escrows offer an advantage over complete anonymity in balancing risk and practicality. While anonymity can also reduce first-mover disadvantages by permanently hiding identities, action escrows allow for conditional de-anonymization. Escrows can be configured such that identities of depositors are revealed to each other, or even publicly, once the trigger criterion is met. This makes escrows especially useful where eventual de-anonymization is required (e.g. when collective action requires physically showing up) or where verified participation is desired (e.g. to determine actual levels of support for a particular movement).

The goal of this paper is to formalize action escrows as a design pattern, and show their broad applicability in encouraging prosocial actions that are dissuaded by first-mover disadvantages. We start by providing a functional typology of situations with first-mover disadvantages in Section 2, revealing how first-mover disadvantages underlie several dilemmas described in CSCW literature on online communities. In doing so, we map the terrain of problems that action escrows can productively address. Next, in Section 3, we introduce the design pattern of an action escrow, grounding it in traditional escrows used in legal and economic processes. Here, we also describe the advantages that action escrows offer over existing behavioral design paradigms in CSCW research—anonymity and extrinsic incentives—that may also be deployed to mitigate first-mover disadvantages. Then, in Section 4, we describe design cases of four deployed social computing systems that instantiate action escrows in order to provide concrete examples of how the pattern can be applied in practice and reveal the underlying design space. Finally, in Section 5, we discuss strategies to mitigate potential risks of using action escrows and highlight opportunities to use escrow mechanisms to address other long-standing CSCW challenges.

This paper makes four main contributions:

(1) We **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, we also reveal its potential to resolve numerous online community issues rooted in first-mover disadvantages.
(2) We **show how action escrows can be applied in practice.** To inform future implementations, we provide design cases of deployed research prototypes and publicly available systems that instantiate action escrows, and outline the design space of action escrows. Viewing these systems through our 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.
(3) We **characterize the limitations of action escrows.** We reflect on the limitations and risks of introducing action escrows into online communities, and identify how the risks can be mitigated.
(4) We **synthesize broader opportunities for escrow mechanisms** to address CSCW challenges beyond first-mover disadvantages.

2 First-Mover Disadvantages

We use the term *first-mover disadvantage* to characterize 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 [49]. 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 [34, 37]. 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 CSCW contexts. For instance, *groupthink* [26, 27, 29]—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 [12], 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* [39] in online communities, 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 [70]. Similarly, the *online authenticity paradox* [19]—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 [31, 79]—can be traced back to the first-mover disadvantage to authentic self-expression [31, 32, 79]. 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 [69]. However, the strong disincentive to speaking up can prevent any public opposition, causing the change to fizzle out [68].

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 [33, 44, 53, 64, 66, 75]. Between 60% and 70% Americans have been bystanders of misbehavior directed at others online [11, 70], yet only 30% of them report having intervened [11, 70]. Nearly 45% of social media users think people ought to show more of their "real" selves [45], yet underestimate how many others think similarly and rarely aim for authenticity themselves (only 32% report making the effort) [45].

First-mover disadvantages 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 opinions in a community can begin to seem more extreme than it actually is [64], 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* [54, 80]. If no one attempts being authentic, then it can normalize filtered and distorted beauty standards that might actually be unrealistic [78], further discouraging authenticity [31, 32, 79]. Similarly, if no one intervenes in response to misbehavior, it can make misbehavior seem more acceptable in a community than it actually is [39]. 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 [48].

We 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.

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 2). 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.

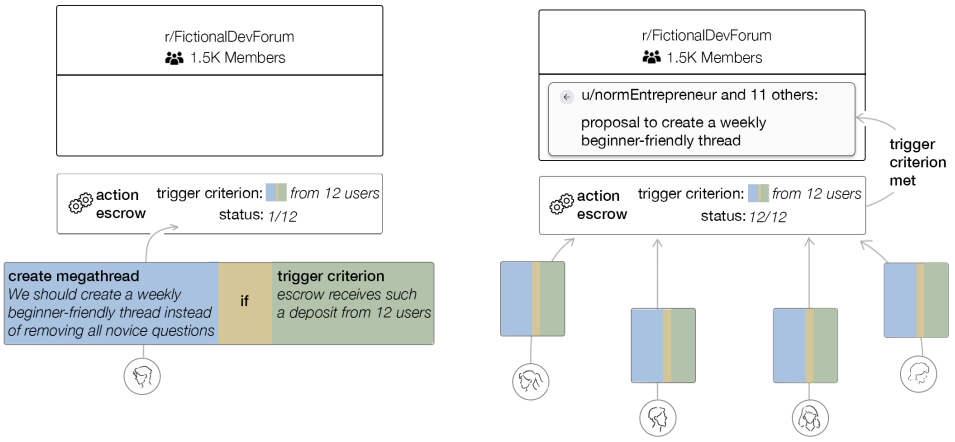

Fig. 2. An hypothetical vignette to illustrate action escrows in, well, action: In r/FictionalDevForum, many subscribers privately questioned the subreddit's ban on beginner questions, but feared downvote brigades if they challenged the status quo established by vocal power users. Jordan used the subreddit's 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.

We present action escrows as a design pattern [2, 6, 20, 35, 46]: a recipe rather than a frozen 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, we provide design cases in Section 4, 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. Here, we describe how action escrows relate to financial and legal escrows, and describe the benefits that action escrows offer over existing CSCW behavior design paradigms. Then, in Section 4, we discuss applications of the pattern.

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 [7] and campus sexual assault reporting [3]. The core function of an escrow is to support "conditional intermediated communication" [3]. 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 [7], 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 [3] 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 CSCW 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.

3.2 Advantages Over Existing CSCW Behavior Design Paradigms

Significant CSCW research has attempted to address many of these problems that we trace back to first-mover disadvantages. Here we outline two influential behavior design approaches that have come out of this work, and the benefits that action escrows offer over each.

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 [43]. Anonymity is especially important and more often used on platforms where the discussion topics or actions are stigmatized and can help assuage embarrassment [62]. Even *perceived* anonymity can be empowering. Perceived anonymity can lessen the spiral of silence effect [76] and even the relative visibility difference of liking versus commenting can affect how much people self-silence [55].

However, even in anonymous communities, social risks do not completely disappear [43]: 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 [36]. 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 [47], implementing point-based reward systems for regular contributions [47], or adopting systems similar to karma requirements whereby subreddits gate privileges until a member demonstrates minimum activity [18]. 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" [8, 50]. Similarly, visualizing interaction imbalances creates social disincentives that simultaneously discourage individuals from dominating or remaining silent [40, 41].

But because these systems do not directly address perceived risks, they can cause people to falsify their preferences while chasing extrinsic incentives [5]. 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 [5, 66, 72, 78]. This can perpetuate groupthink [5], silent majorities [66], online inauthenticity [72, 78], 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 [57, 58]. Action escrows, by contrast, mitigate these risks of false preference signaling.

4 Applying Action Escrows

In this section, we turn to showing how the design pattern of an action escrow can be applied in practice. 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 four illustrative design cases of existing action escrow systems. These cases enlist action escrows to lower first-mover disadvantages in four different contexts: (1) planning a collective action effort; (2) suggesting a new discussion topic; (3) forwarding content into public forums; and (4) admitting romantic interest. Illustrative design cases are commonly deployed in CSCW research [e.g. 1, 15, 24, 65] to show how conceptual ideas apply in practice. Here, they help unlock the design pattern's explanatory power (we can explain *why* the existing systems "work") and generative

power (the cases provide jumping-off points for envisioning new designs). We follow our discussion of the cases with the fuller design space of action escrows, including other parameters that designers can configure.

4.1 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 starting a conversation about a topic that hasn't yet surfaced in a community (e.g. starting a thread about potentially reintroducing Program Committee meetings to the CSCW review process—a bit meta, we know). 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.

4.1.1 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, when initiating a potentially off-topic conversation, 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.

4.1.2 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 subsequent deposits (disclosing *only receipt of first deposit*). The cases we describe next disclose either progress towards trigger or receipt of first deposit.

4.2 Design Cases

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

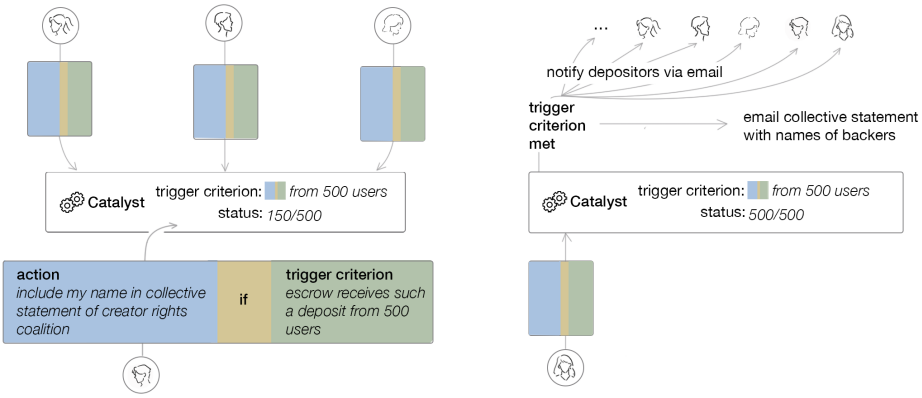

Fig. 3. Catalyst instantiates an action escrow to lower first-mover disadvantages in collective action efforts. Here for instance, creators add their name to a collective statement to protect their rights, with their names only revealed when 500 depositors commit, enabling unified action with reduced individual vulnerability.

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 our paper'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.

4.2.1 ***Catalyst****: Lowering First-Mover Disadvantages in Committing to Collective Action Efforts.* Catalyst [9] 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. Subsequently, 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 3 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 revised, creating substantial 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 [9]. 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). However, like in the usage scenario we describe, Catalyst primarily addresses situations where the initial depositor is aware of and connected to others who share their concern, with hesitation primarily about making public commitments. Yet in many cases, individuals simply don't know who in their community shares their interests and concerns, or whether such like-minded people even exist. Next, we explore how action escrows can address this discovery challenge.

4.2.2 ***Nooks**: Lowering First-Mover Disadvantages in Bringing up New Topics in a Community.*

Nooks [4] 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 (*trigger criterion*). The individual making the first deposit can specify the topic. The application reveals the proposed topic (but not the identity of the depositor) to everyone in the workspace (*interim disclosure*) and waits 24 hours to receive deposits of interest from others in the workspace. 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. Nooks uses *reciprocal deposits* as its trigger criterion and in its interim disclosures reveals *only receipt of the first deposit.* Figure 4 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 multi-national 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

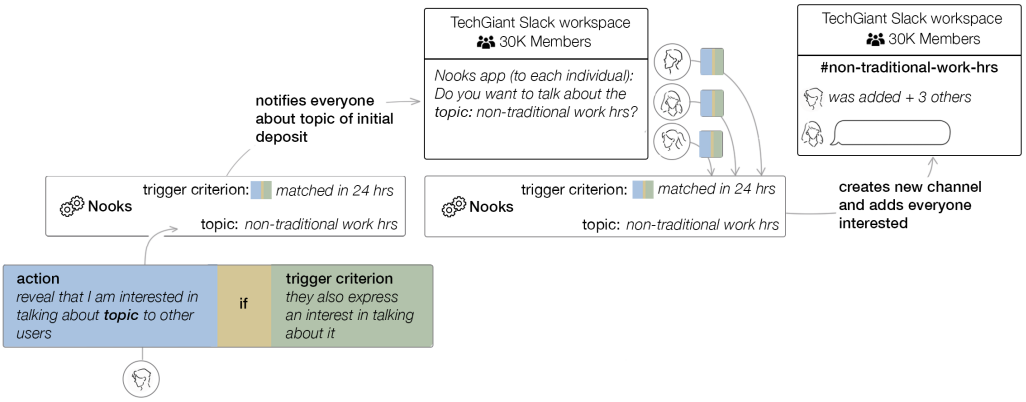

Fig. 4. Nooks instantiates an action escrow to lower first-mover disadvantages in bringing up new topics in a community. Here, a user anonymously proposes a discussion topic about non-traditional work hours. The topic (but not the depositors identity) is shown to all workspace members. Once others express interest within 24 hours, a new channel is created that includes only interested participants, revealing their identities to each other

in person, but is unsure about who might be interested and whether bringing this up would 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.

Discussion: 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* [59], 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. While Nooks facilitates the creation of these private spaces, next, with Burst, we explore how ideas can move from private conversations back into public forums.

4.2.3 ***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.

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, 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

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

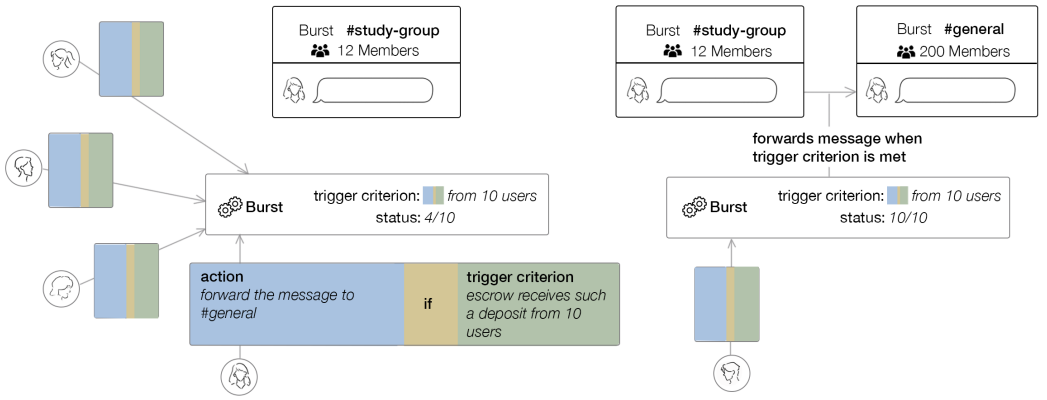

Fig. 5. Burst instantiates an action escrow to lower first-mover disadvantages in forwarding content from private channels to public forums. Here, a student proposes forwarding a message from a small study group to the #general channel, requiring support from 10 group members. The system shows progress 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.

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.

4.2.4 ***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.*

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

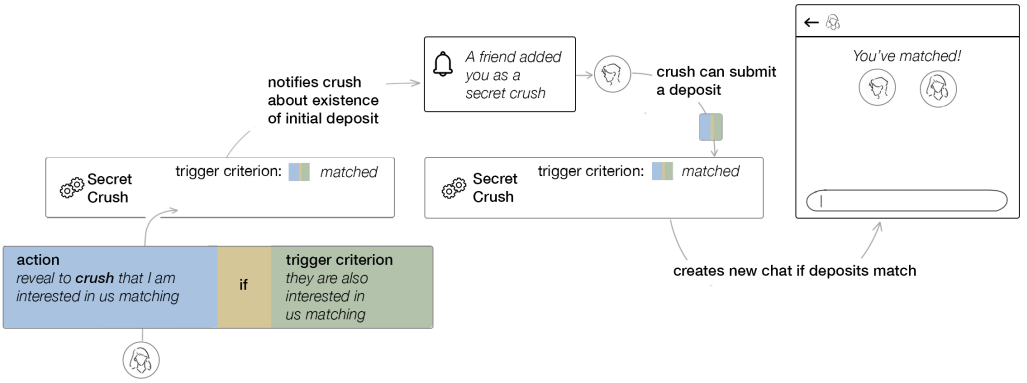

Fig. 6. Secret Crush instantiates an action escrow to lower first-mover disadvantages in admitting romantic interest, similar to familiar dating app matching algorithms but specifically for existing Facebook friends. It uses reciprocal interest as its trigger criterion. Here, when a user adds someone 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

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 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).

4.3 Design Space of Action Escrows

By presenting four distinct contexts 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 7, we summarize a fuller design space, including auxiliary parameters, that can merit explicit consideration when implementing action escrows. Here we discuss these auxiliary parameters and the potential choices for each.

4.3.1 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

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)*	

Fig. 7. The design space of action escrows.

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 subsequent 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 [52] and recent advances in large language models [38, 71]. For example, in public counterspeech applications [51], 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.

4.3.2 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.

4.3.3 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.

4.3.4 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 Discussion

So far, we have introduced the design pattern of action escrows and described the broad range of problems they can address: those with first-mover disadvantages. To inform future applications of the pattern, we have provided concrete cases of existing systems that apply the pattern, and have teased out an underlying design space. Throughout, we have also tried to reveal the relationships between previously disconnected problems (*silent majorities*, *critical mass*) and their technical remedies (Nooks, Catalyst), exposing common roots in first-mover disadvantages. In this section, we first reflect on action escrows' limitations in achieving coordinated action. Then, we discuss potential risks of introducing action escrows in communities, while identifying design approaches to mitigate these risks. Finally, we broaden our focus beyond action escrows and first-mover disadvantages to synthesize how escrow mechanisms can address a broader-set of CSCW challenges, and explore the gap between their theoretical utility and practical adoption.

5.1 Limitations of Action Escrows in Achieving Coordinated Action

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 [63].

5.2 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) [67], pseudonymity options that persist even after threshold activation, or social signals [25] that help participants gauge the trustworthiness of other escrow members before full identity disclosure occurs.

Escrow mechanisms to address CSCW challenges and opportunities

What is the challenge or opportunity addressed?	Lowering first-mover disadvantages **(action escrows)**	Translucence into privately-held opinions across community **(5.3.1)**	Reinforcing community standards **(5.3.2)**	Supporting safe interactions **(5.3.3)**	Supporting data-driven collective action **(5.3.4)**	...
What is withheld by escrow?	execution of socially risky action	privately elicited opinions	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	opinions are aggregated	user commits to norm	all parties independently initiate contact	data is tranformed to be non-identifying	???
Examples	Catalyst [9], Nooks [4], Burst, Facebook's "Secret Crush"	Empathosphere [30]	Commit [56], BeReal	Tinder	Gig2Gether [23]	???

Fig. 8. An overview of escrow mechanisms applied to address CSCW challenges. This is not an exhaustive list; additional escrow applications beyond those explicitly documented here may exist or be potential directions for exploration.

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.

5.3 Beyond Action Escrows: Escrow Mechanisms for other CSCW challenges

Throughout this paper, 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 CSCW 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 8 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 CSCW challenge and explain how escrow mechanisms can be formulated to address it.

5.3.1 Escrows for Translucence Into Privately-Held Opinions. Escrow mechanisms present a novel opportunity to provide translucence [13, 17] 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 [30] 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.

5.3.2 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 [56] 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.

5.3.3 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 [24, 60, 61].

5.3.4 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 [14, 21, 22]. Here, the escrow agent withholds the *permission* to forward a user's data donation until the data is transformed to be non-identifying [77]. Gig2Gether [23] 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 [10, 74].

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

Through this paper, we have 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.

5.4.1 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[3]. 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 [16, 42]. 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.

5.4.2 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 [73]. Despite its mathematical elegance, the mechanism's optimal strategy remains invisible to users without specialized knowledge: there's nothing in the auction description itself that guides users toward truthful bidding or makes the benefits of honesty apparent [28]. 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 [28]. 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.

6 Conclusion

This paper 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 CSCW challenges like silent majorities and collective action failures. Our analysis has bridged previously disconnected systems—from Catalyst to Kickstarter to Nooks to Tinder—revealing their shared conceptual foundations. While action escrows are not without limitations, understanding their design space can enable thoughtful implementations that balance their coordination benefits with potential risks. As activities in online communities increasingly flow beyond digital boundaries to shape political movements, social

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

institutions, and civic discourse, we envision action escrows not merely as features for online platforms but as mechanisms for re-engaging dormant voices.

References

[1] Paul M Aoki and Allison Woodruff. 2005. Making space for stories: ambiguity in the design of personal communication systems. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 181–190.

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

[3] Ian Ayres and Cait Unkovic. 2012. Information escrows. *Mich. L. Rev.* 111 (2012), 145.

[4] Shreya Bali, Pranav Khadpe, Geoff Kaufman, and Chinmay Kulkarni. 2023. 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*. 1–18.

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

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

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

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

[9] Justin Cheng and Michael Bernstein. 2014. Catalyst: Triggering Collective Action with Thresholds. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (Baltimore, Maryland, USA) *(CSCW '14)*. Association for Computing Machinery, New York, NY, USA, 1211–1221. doi:10.1145/2531602.2531635

[10] Alicia DeVrio, Motahhare Eslami, and Kenneth Holstein. 2024. 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*. 1093–1106.

[11] Maeve Duggan. 2017. Online harassment 2017. (2017).

[12] Linn Van Dyne, Soon Ang, and Isabel C Botero. 2003. Conceptualizing employee silence and employee voice as multidimensional constructs. *Journal of management studies* 40, 6 (2003), 1359–1392.

[13] Thomas Erickson and Wendy A. Kellogg. 2000. Social Translucence: An Approach to Designing Systems That Support Social Processes. *ACM Trans. Comput.-Hum. Interact.* 7, 1 (mar 2000), 59–83. doi:10.1145/344949.345004

[14] Daniel Franzen, Claudia Müller-Birn, and Odette Wegwarth. 2024. 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 (2024), 1–56.

[15] Seth Frey, PM Krafft, and Brian C Keegan. 2019. " This Place Does What It Was Built For" Designing Digital Institutions for Participatory Change. *Proceedings of the ACM on human-computer interaction* 3, CSCW (2019), 1–31.

[16] Yue Fu, Sami Foell, Xuhai Xu, and Alexis Hiniker. 2024. 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*. 1–17.

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

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

[19] Oliver L Haimson, Tianxiao Liu, Ben Zefeng Zhang, and Shanley Corvite. 2021. 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 (2021), 1–18.

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

[21] Jane Hsieh, Angie Zhang, Seyun Kim, Varun Nagaraj Rao, Samantha Dalal, Alexandra Mateescu, Rafael Do Nascimento Grohmann, Motahhare Eslami, and Haiyi Zhu. 2024. 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*. 697–700.

[22] Jane Hsieh, Angie Zhang, Mialy Rasetarinera, Erik Chou, Daniel Ngo, Karen Lightman, Min Kyung Lee, and Haiyi Zhu. 2024. Supporting Gig Worker Needs and Advancing Policy Through Worker-Centered Data-Sharing. *arXiv preprint arXiv:2412.02973* (2024).

[23] Jane Hsieh, Angie Zhang, Sajel Surati, Sijia Xie, Yeshua Ayala, Nithila Sathiya, Tzu-Sheng Kuo, Min Kyung Lee, and Haiyi Zhu. 2025. Gig2Gether: Data-sharing to Empower, Unify and Demystify Gig Work. *arXiv preprint arXiv:2502.04482* (2025).

[24] Jane Im, Jill Dimond, Melody Berton, Una Lee, Katherine Mustelier, Mark S Ackerman, and Eric Gilbert. 2021. 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*. 1–18.

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

[26] Nassim JafariNaimi and Eric M Meyers. 2015. 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*. 1872–1881.

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

[28] Klaus Peter Kaas and Heidrun Ruprecht. 2006. 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 (2006), 37–55.

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

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

[31] JaeWon Kim, Soobin Cho, Robert Wolfe, Jishnu Hari Nair, and Alexis Hiniker. 2025. Privacy as Social Norm: Systematically Reducing Dysfunctional Privacy Concerns on Social Media. *Proceedings of the ACM on Human-Computer Interaction* 9, 2 (2025), 1–39.

[32] JaeWon Kim, Robert Wolfe, Ramya Bhagirathi Subramanian, Mei-Hsuan Lee, Jessica Colnago, and Alexis Hiniker. 2025. Trust-Enabled Privacy: Social Media Designs to Support Adolescent User Boundary Regulation. *arXiv preprint arXiv:2502.19082* (2025).

[33] Mihee Kim. 2016. 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 (2016), 696–702.

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

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

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

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

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

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

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

[41] Gilly Leshed, Diego Perez, Jeffrey T Hancock, Dan Cosley, Jeremy Birnholtz, Soyoung Lee, Poppy L McLeod, and Geri Gay. 2009. 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*. 537–546.

[42] Yihe Liu, Anushk Mittal, Diyi Yang, and Amy Bruckman. 2022. 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*. 1–13.

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

[44] Rijul Magu, Nivedhitha Mathan Kumar, Yihe Liu, Xander Koo, Diyi Yang, and Amy Bruckman. 2024. Understanding Online Discussion Across Difference: Insights from Gun Discourse on Reddit. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–28.

[45] MarketingCharts. 2022. *Consumer Trend: Gen Z Seeks More Authenticity in Social Media.* https://www.marketingcharts.com/digital/social-media-119371 Accessed: 2025-05-05.

[46] David Martin, Mark Rouncefield, and Ian Sommerville. 2002. 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*. 235–242.

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

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

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

[50] Benedikt Morschheuser, Alexander Maedche, and Dominic Walter. 2017. Designing cooperative gamification: Conceptualization and prototypical implementation. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. 2410–2421.

[51] Jimin Mun, Cathy Buerger, Jenny T Liang, Joshua Garland, and Maarten Sap. 2024. 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*. 1–22.

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

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

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

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

[56] Lindsay Popowski, Yutong Zhang, and Michael S. Bernstein. 2024. Commit: Online Groups with Participation Commitments. *Proc. ACM Hum.-Comput. Interact.* 8, CSCW2, Article 488 (Nov. 2024), 28 pages. doi:10.1145/3687027

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

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

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

[60] Li Qiwei, Francesca Lameiro, Shefali Patel, Cristi Isaula-Reyes, Eytan Adar, Eric Gilbert, and Sarita Schoenebeck. 2024. Feminist Interaction Techniques: Social Consent Signals to Deter NCIM Screenshots. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology*. 1–14.

[61] Li Qiwei, Allison McDonald, Oliver L Haimson, Sarita Schoenebeck, and Eric Gilbert. 2024. The Sociotechnical Stack: Opportunities for Social Computing Research in Non-consensual Intimate Media. *Proceedings of the ACM on Human-Computer Interaction* 8, CSCW2 (2024), 1–21.

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

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

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

[65] Nathan Schneider, Primavera De Filippi, Seth Frey, Joshua Z Tan, and Amy X Zhang. 2021. Modular politics: Toward a governance layer for online communities. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–26.

[66] William Small Schulz. 2024. *Warped Words How Online Speech Misrepresents Opinion.* Ph. D. Dissertation. Princeton University.

[67] Nouran Soliman, Hyeonsu B Kang, Matthew Latzke, Jonathan Bragg, Joseph Chee Chang, Amy Xian Zhang, and David R Karger. 2024. Mitigating Barriers to Public Social Interaction with Meronymous Communication. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*. 1–26.

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

[69] Cass R Sunstein. 2019. *How change happens.* Mit Press.

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

[71] 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. 2024. AI can help humans find common ground in democratic deliberation. *Science* 386, 6719 (2024), eadq2852.
[72] José Van Dijck. 2013. 'You have one identity': performing the self on Facebook and LinkedIn. *Media, culture & society* 35, 2 (2013), 199–215.
[73] William Vickrey. 1961. Counterspeculation, auctions, and competitive sealed tenders. *The Journal of finance* 16, 1 (1961), 8–37.
[74] Nicholas Vincent, Hanlin Li, Nicole Tilly, Stevie Chancellor, and Brent Hecht. 2021. 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*. 215–227.
[75] Emily K Vraga, Kjerstin Thorson, Neta Kligler-Vilenchik, and Emily Gee. 2015. How individual sensitivities to disagreement shape youth political expression on Facebook. *Computers in Human Behavior* 45 (2015), 281–289.
[76] Tai-Yee Wu and David J Atkin. 2018. 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 (2018), 4512–4532.
[77] Siyuan Xia, Zhiru Zhu, Chris Zhu, Jinjin Zhao, Kyle Chard, Aaron J Elmore, Ian Foster, Michael Franklin, Sanjay Krishnan, and Raul Castro Fernandez. 2023. Data station: delegated, trustworthy, and auditable computation to enable data-sharing consortia with a data escrow. *arXiv preprint arXiv:2305.03842* (2023).
[78] Joanna C Yau and Stephanie M Reich. 2019. "It's just a lot of work": Adolescents' self-presentation norms and practices on Facebook and Instagram. *Journal of research on adolescence* 29, 1 (2019), 196–209.
[79] Dorothy Zhao, Mikako Inaba, and Andrés Monroy-Hernández. 2022. Understanding teenage perceptions and configurations of privacy on instagram. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–28.
[80] Dora Zhao, Diyi Yang, and Michael S Bernstein. 2025. Mapping the Spiral of Silence: Surveying Unspoken Opinions in Online Communities. *arXiv preprint arXiv:2502.00952* (2025).

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009