Dropping the F-bomb:

Algorithmic Vocabulary Control and Novelty in User Community Innovation

Abstract

Organizations increasingly deploy AI to regulate how members speak and write. I argue that such *algorithmic vocabulary control* disrupts emergent vocabulary structures and can unintentionally harm the novelty of ideas generated in organizations by constraining feedback. I theorize three cognitive mechanisms: (1) reduced vocabulary richness in feedback constrains associative recombination; (2) narrowed affective expression diminishes the salience of feedback; and (3) heightened self-monitoring slows the exchange of feedback. I test these predictions using the staggered rollout of an automated profanity filter on Steam Workshop—an online community where users create and refine new game content—analyzing 7,145 idea updates and 61,175 feedback comments across 1,045 communities. Difference-in-differences estimates show that idea novelty declines after the filter is introduced, and causal-mediation analyses align with the theory. I contribute to research on organizational design and control, vocabularies, algorithmic management, and user innovation by showing that algorithmic interventions in organizational vocabularies can inadvertently curb creative outcomes.

Jay Park
Ph.D. Candidate
Department of Strategy and Entrepreneurship
The Paul Merage School of Business
University of California, Irvine
4293 Pereira Dr. Irvine, CA 92697
jay.park@uci.edu

October 16, 2025

DRAFT

INTRODUCTION

A fundamental question in organization design is how to integrate the efforts of different organizational members to improve coordination and performance (Joseph & Sengul, 2025). To achieve this, organizations consistently monitor and evaluate members' behavior, such as through standardized work practices (cf. Cardinal et al., 2017). One important target of such behavioral control is that of regulating organizational vocabularies: the shared system of words commonly used by social collectives (Loewenstein et al, 2018). Organizations can implement such control by constraining how members speak, write, and interact, through requiring standardized vocabularies or limiting the use of certain terms, acronyms, or abbreviations (Neeley et al., 2012; Kroon et al., 2015).

To lower the costs of implementing such controls (Cardinal et al., 2017; Gibbons, 2005; Sitkin et al., 2010), firms are turning to algorithms and other intelligent technologies. Algorithmic control systems can detect, flag, substitute, and pre-compose vocabularies at scale, nudge writers through autocomplete/templates, and generate standardized text, thereby reducing marginal oversight costs while ensuring uniformity (Allen & Choudhury, 2022; Cameron, 2022, 2024; Kellogg et al., 2020; Kim et al., 2024; Koo, 2025). Algorithms are intended to streamline communication: For instance, they filter non-essential content (such as colloquialisms or emphatics) or filter toxicity (such as profanity). This trend is evident across diverse organizational settings: for example, Goldman Sachs uses compliance filters to block prohibited terms (e.g., swear words) in analyst communications (WSJ, 2010). In workplace tools, Google Workspace uses "smart compose" features that accelerate writing but nudge teams toward similar terms (Chen et al., 2019). On digital platforms, Facebook and YouTube rely on AI moderation filters that suppress toxic content (Oh & Downey, 2025).

Although algorithmic vocabulary control is often promoted as beneficial for firms (He et al., 2024), its organizational impact remains insufficiently understood. Two properties of vocabularies are responsible. First, organizational vocabularies are inherently *emergent:* they develop through ongoing interactions and adapt over time to a particular organizational context (Loewenstein, 2017; Loewenstein & Ocasio, 2005). By contrast, algorithmic vocabulary control is *rule-based*: it standardizes expression, substitutes terms, and enforces compliance at scale. Given that research demonstrates that control mechanisms frequently generate unintended consequences, such as engaging in deviant behaviors (Anteby & Chan, 2018), reduced engagement (Rahman, 2021), and lower productivity (Tong et al., 2021), what occurs when an emergent vocabulary is subjected to a rule-based algorithmic control remains essentially a 'black-box' for researchers (Shrestha et al., 2021).

Second, the vocabularies individuals use guide how they think (Boroditsky, 2011), what they are likely to think about (Gentner & Goldin-Meadow, 2003), and what they are likely to attend to (Ocasio & Joseph, 2008; Ocasio et al., 2018; Vuori, 2024). Vocabularies not only enable members to frame feedback and initiatives (Kaplan, 2008), socialize issues (Dutton & Ashford, 1993), and engage in sensegiving (Gioia & Chittipeddi, 1991), but also exert a causal—albeit not determinative—influence on their own cognition (Loewenstein et al., 2012). When emergent vocabularies are subjected to algorithmic control, the implications may extend to the cognition underlying communication, which in turn has implications for coordination (Okhuysen & Bechky, 2009) and innovation (Monge et al., 1992), the subject of this study.

The primary thesis proffered in this study is that when algorithmic vocabulary control is imposed on emergent vocabularies, it reconfigures the organization's overall *vocabulary structure*—the patterned system of word frequencies, word-to-word ties, and word-to-example

relations that guide decision making (Loewenstein et al., 2012). This reconfiguration manifests in three observable patterns in cognition. First, algorithmic vocabulary control narrows acceptable terms: members' expression collapses toward simple, repeated "safe" terms, shrinking the space for recombination of ideas. Second, algorithmic vocabulary control reduces emotional nuance: as colloquialisms, emphatics, and affect-rich vocabularies recede, members narrow their affective expression, weakening the salience of ideas. Third, algorithmic vocabulary control increases the cognitive load: members engage in self-monitoring to avoid violations, which slows the exchanges of ideas. Together, these shifts reveal the paradox of algorithmic control: while it streamlines communication, it may reduce the novelty that emergent vocabularies cultivate.

I study the impact of algorithmic vocabulary control on novelty in online user innovation communities. Online user innovation communities, forums where users congregate, interact, and generate and diffuse numerous innovations (Shah & Nagle, 2019), offer a compelling setting to examine how vocabulary control shapes behavior and aggregates into organizational consequences (Jeppesen & Frederiksen, 2006; Piezunka & Dahlander, 2019). Because interactions essential to innovation occur almost entirely through written feedback, organizational vocabulary is the primary medium for co-creation (Park et al., 2023; Seo et al., 2021b). In the absence of visual or situational cues, members rely primarily on vocabulary to exchange ideas, convey emotions, and coordinate actions (Bregolin, 2024; Leonardi, 2014, 2018).

Empirically, I examine the staggered rollout of a profanity filter algorithm on Steam Workshop, an online user innovation community where user-innovators propose video game "mods" (modifications to an existing video-game product, e.g., cosmetics, maps, gameplay tweaks, story extensions) and receive peer feedback. In 2020, following criticism over toxic

language, the platform implemented an algorithm that automatically replaced swear words with heart-emoji strings. The adoption was phased across different sub-communities (June vs. October), providing a quasi-experimental setting. Using data on 7,145 updates and 61,175 feedback comments across 1,045 communities, difference-in-differences estimates show the filter reduced the novelty of updates by 49%[1]. Also consistent with my predictions, written feedback became simpler and more repetitive (37% increase in simpler vocabularies; 28% increase in repeated vocabularies), showed reduced emotionality (37.5% reduction), and arrived slowly (60% longer times between feedback). Causal mediation analyses show these mechanisms partially mediate the total effect on novelty.

This study makes four contributions. First, I extend research on organization design by advancing theory on the dynamics of organizational control (Cardinal et al., 2017; Joseph & Sengul, 2025). Many elements of design are in place without explicit planning, and even formally planned designs are often reshaped in practice as work unfolds through emergent activities such as established routines and tacit coordination (e.g., Chown, 2021; Anteby & Chan, 2018). Yet prior work has paid limited attention to emergent organizational vocabularies or to how they interact with planned design choices. Here, I theorize how a rule-based vocabulary control system interacts with an emergent vocabulary by altering participants' underlying cognitive structure and, in turn, shifting the vocabularies they use in feedback exchange.

Second, I contribute to research on algorithmic management. Organizations increasingly rely on algorithms to monitor behavior, enforce rules, and structure interaction (Allen & Choudhury, 2022; Kellogg et al., 2020; Kim et al., 2024). While such systems promise efficiency, they often generate unintended consequences (Burrell, 2016; Rahman, 2021). By

[1] In a post-hoc analysis, I also find that the introduction of algorithmic vocabulary control reduced the *frequency* of updates by 40%.

examining algorithmic vocabulary control, I show how these algorithms can narrow the richness of expression, suppress emotional nuance, and add cognitive load, therefore reducing the innovative output of the organizations. By focusing on vocabulary structure, I advance our understanding of this social process and not only show *whether* algorithmic control of an organization's vocabulary impacts organizational behavior, but also *how,* based on the mediating effects of reduced feedback richness, salience, and longer feedback delays.

Third, I contribute to research on vocabularies in organizations. Although vocabularies shape how information and attention are distributed (Eklund & Mannor, 2021; Joseph et al., 2023; Ocasio et al., 2018), their direct impact on organizational outcomes has been understudied and often treated as implicit. Vocabularies have typically served as methodological proxies for constructs such as institutional logics (Ocasio et al., 2015; Thornton & Ocasio, 1999), culture (Koçak & Puranam, 2024; Marchetti & Puranam, 2022), or mental models (Carley, 1997), rather than being theorized as constructs in their own right (Loewenstein & Ocasio, 2005). Leveraging the exogenous shock of the profanity filter, I disentangle vocabularies from related constructs and show how changes in vocabulary structure directly shape organizational outcomes.

Fourth, I contribute to the research on innovation in user communities (Fisher, 2019; Miller et al., 2009; Seo et al., 2021a). Whereas this stream of research has examined what drives exploratory or novel contributions, much of this has focused on the characteristics of the user-innovators (e.g., Autio et al., 2013; Burtch et al., 2022; Shah & Nagle, 2019) or the structure of the community (Dahlander & Frederiksen, 2012; Park et al., 2023; Piezunka & Dahlander, 2015), with far less focus on the feedback itself (see Seo et al., 2021 for an exception). My key findings highlight the importance of written feedback exchanged in online communities as the driving engine of the novelty these communities generate.

THEORETICAL DEVELOPMENT

Online User Communities and The Role of Organizational Vocabularies

An important source of innovation for organizations is online user communities (Baldwin & Von Hippel, 2011; Faraj et al., 2011; Fisher, 2019; O'mahony & Ferraro, 2007; Von Hippel & Von Krogh, 2003; Von Krogh & Von Hippel, 2006). Such communities rely on entrepreneurial contributors (i.e., user innovators) who advance new ideas and interact with other users who provide written feedback and make suggestions for changes to their ideas (Goldman & Gabriel, 2005; Shah & Nagle, 2019; Shah & Tripsas, 2007).

Communities constitute a social structure that exhibits organizing principles that differ from the traditional firm-based model (Lee & Cole, 2003; Safadi et al., 2021), in that knowledge is public, membership is open and voluntary, interests are shared, and feedback is targeted to improving ideas (Park et al., 2023; Piezunka & Dahlander, 2019). User innovators benefit from personal use of their ideas, recognition within the community, and, in some cases, financial gain (Shah & Tripsas, 2007; Von Hippel, 2005). Online user communities offer organizations access to distant knowledge at a lower cost (Afuah & Tucci, 2013; Dahlander & Piezunka, 2014; Jeppesen & Lakhani, 2010), while community members gain a sense of shared identity (Nagaraj & Piezunka, 2020; Ren et al., 2007; Zhang & Zhu, 2011) and free access to content (Dahlander & Frederiksen, 2012; Levine & Prietula, 2014).

In online user communities, user innovators and peer community members engage in an interactive process aimed at enhancing the novelty of ideas. Most of these communities provide a discussion forum for user innovators to generate and diffuse their innovations, and for their peers to interact and provide written feedback (Shah & Nagle, 2019). To gain community support for their ideas, user innovators provide descriptions of their ideas that focus on selected salient

features of their ideas (Cornelissen & Werner, 2014), and peer community members contribute written feedback.

Feedback providers (i.e., the peer community members) engage in efforts to improve the novelty and usefulness of an idea (Harrison & Rouse, 2015). Communication among users in these communities is organized around discussion "threads," where user innovators open new discussion threads to describe and showcase a new idea (Autio et al., 2013). The literature on user communities underscores its interactive, iterative, and consequential nature, wherein the viability of innovative work depends on exposing ideas to feedback from community members and adapting them accordingly (Hampel et al., 2020; Park et al., 2023).

The content of these interactions is shaped by organizational vocabularies. Following Loewenstein et al. (2012), I define organizational vocabularies as the shared system of words commonly used by a social collective. A Neo-Whorfian view (Loewenstein et al., 2012) holds that people are more likely to think in particular ways when a particular vocabulary is available. That is, vocabulary can shape *how* individuals think (Boroditsky, 2011) and *what* they are likely to think about (Gentner & Goldin-Meadow, 2003). This view aligns with research on the role of communication in attention dynamics (Ocasio et al., 2018), where vocabulary serves as a mechanism for directing attention. For instance, the presence of terms like "business models" not only directs focus to specific components of organizational strategy but also influences how those strategies are conceptualized and evaluated. Similarly, in online user communities, written feedback from community members that introduces a particular vocabulary can shape how user-innovators pursue the novelty of their innovations. Three features of organizational vocabularies in written feedback are particularly important in generating novel ideas in online user communities: First, organizational vocabularies enable richer feedback that supports the

divergent thinking behind creative performance. Second, organizational vocabularies sharpen the salience of feedback, quickly flagging problematic features with minimal text. Third, organizational vocabularies facilitate coordination by reducing collective interpretation costs and allowing timely feedback.

Algorithmic Vocabulary Control and Novelty in User Communities

While organizational vocabularies foster novelty, they can also create potential for communication problems (c.f., Carlie, 2002). For instance, vocabulary that allows coordination can become locally specialized within functions, making the translations across boundaries costly (Bechky, 2003). At the same time, vocabularies that help convey emotions and build in-group bonds can drift into a repertoire of insults toward out-groups, alienating others and rendering the communication environment toxic (Gillespie, 2018; Madhyastha et al., 2023).

In such contexts, organizations often institute vocabulary control. For instance, they may establish a common language (*lingua franca)* after a cross-national merger (Neeley et al., 2012; Kroon et al., 2015) or introduce a planned vocabulary to streamline and align agendas (Ocasio & Joseph, 2008). To address toxicity, digital platforms may introduce moderation filters that suppress the use of toxic vocabulary (Oh & Downey, 2025; Gillespie, 2018).

However, controlling vocabulary also introduces significant design and monitoring costs: codifying approved vocabularies and substitutions, training and recertifying members, auditing for noncompliance, and updating the vocabularies as work and rules evolve. To economize on these costs while achieving speed and uniformity, organizations increasingly deploy AI and algorithmic systems that detect, flag, substitute, mask, and pre-compose language. Examples include profanity filters used to block restricted terms in finance communications (e.g., Goldman

Sachs, 2011), standardized phrasing tools for clinical documentation (e.g., DeepScribeAI), auto-completion that nudges writers toward canonical wording (e.g., Gmail Smart Compose), and text-moderation filters that screen toxic or non-compliant vocabulary in platforms (e.g., YouTube, Facebook).

Research shows that using AI and algorithms for control often generates unintended consequences in organizations for certain tasks (Jia et al., 2024; Choi et al., 2025). Even when they yield objectively better decisions—some behaviors do not benefit as much as others—and can trigger underperformance, such as a decrease in member satisfaction, member disengagement, and underproductivity (Rahman, 2021; Allen & Choudhury, 2022; Jia et al., 2023). These unintended consequences may occur because the features of algorithmic control do not align with the features of the subjects it controls.

In particular, emergent activities that are built upon and evolved organically through frequent interactions, such as standard operating procedures or tacit coordination, may not benefit from algorithmic control (Jia et al., 2024). It is because algorithms are often rule-bound in ways that underweight context, and are opaque, offering limited rationale for their outputs (Burrell, 2016; Rahman, 2021). At the same time, organizational vocabularies are emergent. They develop through ongoing interactions and adapt over time to a particular organizational context (Loewenstein, 2017).

In what follows, I argue that when introduced in online user-innovation communities, algorithmic vocabulary control can reduce novelty through three mechanisms: (1) it narrows the richness of feedback by narrowing vocabulary diversity and complexity, limiting associative recombination; (2) it diminishes the salience of feedback by stripping affective cues that convey

stance and urgency; (3) it slows feedback cycles by increasing self-monitoring and cognitive load. Below, I discuss each of these and consider the direct and mediating impact on novelty.

Hypotheses Development

Organizational vocabulary, Richer feedback, and Algorithmic vocabulary control

Organizational vocabularies allow richer feedback (i.e., feedback with greater variety and complexity) (Segundo-Marcos et al., 2025; Kuznetsova et al., 2013) that supports the divergent thinking behind creative performance. As members continuously exchange feedback and adopt one another's expressions, they develop a richer set of common vocabulary (Loewenstein, 2017). Over time, this interaction produces a vocabulary structure in feedback (i.e., patterned word frequencies, word-to-word ties, and word-to-example relations) that is more complex and diverse.

For example, vocabulary structure allows members to make unexpected connections and novel recombination. Through continuous exchange, the thread of feedback develops word-to-word ties (i.e., recurring co-occurrence of different terms in written feedback that link concepts). As feedback providers constantly use different vocabulary together, they generate relationships between concepts that might otherwise remain separate. In communities with well-developed organizational vocabularies, these word-to-word ties may guide feedback receivers to make connections between seemingly distinct concepts. Such ties invite associative thinking, helping user innovators to make connections across different domains.

When vocabulary structure allows word-to-example relations (i.e., the systematic link between abstract terms to concrete instances), feedback receivers can easily make abstract ideas actionable across contexts. In communities with strong word-to-example relations, members can

navigate between multiple exemplars to show how the feedback has been applied. This allows user innovators to mix elements from different examples to produce novel solutions.

Implementing algorithmic vocabulary control, however, can undermine the richness of feedback by disrupting the existing vocabulary structure. Algorithmic vocabulary control narrows the range of expression, confining exchanges to a smaller set of terms in formulating feedback. Imposed constraints on feedback increase cognitive demands on users as they adapt to the new linguistic limitations, often leading to the use of simpler and repeated vocabulary. When familiar terms are no longer available, users must exert additional cognitive effort to reformulate their ideas using the remaining vocabulary. From an attentional processing perspective, this challenge reflects a problem of "searching for and processing relevant information when such searches are costly and decision makers are boundedly rational" (Lant & Shapira, 2000, 2001).

Restricting linguistic options heightens cognitive costs in identifying appropriate terms. As cognitive resources are finite, this increased demand can deplete the attentional resources available for other functions, such as effectively communicating and developing complex ideas. Correspondingly, studies on cognitive load (Sweller, 1988) and activity load (Castellaneta & Zollo, 2015) demonstrate how increased cognitive or activity load can strain an organization's attention capacity, adversely affecting decision-making performance in focal tasks. To manage this additional cognitive burden, users may simplify their language by using shorter sentences, straightforward syntax, and repeated vocabulary. This reduction in complexity allows individuals to preserve their limited cognitive resources while still regularly participating in the community.

Research on language and cognitive effort supports this tendency. Biber and Finegan (1991) found that as communicative constraints increase, individuals shift toward less cognitively demanding linguistic structures. Similarly, Alter and Oppenheimer (2009)

demonstrated that people are more likely to rely on simpler, easier-to-process information when under cognitive strain, highlighting how linguistic simplification is a natural response to increased mental demands that accompany the introduction of vocabulary constraints. For example, in a study of second-language writing, Elis and Yuan (2004) found that under strict time pressure with a minimum-length requirement, participants produced texts with lower syntactic complexity and smaller vocabularies. Similarly, Kormos (2000) demonstrated that English learners given strict guidelines (e.g., No notes allowed) wrote less fluently and used simpler syntax.

As a result, feedback receivers (i.e., user innovators) encounter feedback that is less complex and more repetitive; this shift steers them toward familiar, conventional solutions and dampens novelty.

Organizational vocabulary, Salience of feedback, and Algorithmic vocabulary control

Organizational vocabularies allow members to co-develop affective expressions, thus sharpening the salience of feedback. When members of the community share affective expressions, such as colloquialisms (e.g., “my bad”, “no worries”), emphatics (“legit”, “sick”), and mild profanity (“this is f**king great”), they can efficiently convey stance and emotion, quickly flagging outstanding features with minimal text (Fayard & DeSanctis, 2010; Huy, 2011; Jay & Janschewitz, 2008). By making key information salient and functioning as a “potent emotion-arousing symbol”, such organizational vocabularies channel collective attention to a particular feature (Huy, 2011; Ludwig et al., 2014). Emotional intensity in feedback significantly impacts attentional processing (Huy, 2012; Lerner et al., 2015; Scherer & Moors, 2019), as the

strength of emotions directly influences the extent to which they capture attention. The stronger the emotional intensity, the greater its effect on focus and engagement (Vuori, 2024).

This influence is particularly pronounced in contexts such as when evaluating the quality (or novelty) of ideas, where clear-cut assessments of the feedback providers are not readily apparent (Vuori & Huy, 2022). In such situations, emotional intensity in feedback can drive feedback receivers to engage deeply with a topic and persist in their efforts. When receiving feedback that uses more affective expression and thus is salient, user innovators may infer that the feedback provider has strong feelings and devote more attention and cognitive effort to understanding reactions and how the idea might be improved, experiment with alternative configurations, and yield exploratory revisions—i.e., novel recombination (Huy, 2011; Ludwig et al., 2014).

However, when algorithmic vocabulary control is implemented, it may reduce the salience of feedback by limiting the feedback provider's ability to articulate certain intense or emotionally charged ideas, thereby reducing attentional engagement of discussion (Vuori et al., 2018). For example, in a gaming community, where swear words might have been used to emphasize frustration with an idea or excitement over a new update, their absence may shift conversations to less emotional or more neutral topics. Without these emotionally charged linguistic tools, conversations may become more generic, as users default to universally accepted phrases (Fayard & DeSanctis, 2010; Huy, 2011; Jay & Janschewitz, 2008). For feedback receivers (i.e., user innovators), muted signals obscure urgency and importance, lowering the cognitive effort to respond to the feedback and revise the idea; decreasing the likelihood of user innovators proposing novel ideas (Huy, 2011; Ludwig et al., 2014).

Organizational vocabulary, Timely feedback, and Algorithmic vocabulary control

Organizational vocabularies facilitate coordination, reducing the burden of cognitive processing and allowing timely feedback (Okhuysen & Bechky, 2009). When feedback providers and receivers use familiar terms, they process information in a way that Levinthal and Rerup (2006) refer to as "less mindful information processing", who distinguish mindful (or controlled) information processing from less mindful (or automatic) information processing and argue that the latter involves more routinized behavior. That is, "when fewer cognitive processes are activated less often, the resulting state is one of mindlessness characterized by reliance on past categories, acting on 'automatic pilot,' and fixation on a single perspective without awareness that things could be otherwise." (Weick et al., 1999, p.90). With organizational vocabularies, members can spend less time explaining terms or clarifying misunderstandings (Bechky, 2003; Levinthal & Rerup, 2006; Weick et al., 1999).

When algorithmic vocabulary control is implemented, it will require members to be more thoughtful of their expression and cause members to become more conscious of the delivery of their messages (Weick & Sutcliffe, 2006), resulting in slowing feedback cycles (Levinthal & Rerup, 2006; Weick et al., 1999; Bechky, 2003; Okhuysen & Bechky, 2009). Especially given that algorithms are often opaque (Rahman, 2021), where the rules and rationale are hidden, unpredictable, and frequently changing, members must constantly regulate their expression in anticipation of how algorithms might interpret or penalize specific words. This mirrors Snyder's (1974) definition of self-monitoring as the extent to which people "observe and control how they present themselves in social situations". Imposed algorithmic vocabulary control, members must pre-scan wording, monitor borderline terms, and clarify potential misunderstandings (Burrell, 2006, 2016). These self-monitoring steps may lengthen the interval between feedback; as delays

accumulate, feedback–revision iterations within a given window decrease, narrowing opportunities for timely recombination and lowering novelty.

Further, when feedback is delayed, receivers cannot build real-time associative links, weakening coordination with the peer users (Faraj and Sproull, 2000). For instance, when feedback arrives later, earlier ideas may have already diverged, reducing opportunities for joint discovery. Additionally, without rapid reactions to resolve uncertainty about how others value work, creators may postpone bold changes and, opting for incremental changes, thereby dampening the risk-taking that underwrites novelty (Edmondson, 1999).

In sum, I offer the following baseline and mediating hypotheses:

Hypothesis 1 (H1 - Baseline). *Imposing an algorithmic vocabulary control decreases the novelty of ideas generated in online user communities.*

Hypothesis 2 (H2). *Imposing algorithmic vocabulary control reduces vocabulary complexity and vocabulary diversity in written feedback, thereby decreasing the novelty of ideas generated in online user communities.*

Hypothesis 3 (H3). *Imposing algorithmic vocabulary control reduces affective expression in written feedback, thereby decreasing the novelty of ideas generated in online user communities.*

Hypothesis 4 (H4). *Imposing an algorithmic vocabulary control increases delays between the exchange of feedback, thereby decreasing the novelty of ideas generated in online user communities.*

METHODS

Empirical Context: Steam and Steam Workshop

To test my hypotheses, I analyze field data on the introduction of an algorithmic vocabulary control (i.e., profanity filter) in Steam Workshop in 2020. Launched in September 2003, Steam is an online video gaming distribution platform developed by Valve Corporation. Often referred to as the "iTunes of Gaming," Steam has become a central platform for purchasing, downloading, and managing games. With over 120 million monthly active users and a library exceeding 50,000 games, it is the most profitable digital gaming platform, generating over $6 billion annually (Statista, 2023).

Steam launched Steam Workshop in 2011 to empower its user community by providing a dedicated space for user innovators to contribute user-generated content. Also known as the "mods", these contents are modifications to original games that allow user innovators to create new items, characters, or even entire storylines. For instance, in the online shooting game *Counter-Strike: Global Offensive (CS:GO)*, user innovators have created custom maps and weapon skins, significantly expanding the gameplay experience. Designed to "give players a direct pipeline to help shape their favorite games" (Steam, 2011), the Workshop fosters innovation and collaboration by facilitating the creation and sharing of mods. As of 2023, the Workshop hosts over 1.2 million unique mods created by more than 500,000 user innovators, covering a wide range of contributions and spanning over 1,000 games (Statista, 2023).

In the Steam Workshop, social interaction is crucial for fostering innovation. The platform enables user innovators and other community members to exchange feedback, rate submissions, and subscribe to updates, creating an environment where communication is critical for the community's vitality. Similar to other online user communities, such as Wikipedia or Reddit, the Steam Workshop relies heavily on user contributions and collective participation. However, the Steam Workshop introduces a more competitive dynamic that sets it apart. Contributors in the Workshop vie for attention, as visibility and community endorsement are critical to success. Creators actively compete to have their mods recognized by the community, making the novelty in their ideas especially important.

Algorithmic Vocabulary Control in Steam Workshop

Steam Workshop, like many online gaming-related platforms, has struggled with issues of toxic behavior and the corresponding use of profanity. This has become a significant social issue, with many newspaper articles publicly criticizing these behaviors (e.g., *The Guardian*, 2020; *Wired*, 2020). In this context, in June 2020, Steam introduced a vocabulary control algorithm—the "profanity filter"—to foster a safer and more inclusive environment for its community. The filter used an algorithm to automatically detect and replace offensive words in community chats with symbols (e.g., ♡♡♡♡). Violators were flagged and reported, and persistent offenders faced a lifetime ban from the community. Initially, Steam implemented this feature in two Valve-developed games, *Dota 2* and *CS:GO*, as an experimental rollout. By October 2020, the feature was extended to other games on the platform. According to Steam's official statement, the filter aimed to "address toxicity and create a welcoming space for all

players" (Steam, 2020). The implementation was abrupt and lacked advance communication, catching many community members by surprise.

The implementation of the profanity filter creates a unique context for examining how algorithmic vocabulary control influences activities within user communities. Additionally, the phased implementation of this feature, starting with Valve's own games and later expanding to others, offers valuable opportunities for comparative analysis. FIGURE 1 illustrates the implementation timeline.

As such, I treat this context as a quasi-natural experiment, with the implementation of the profanity filter serving as the treatment and games without the filter as the control group. The staggered rollout allows for comparisons between treated and untreated communities, providing insights into how algorithmic vocabulary control affects user interactions and novelty.

INSERT FIGURE 1 ABOUT HERE

Data and Sample

The comprehensive dataset includes feedback comments, update notes, and community characteristics for both treatment games (CS:GO and DOTA 2) and control games (Elder Scrolls V: Skyrim and Garry's Mod) spanning the period from February 2020 (t-4) to September (t+3)[2]. Two games, Elder Scrolls V: Skyrim and Garry's Mod, were chosen as ideal control groups due to their similarities with the treatment games in terms of genre, revenue, and community size. Each mod within a game corresponds to a distinct community, which I consider the basic unit of observation for the analysis. During the study period, a total of 1,074 communities were active.

[2] I broaden the study window to t-12 to t+12 to examine the long-term implication in the post-hoc analyses.

I collected and structured two datasets for the analysis. First, to assess the change in novelty over time, I analyzed "change notes" posted by mod creators. These are written descriptions detailing updates to mods – for example, fixing bugs, introducing new characters, or adding new storylines. In total, I collected 7,115 of these change notes across all the mods, aggregated at the monthly level. Second, to test the proposed mechanisms, I gathered 61,175 written feedback posts from the members of the Workshop, also aggregated at the monthly level.

Empirical Strategy

To examine the effect of a vocabulary control on novelty in community-driven innovations, I employ both the traditional Two-Way Fixed-Effects DiD (TWFE) and Staggered Difference-in-differences (DiD) approach (CS21) (Angrist & Pischke, 2009; Callaway & Sant'Anna, 2021; Kang & Eklund, 2024). The rollout of the profanity filter, which began in June 2020, serves as the treatment point for the initial set of communities. Other communities on the platform that had not yet been treated serve as the control group.

Measures

Preprocessing Text Data

In the study, I employed a series of standard preprocessing techniques (see Jurafsky & Martin, 2014; Manning et al., 2008, for an overview) to address the complexity of textual data while ensuring the substantive content was preserved for analysis (Denny & Spirling, 2018). First, I parsed the text into individual tokens, such as words, to facilitate structured analysis, reducing the average document length by 49.1%. Non-alphabetical expressions, including numbers, punctuation, emojis, and special characters (e.g., #, %, &, $), were removed to simplify

the corpus, with punctuation entirely eliminated. Then, I lowercased (decapitalized) the words to ensure uniformity and prevent inconsistencies, such as treating "Game" and "game" as distinct entities. Next, I removed stopwords (common words like "the," "is," and "and"), which primarily serve grammatical functions and add little meaning to the analysis. I then applied lemmatization (Manning, 2008), a technique that reduces words to their base or dictionary form (e.g., "playing," "played," and "plays" were all reduced to "play"), which minimized vocabulary variation and reduced the vocabulary size by 67.9%. Finally, domain-specific elements that could distort the analysis, such as URLs, usernames, and other irrelevant text, were excluded.

Dependent Variable: Novelty (H1)

I use the extent of *Cosine Distance* observed in the change notes as a proxy for the user innovator's novelty. Following prior studies that have measured the novelty of ideas using text-mining methods (e.g., Park et al., 2023; Angus, 2019; Schweisfurth et al., 2023; Miric et al., 2023; Ploog & Rietveld, 2024), I employ cosine distance to quantify the difference between numeric representations of textual content in vector space.

I calculated the cosine distance between verbal descriptions of updates to modifications or "change notes". Specifically, the distance between verbal descriptions *i* and *j* was measured as $d_{ij} = 1 - \frac{\overrightarrow{v_i} \cdot \overrightarrow{v_j}}{\|\overrightarrow{v_i}\| \|\overrightarrow{v_j}\|}$, where $\overrightarrow{v_i}$ and $\overrightarrow{v_j}$ are vectorized verbal descriptions, respectively. When two verbal descriptions of "change notes" are identical in terms of vocabulary used, d_{ij} will become 0. As the descriptions become more distinct, d_{ij} increases. To construct these vector representations, I use both of the two widely adopted methods: Term Frequency-Inverse Document Frequency (TF-IDF) and Bidirectional Encoder Representations from Transformers (BERT). Additionally, for the cosine dissimilarity measure, defining the reference set is critical

(Burtch et al., 2022). I use both alternative reference sets to assess the user innovator's novelty in contributions: An update may be considered novel relative to the same entrepreneur's initial update (*Cosine Dissimilarity (First)*) or relative to the most recent update (*Cosine Dissimilarity (Recent)*)[3].

Mediating Variables:

***Vocabulary Diversity and Vocabulary Complexity in Written Feedback* (H2)**

To test H2, I construct two variables: First, to measure the complexity of feedback, I use the Gunning-Fog index (*Fog Index*). Following many of the strategy and management literature (Callery & Perkins, 2021; Carton et al., 2014; Datar et al., 2024; Fabrizio & Kim, 2019; Guo et al., 2020, 2021; König et al., 2018), I operationalize *Fog Index* using the Gunning-Fog index of the feedback comments. The Gunning-Fog index is a standard linguistic measure used in the literature to analyze the complexity of a document (Lehavy et al., 2011). This index analyzes the readability of a text by measuring two components: the length of sentences and the proportion of complex words. Complex words are defined as words having more than three syllables. The formula is the following:

$$fog = 0.4\left(\frac{words}{sentences} + 100 \times \frac{complex\ words}{words}\right)$$

The output of this function is a positive number. The large number of these variables suggests that the vocabulary is more complex.

[3] I used the first reference set (Cosine Dissimilarity (Recent)) for the main analysis. As a robustness check, I executed the analysis using the second reference set (Cosine Dissimilarity (First)), and the results remained consistent.

Second, I quantified the *Number of Unique Words* per sentence used in each feedback post to measure the extent of vocabulary variety in feedback. This metric captures the diversity of vocabulary employed by users when providing feedback, aligning with prior research that highlights the role of unique vocabularies in analyzing textual diversity (Allen & Choudhury, 2022; Criscuolo et al., 2017; Piezunka & Dahlander, 2015). To compute this, I utilized the LIWC (Linguistic Inquiry and Word Count) software, which analyzes text on multiple dimensions, including linguistic and psychological constructs. Specifically, I leveraged the "WC (Unique Word Count)" feature of LIWC to extract the total number of words used in each feedback post. Subsequently, I identified and isolated the unique words—those that appear only once within a sentence—and calculated the average number of unique words per sentence for each post (Angus, 2019; Devarakonda et al., 2018).

Emotionality of community feedback (H3)

To test H3, which predicts the relationship between algorithmic vocabulary control and reduced affective expression, I operationalize the *emotionality of community feedback.* I use the Natural Language Toolkit (NLTK), a widely used Python library for natural language processing (Choudhury et al., 2019; Momtaz, 2021). Specifically, I rely on its sentiment analysis resources to assign each feedback post an emotionality score. These scores capture both the valence (positive vs. negative tone) and the strength of the expressed emotion. For example, a comment such as “This idea is brilliant and incredibly inspiring!” would receive a strong positive score, indicating high emotionality in a positive direction. By contrast, “This suggestion is frustrating and doesn’t make sense” would be classified as strongly negative, while a neutral remark like “Thanks for sharing your idea” would receive a low emotionality score.

Delays in community feedback (H4)

To test H4, which predicts the relationship between algorithmic vocabulary control and delayed feedback, I operationalize delays in community feedback as the *elapsed number of hours* between sequential feedback posts within a discussion thread. For instance, if a user posts design feedback at 10:00 a.m. and a different user posts another feedback at 4:00 p.m., the delay is measured as six hours. Longer elapsed times reflect greater delays in feedback, suggesting slower responsiveness, whereas shorter intervals indicate more immediate interaction.

Independent Variables: Treatment × Post

To employ the difference-in-difference (DiD) approach to test the hypotheses, I first define an indicator variable *Treatment* that takes a value of 1 for workshops in the treatment group (communities in Dota 2 and CS:GO, which experienced the vocabulary control in June 2020) and 0 for the control group (communities in Elder Scrolls V: Skyrim and Garry's Mod). I then define an indicator variable *Post* that takes a value of 1 when a workshop experienced the vocabulary control and 0 otherwise. Finally, I create an interaction variable *Treatment* × *Post*, which multiplies *Treatment* and the *Post* variables; the coefficient of this variable represents the effect of the vocabulary control on both the mediating variables and the dependent variable.

Control Variables

I control for a variety of factors that may impact the user innovator's novelty. First, I control for *User Retention*, which measures the proportion of users who continue to engage with the platform over time. High user retention indicates a stable and engaged user base, which can

influence the generation of novel ideas by maintaining a consistent pool of contributors. Second, I include a control for the *Community Size*, representing the total number of active users in the community at a given point in time. Larger communities may facilitate a greater diversity of ideas and perspectives, potentially influencing the user innovator's novelty. Third, I control for *Community Age*, which reflects the duration (in years) since the community was first established. Older communities may have more established norms and shared knowledge, which could either constrain or enable the emergence of novel contributions. Time (year, month) fixed effects, game fixed effects, and creator fixed effects are added by including dummies (Allison & Waterman, 2002). These controls account for unobserved heterogeneity that might otherwise bias the estimates.

Empirical Specification

To study the impact of vocabulary control on the user innovator's novelty, I utilize the following Difference-in-Differences (DiD) model:

$$Y_{it} = \alpha_0 + \alpha_1 Treatment_i + \alpha_2 Post_t + \beta Treatment_i \times Post_t + \sigma_i + \gamma_i + \mu_t + \epsilon_{it}$$

Y_{it} captures the changes in the user innovator's novelty in a community i in a year t. $Treatment_i$ is a dummy variable equal to 1 if the community i is part of the treated group (those communities affected by the profanity filter) and 0 otherwise. $Post_t$ is a dummy variable equal to 1 if the observation year is after the introduction of the profanity filter in June 2020 and 0 otherwise. σ_i denotes creator-fixed effects, which control for any unobserved characteristics specific to each creator that might influence innovation novelty consistently across time. γ_i represents game-fixed effects, controlling for factors specific to each game that could affect the observed outcomes, such as the game's complexity or its general appeal to creators.

μ_t represents year- and month-fixed effects, accounting for any unobserved time trends or external factors affecting all games and creators, each month, such as changes in overall platform usage or general industry trends.

RESULTS

Table 1 reports the summary statistics and the pair-wise correlation. The positive correlation between the *Fog Index* and the *Number of Unique Words* suggests that feedback with greater vocabulary variety tends to be more complex, reflecting a connection between vocabulary diversity and complexity. *Community Size* and *User Retention* are positively correlated, indicating that larger communities are better able to sustain user engagement over time.

INSERT TABLE 1 ABOUT HERE

I first conducted a TWFE analysis and staggered difference-in-differences (CS21) analysis[4] to assess the relationship between vocabulary control and the novelty of community-driven innovation. TABLE 2 (TWFE) reports the results, in which the dependent variable is novelty. Additionally, in TABLE 2A, I measure cosine distance using different benchmarks and methods. Column (1) uses the most recent update (TF-IDF), Column (2) uses the first update (TF-IDF), Column (3) uses the most recent update (BERT), and Column (4) uses the first update (BERT).

INSERT TABLE 2 and TABLE 2A ABOUT HERE

The estimates for Treat × Post range from –0.059 to –0.034. These differences correspond to an average monthly decline in novelty of approximately 24.6% to 43% per

[4] Alternatively, I conducted a TWFE analysis. Results are reported in the appendix.

community over the three months following the treatment. FIGURE 2 graphically illustrates the dynamic effects alongside the pretreatment trend.

INSERT FIGURE 2 ABOUT HERE

Test of Mediating Hypotheses

I test the three mechanisms (*decreased richness of feedback, decreased salience of feedback, and increased delay of feedback)*, and causal mediation analyses using bootstrapping methods confirm the mediation.

Decreased Richness of Feedback

In Table 3, I assess the impact of the algorithmic vocabulary control on decreases in the richness of feedback. Supporting H2, Model 1 shows that the Treatment (*Treatment* × *Post*) significantly decreases the *Fog Index* ($\beta = -0.141$, $p = 0.057$), indicating that feedback becomes less complex. In Model 2, the *Number of Unique Words* decreases significantly ($\beta = -0.253$, $p = 0.031$), suggesting that feedback becomes less varied.

INSERT TABLE 3 ABOUT HERE

To illustrate the effects of algorithmic vocabulary control on the complexity of feedback, Figures 3 and 4 provide a graphical representation of the results. Figure 3 shows the changes in vocabulary complexity over time, comparing the treatment and control groups before and after the introduction of the profanity filter feature. The treatment group experienced a notable decline in the vocabulary complexity post-treatment, indicating reduced complexity in feedback.

Figure 4 further examines the treatment effect on the vocabulary diversity, displaying relative changes across periods before and after the algorithmic vocabulary control. As shown in Figure 4, the treatment experienced a notable decline in vocabulary diversity post-treatment.

INSERT FIGURES 3 AND 4 ABOUT HERE

Decreased Salience of Feedback (H3) and Increased Delays of Feedback (H4)

Table 4 presents the two-way fixed effects (TWFE) OLS regression estimates of the treatment's effect on the salience of feedback and delays in feedback, using an event window of $[-4,3]$. Consistent with H3, the results in Model 1 indicate that algorithmic vocabulary control significantly reduced the emotionality of written feedback ($\beta = -0.090$, $p < 0.10$), suggesting a decline in affective expressions in written feedback that typically signal emphasis and engagement. Additionally, Consistent with H4, the time between feedback events increased markedly ($\beta = 15.198$, $p < 0.05$), implying slower feedback exchanges after implementation. Together, these results suggest that algorithmic vocabulary control lowered the salience of feedback and increased the delays in feedback.

INSERT TABLE 4 ABOUT HERE

Causal Mediation Analysis

To test for causal mediation, I conducted mediation analyses using bootstrapping methods (e.g., Aguinis et al., 2017; Tuggle et al., 2024). Tables 5–6 report the results for H2, which predicted that *Fog Index* and *Number of Unique Words* mediate the relationship between Treatment × Post and the novelty of ideas.

INSERT TABLES 5 AND 6 ABOUT HERE

The bootstrapping-mediated regression analysis results show significant indirect effects for both linguistic mediators. In Table 5, for *Fog Index*, the indirect effect is significant (B = –0.0008, $p < 0.001$), mediating approximately 0.7% of the total effect (B = –0.1207, $p < 0.001$). In Table 6, for *Number of Unique Words*, the indirect effect is also significant (B = –0.0044, $p < 0.001$), mediating around 3.6% of the total effect (B = –0.1211, $p < 0.001$). Across both models, the direct effects of *Treatment* × *Post* on novelty remain negative and statistically significant (*Fog Index:* B = –0.1215, $p < 0.001$; *Number of Unique Words:* B = –0.1255, $p < 0.001$), indicating partial mediation. These results are robust to the Sobel test (Tuggle et al., 2024). Overall, Tables 5 and 6 suggest that algorithmic vocabulary control partially reduces idea novelty through its negative effects on both vocabulary complexity and vocabulary diversity.

Tables 7–8 report the results that examine whether *Emotionality* (H3) and *Delays in community feedback* (H4) mediate the relationship between *Treatment* × *Post* and *novelty in user innovators' ideas*.

--Insert Tables 7 and 8 About Here—

Table 7 shows that the indirect effect via *Emotionality* is negative but is not statistically significant (B = –0.0067). The direct effect remains large and negative (B = –0.0921, $p < 0.001$), and the total effect (B = –0.0854, $p < 0.001$), however, indicates that *Emotionality* does not significantly mediate the treatment's impact on novelty.

By contrast, Table 8 shows a strong and significant indirect effect through *Delays in feedback* (B = –0.0277, $p < 0.001$), which accounts for about 32.7% of the total effect. The direct effect of *Treatment* × *Post* remains negative and significant (B = –0.1122, $p < 0.001$), and the total effect is also negative (B = –0.0845, $p < 0.001$). This indicates that algorithmic vocabulary control reduces idea novelty partly by slowing the pace of feedback exchange in the community.

Across all models, the findings support partial mediation: 0.7% of the total effect is explained by decreased vocabulary complexity, 3.6% by reduced vocabulary diversity, and 32.7% by delays in feedback, while emotionality shows no significant mediating effect.

POST HOC ANALYSES

To complement the main findings, I conducted several post-hoc analyses.

First, I test an alternative dependent variable—*frequency of updates*. The analysis shows that vocabulary control not only reduces the novelty of contributions but also dampens the overall pace of activity, leading to fewer updates over time.

Second, I examine boundary conditions. I find that the decline in novelty is concentrated in *younger communities*, in groups that *swore less before* the intervention, and in communities with a majority of *non-native speakers*. By contrast, older, high-swear, and native-majority communities appear more resilient, suggesting that restrictions are especially significant where repertoires are less established.

Third, I investigate long-term effects. While some communicative dimensions (e.g., use of complex words and unique words) show partial recovery, novelty and update activity remain persistently suppressed, with no return to pre-treatment levels. This indicates that the negative effect is enduring, even as other aspects of communication may rebound.

Fourth, I explore adaptation strategies. Over time, users developed creative workarounds through misspellings, inventive variants, and replacement words. At the same time, the decrease in profanity led to a decline in the words that had typically co-occurred with it—such as *fix, map, game, work,* and *update*. This shift suggests that communication content became less substantive and more detached from concrete, task-focused discussions.

Fifth, to address the alternative explanation that the observed effects may reflect changes in the composition of users rather than behavioral change, I compare the member retention rate pre- and post-treatment.

DISCUSSION

This study examines the impact of vocabulary control on novelty in online user communities. In particular, I examine how a shift in user community vocabulary (i.e., through the imposition of vocabulary control) affects how user innovator members respond to community feedback and express novelty in the description of their contributions. Drawing on theories of attention dynamics and vocabularies and by analyzing data from Steam Workshop, I found that the introduction of a vocabulary control algorithm (profanity filter) disrupted shared vocabularies, reducing both complexity and variety. These changes negatively impacted the novelty of user innovators' contributions, highlighting how constraints on community vocabularies shape co-creation dynamics. The mediation analysis confirmed that the decline in vocabulary complexity and variety mediated the relationship between vocabulary control and the reduction in idea novelty.

The dynamic ABV highlights the role of social interactions and communicative practices as factors influencing attentional processing (Ocasio et al., 2018; Vuori, 2024). I argue that the introduction of vocabulary control significantly influences user communities by shaping attentional processing through both cognition and emotion. Specifically, restricting key vocabulary words compels users to simplify and limit the variety of their feedback. The additional cognitive effort required to adapt to these controls leads to a focus on maintaining communication efficiency, further encouraging linguistic simplification. The constraint on

emotionally laden words (i.e., profanity) reduces emotional intensity in conversations, ultimately reducing attentional engagement and fostering more uniform and less diverse discussions.

I make several contributions to the literature. First, the work contributes to research on organizational design by exploring the implications of vocabulary control for innovation outcomes. Existing research (see Cardinal et al., 2017, for a recent review) finds that organizational control significantly impacts important organizational outcomes such as employee satisfaction (Gregory et al., 2013), effective information flows (Gowen III et al., 2006; Lin & Germain, 2003), and team performance (Kreutzer et al., 2015). However, studies on how control impacts innovation outcomes are limited. Of the 108 organizational control studies Cardinal et al., (2017) reviewed, only 15 studies have included adaptability (e.g., innovation) outcomes, the lowest level of inclusion among the four types (adaptability, human relations, process, rational goal) of outcomes. Even within the handful of studies, they offer mixed findings. Some show positive outcomes (Cardinal, 2001; Goodale et al., 2011), while others find that controls may hinder innovation (Ogbonna & Wilkinson, 2003).

The mixed findings may exist because existing studies heavily emphasize output and input controls, focusing less on how organizations use behavior controls on innovation (Cardinal et al., 2017). Most studies assess innovation outcomes at the end of a project (i.e., output controls), such as the number of new drugs (Cardinal, 2001), or the input of an R&D project (i.e., input controls), including scientists' knowledge diversity and professionalization (Cardinal, 2001), without delving much into how controls guide members' ongoing activities during the innovation process. This study introduces vocabulary control as a unique form of behavioral control that directly influences cognitive and linguistic processes within an organization. While controls are typically intended to enhance efficiency or maintain decorum (e.g., Neeley, 2017;

Ogbonna & Wilkinson, 2003), the findings suggest that such measures can inadvertently stifle creativity by reducing the cognitive resources and expressive tools available for innovation. This highlights the need for organizations to carefully consider the unintended consequences of vocabulary control mechanisms, particularly in innovation-driven settings.

This study also advances understanding within the user community literature, particularly in how feedback dynamics shape innovation. While previous research has examined how community or contributor characteristics influence user-generated contributions (e.g., Dahlander & Frederiksen, 2012; Autio et al., 2013; Burtch et al., 2022; Shah & Nagle, 2019; Park et al., 2023), the findings underscore the critical role of feedback content in fostering novelty. Specifically, I show that constraints on community vocabulary reduce the cognitive and emotional richness of feedback, hindering the co-creation process between user innovators and their peers. These findings complement prior work by Fisher (2019) and Seo et al. (2021), emphasizing that feedback complexity and variety are central to sustaining innovative output in online communities.

By using the dynamic ABV lens (Ocasio et al., 2018; Vuori, 2024) I offer a new understanding of the underlying mechanisms that shape the co-creation interactions between user innovators and the user community. The mediation analysis uncovers new sources of empirical variation in the relationship between community vocabularies and user innovator novelty stemming from the cognitive and emotional effects on attentional processing. This augmentation to the theory suggests that future research may want to better understand these mediation effects and recognize the important role of vocabulary in directing attention (Loewenstein et al., 2012).

It is also worth noting that the use of the exogenous shock in the Steam community and DID approach helps better establish the causal effects of vocabulary choices on entrepreneurial

framing. Although this relationship is posited in much of the organizational literature on linguistics (Clarke & Cornelissen, 2011; Mantere, 2013), the claims of causality have not been fully established. This is significant because it calls into question whether vocabulary choice is better interpreted as an antecedent or an outcome of behavior. Most studies conflate the two because it's difficult to separate a vocabulary change from a change in cognition and practices. Many studies have attempted to demonstrate this relationship using verbal theory (Clarke & Cornelissen, 2011; Lounsbury & Glynn, 2001), content analysis (Allison et al., 2015; Allison et al., 2013), case studies (Ansari et al., 2016; Weber, 2008), or other econometric approaches to deal with determining causal relationships, but, to my knowledge, this has not been tested in a quasi-natural experiment that deals more fully with bias.

While the study provides valuable insights, it is not without limitations. First, the profanity filter represents a specific type of vocabulary control, and its effects may not generalize to other forms of vocabulary control, such as shifts in organizational jargon or mandated linguistic policies (e.g., English speaking only). Second, the focus on an online gaming platform limits the generalizability of the findings to other contexts, such as corporate settings or offline communities. Future research could explore how different types of vocabulary control influence innovation across various organizational and cultural settings. Additionally, while I examined vocabulary complexity and variety as mediators, other factors, such as perceived legitimacy, may also play a role and warrant further investigation.

CONCLUSION

This study demonstrates the impact of vocabulary control on the co-creation dynamics within online user communities. By showing how disruptions to shared vocabularies reduce the

complexity and variety of feedback, I highlight the critical role of vocabulary control in fostering user innovation. These findings offer important implications for the design of platforms and organizations more generally, suggesting that while vocabulary control mechanisms may enhance civility, they can also inhibit the richness of interactions that drive creative outcomes. Future research should continue to examine the delicate balance between vocabulary moderation and innovation in organizational and community contexts.

REFERENCES

Afuah, A., & Tucci, C. L. (2013). Value capture and crowdsourcing. *Academy of management Review*, *38*(3), 457-460.

Aguinis, H., Edwards, J. R., & Bradley, K. J. (2017). Improving our understanding of moderation and mediation in strategic management research. *Organizational research methods*, *20*(4), 665-685.

Allen, R., & Choudhury, P. (2022). Algorithm-augmented work and domain experience: The countervailing forces of ability and aversion. *Organization Science*, *33*(1), 149-169.

Allison, P. D., & Waterman, R. P. (2002). 7. Fixed-effects negative binomial regression models. *Sociological methodology*, *32*(1), 247-265.

Allison, T. H., Davis, B. C., Short, J. C., & Webb, J. W. (2015). Crowdfunding in a prosocial microlending environment: Examining the role of intrinsic versus extrinsic cues. *Entrepreneurship Theory and Practice*, *39*(1), 53-73.

Allison, T. H., McKenny, A. F., & Short, J. C. (2013). The effect of entrepreneurial rhetoric on microlending investment: An examination of the warm-glow effect. *Journal of Business Venturing*, *28*(6), 690-707.

Alter, A. L., & Oppenheimer, D. M. (2009). Suppressing secrecy through metacognitive ease: Cognitive fluency encourages self-disclosure. *Psychological science*, *20*(11), 1414-1420.

Amabile, T. M. (1997). Entrepreneurial creativity through motivational synergy. *The journal of creative behavior*, *31*(1), 18-26.

Angus, D. (2019). Recurrence methods for communication data, reflecting on 20 years of progress. *Frontiers in Applied Mathematics and Statistics*, *5*, 54.

Ansari, S., Garud, R., & Kumaraswamy, A. (2016). The disruptor's dilemma: TiVo and the US television ecosystem. *Strategic management journal*, *37*(9), 1829-1853.

Autio, E., Dahlander, L., & Frederiksen, L. (2013). Information exposure, opportunity evaluation, and entrepreneurial action: An investigation of an online user community. *Academy of Management Journal*, *56*(5), 1348-1371.

Baldwin, C., & Von Hippel, E. (2011). Modeling a paradigm shift: From producer innovation to user and open collaborative innovation. *Organization Science*, *22*(6), 1399-1417.

Bammens, Y., & Collewaert, V. (2014). Trust between entrepreneurs and angel investors: Exploring positive and negative implications for venture performance assessments. *Journal of management*, *40*(7), 1980-2008.

Biber, D. (1988). *Variation across Speech and Writing*. Cambridge University Press.

Biber, D., & Finegan, E. (1991). *Multi-dimensional Analyses of Author's style: Some Case Studies from the 18th Century*.

Blakemore, D. (2011). On the descriptive ineffability of expressive meaning. *Journal of Pragmatics*, *43*(14), 3537-3550.

Boroditsky, L. (2011). How language shapes thought. *Scientific American*, *304*(2), 62-65.

Burtch, G., He, Q., Hong, Y., & Lee, D. (2022). How do peer awards motivate creative content? Experimental evidence from Reddit. *Management Science*, *68*(5), 3488-3506.

Callery, P. J., & Perkins, J. (2021). Detecting false accounts in intermediated voluntary disclosure. *Academy of Management Discoveries*, *7*(1), 40-56.

Cardinal, L. B. (2001). Technological innovation in the pharmaceutical industry: The use of organizational control in managing research and development. *Organization Science*, *12*(1), 19-36.

Cardinal, L. B., Kreutzer, M., & Miller, C. C. (2017). An aspirational view of organizational control research: Re-invigorating empirical work to better meet the challenges of 21st century organizations. *Academy of Management Annals*, *11*(2), 559-592.

Carton, A. M., Murphy, C., & Clark, J. R. (2014). A (blurry) vision of the future: How leader rhetoric about ultimate goals influences performance. *Academy of Management Journal*, *57*(6), 1544-1570.

Castellaneta, F., & Zollo, M. (2015). The dimensions of experiential learning in the management of activity load. *Organization Science*, *26*(1), 140-157.

Cattani, G., Deichmann, D., & Ferriani, S. (2022). *The generation, recognition and legitimation of novelty*. Emerald Publishing Limited.

Cattani, G., Deichmann, D., Ferriani, S., & Snihur, Y. (2024). Framing Novelty: A Linguistic Approach to the Understanding of Entrepreneurship, Creativity, and Innovation. *Strategic entrepreneurship journal*.

Clarke, J., & Cornelissen, J. (2011). Language, communication, and socially situated cognition in entrepreneurship. *Academy of management Review*, *36*(4), 776-778.

Clarke, J. S., Cornelissen, J. P., & Healey, M. P. (2019). Actions speak louder than words: How figurative language and gesturing in entrepreneurial pitches influences investment judgments. *Academy of Management Journal*, *62*(2), 335-360.

Cornelissen, J. P., Clarke, J. S., & Cienki, A. (2012). Sensegiving in entrepreneurial contexts: The use of metaphors in speech and gesture to gain and sustain support for novel business ventures. *International small business journal*, *30*(3), 213-241.

Cornelissen, J. P., Holt, R., & Zundel, M. (2011). The role of analogy and metaphor in the framing and legitimization of strategic change. *Organization Studies*, *32*(12), 1701-1716.

Cornelissen, J. P., & Werner, M. D. (2014). Putting framing in perspective: A review of framing and frame analysis across the management and organizational literature. *Academy of Management Annals*, *8*(1), 181-235.

Crilly, D., Hansen, M., & Zollo, M. (2016). The grammar of decoupling: A cognitive-linguistic perspective on firms' sustainability claims and stakeholders' interpretation. *Academy of Management Journal*, *59*(2), 705-729.

Criscuolo, P., Dahlander, L., Grohsjean, T., & Salter, A. (2017). Evaluating novelty: The role of panels in the selection of R&D projects. *Academy of Management Journal*, *60*(2), 433-460.

Cutolo, D., Ferriani, S., & Cattani, G. (2020). Tell me your story and I will tell your sales: A topic model analysis of narrative style and firm performance on Etsy. In *Aesthetics and Style in Strategy* (pp. 119-138). Emerald Publishing Limited.

Dahlander, L., & Frederiksen, L. (2012). The core and cosmopolitans: A relational view of innovation in user communities. *Organization Science*, *23*(4), 988-1007.

Dahlander, L., & Piezunka, H. (2014). Open to suggestions: How organizations elicit suggestions through proactive and reactive attention. *Research policy*, *43*(5), 812-827.

Datar, A., Amore, M. D., & Fosfuri, A. (2024). Strategic patent disclosure: Unraveling the influence of temporal preferences. *Strategic Organization*, 14761270241299756.

De Stobbeleir, K. E., Ashford, S. J., & Buyens, D. (2011). Self-regulation of creativity at work: The role of feedback-seeking behavior in creative performance. *Academy of Management Journal*, *54*(4), 811-831.

Denny, M. J., & Spirling, A. (2018). Text preprocessing for unsupervised learning: Why it matters, when it misleads, and what to do about it. *Political analysis*, *26*(2), 168-189.

Devarakonda, S. V., McCann, B. T., & Reuer, J. J. (2018). Marshallian forces and governance externalities: Location effects on contractual safeguards in research and development alliances. *Organization Science*, *29*(6), 1112-1129.

Durand, R., & Huysentruyt, M. (2022). Communication frames and beneficiary engagement in corporate social initiatives: Evidence from a randomized controlled trial in France. *Strategic management journal*, *43*(9), 1823-1853.

Fabrizio, K. R., & Kim, E.-H. (2019). Reluctant disclosure and transparency: Evidence from environmental disclosures. *Organization Science*, *30*(6), 1207-1231.

Faraj, S., Jarvenpaa, S. L., & Majchrzak, A. (2011). Knowledge collaboration in online communities. *Organization Science*, *22*(5), 1224-1239.

Fayard, A. L., & DeSanctis, G. (2010). Enacting language games: the development of a sense of 'we-ness' in online forums. *Information Systems Journal*, *20*(4), 383-416.

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. *Administrative Science Quarterly*, *48*(1), 94-118.

Fischer, E., & Reuber, A. R. (2014). Online entrepreneurial communication: Mitigating uncertainty and increasing differentiation via Twitter. *Journal of Business Venturing*, *29*(4), 565-583.

Fisher, G. (2019). Online communities and firm advantages. *Academy of management Review*, *44*(2), 279-298.

Foolen, A. (2015). Word valence and its effects. In *Emotion in language: Theory–research–application* (pp. 241-256). John Benjamins Publishing Company.

Forbes. (2020). Lessons learned from executing successful virtual hackathons. *Forbes*. https://www.forbes.com/councils/forbestechcouncil/2020/06/11/lessons-learned-from-executing-successful-virtual-hackathons/

Forbes. (2023). How Hackathons Are Reshaping Our Society. *Forbes*.

Gafni, H., Marom, D., & Sade, O. (2019). Are the life and death of an early-stage venture indeed in the power of the tongue? Lessons from online crowdfunding pitches. *Strategic entrepreneurship journal*, *13*(1), 3-23.

Gentner, D., & Goldin-Meadow, S. (2003). Language in mind. In: Cambridge, MA: MIT Press.

Gillespie, T. (2018). *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

Goldman, R., & Gabriel, R. P. (2005). *Innovation happens elsewhere: Open source as business strategy*. Morgan Kaufmann.

Goodale, J. C., Kuratko, D. F., Hornsby, J. S., & Covin, J. G. (2011). Operations management and corporate entrepreneurship: The moderating effect of operations control on the antecedents of corporate entrepreneurial activity in relation to innovation performance. *Journal of operations management*, *29*(1-2), 116-127.

Goodman, J. S., Wood, R. E., & Hendrickx, M. (2004). Feedback specificity, exploration, and learning. *Journal of Applied Psychology*, *89*(2), 248.

Gowen III, C. R., Mcfadden, K. L., Hoobler, J. M., & Tallon, W. J. (2006). Exploring the efficacy of healthcare quality practices, employee commitment, and employee control. *Journal of operations management*, *24*(6), 765-778.

Gregory, R. W., Beck, R., & Keil, M. (2013). Control balancing in information systems development offshoring projects. *Mis Quarterly*, 1211-1232.

Guardian (2020). Is the video games industry finally reckoning with sexism? https://www.theguardian.com/games/2020/jul/22/is-the-video-games-industry-finally-reckoning-with-sexism

Guo, W., Sengul, M., & Yu, T. (2020). Rivals' negative earnings surprises, language signals, and firms' competitive actions. *Academy of Management Journal*, *63*(3), 637-659.

Guo, W., Sengul, M., & Yu, T. (2021). The impact of executive verbal communication on the convergence of investors' opinions. *Academy of Management Journal*, *64*(6), 1763-1792.

Hampel, C. E., Tracey, P., & Weber, K. (2020). The art of the pivot: How new ventures manage identification relationships with stakeholders as they change direction. *Academy of Management Journal*, *63*(2), 440-471.

Hargadon, A. B., & Bechky, B. A. (2006). When collections of creatives become creative collectives: A field study of problem solving at work. *Organization Science*, *17*(4), 484-500.

Harrison, S. (2011). Organizing the cat? Generative aspects of curiosity in organizational life.

Harrison, S. H., & Dossinger, K. (2017). Pliable guidance: A multilevel model of curiosity, feedback seeking, and feedback giving in creative work. *Academy of Management Journal*, *60*(6), 2051-2072.

Harrison, S. H., & Rouse, E. D. (2015). An inductive study of feedback interactions over the course of creative projects. *Academy of Management Journal*, *58*(2), 375-404.

He, V. F., Puranam, P., Shrestha, Y. R., & von Krogh, G. (2020). Resolving governance disputes in communities: A study of software license decisions. *Strategic management journal*, *41*(10), 1837-1868.

Hobbs, P. (2013). Fuck as a metaphor for male sexual aggression. *Gender & Language*, *7*(2).

Huy, Q. N. (2011). How middle managers' group-focus emotions and social identities influence strategy implementation. *Strategic management journal*, *32*(13), 1387-1410.

Huy, Q. N. (2012). Emotions in strategic organization: Opportunities for impactful research. *Strategic Organization*, *10*(3), 240-247.

Jay, T. (2000). *Why we curse*. John Benjamins.

Jay, T., & Janschewitz, K. (2008). The pragmatics of swearing.

Jeppesen, L. B., & Frederiksen, L. (2006). Why do users contribute to firm-hosted user communities? The case of computer-controlled music instruments. *Organization Science*, *17*(1), 45-63.

Jeppesen, L. B., & Lakhani, K. R. (2010). Marginality and problem-solving effectiveness in broadcast search. *Organization Science*, *21*(5), 1016-1033.

Joseph, J., & Sengul, M. (2025). Organization design: Current insights and future research directions. *Journal of management*, *51*(1), 249-308.

Jurafsky, D., & Martin, J. H. (2014). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition. In.

König, A., Mammen, J., Luger, J., Fehn, A., & Enders, A. (2018). Silver bullet or ricochet? CEOs' use of metaphorical communication and infomediaries' evaluations. *Academy of Management Journal*, *61*(4), 1196-1230.

Kraut, R. E., & Resnick, P. (2012). *Building successful online communities: Evidence-based social design*. Mit Press.

Kreutzer, M., Walter, J., & Cardinal, L. B. (2015). Organizational control as antidote to politics in the pursuit of strategic initiatives. *Strategic management journal*, *36*(9), 1317-1337.

Kroon, D. P., & Rouzies, A. (2015). Reflecting on the use of mixed methods in M&A studies. In *The Routledge companion to mergers and acquisitions* (pp. 197-220). Routledge.
Lant, T. K., & Shapira, Z. (2000). *Organizational cognition: computation and interpretation*. Psychology Press.
Lant, T. K., & Shapira, Z. (2001). New research directions on organizational cognition. *Organizational cognition: Computation and interpretation*, 367-376.
Lee, G. K., & Cole, R. E. (2003). From a firm-based to a community-based model of knowledge creation: The case of the Linux kernel development. *Organization Science*, *14*(6), 633-649.
Lehavy, R., Li, F., & Merkley, K. (2011). The effect of annual report readability on analyst following and the properties of their earnings forecasts. *The accounting review*, *86*(3), 1087-1115.
Leonardi, P. M., & Jackson, M. H. (2004). Technological determinism and discursive closure in organizational mergers. *Journal of Organizational Change Management*, *17*(6), 615-631.
Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual review of psychology*, *66*(1), 799-823.
Levine, S. S., & Prietula, M. J. (2014). Open collaboration for innovation: Principles and performance. *Organization Science*, *25*(5), 1414-1433.
Levinthal, D., & Rerup, C. (2006). Crossing an apparent chasm: Bridging mindful and less-mindful perspectives on organizational learning. *Organization Science*, *17*(4), 502-513.
Lin, X., & Germain, R. (2003). Organizational structure, context, customer orientation, and performance: lessons from Chinese state-owned enterprises. *Strategic management journal*, *24*(11), 1131-1151.
Lix, K., Goldberg, A., Srivastava, S. B., & Valentine, M. A. (2022). Aligning differences: Discursive diversity and team performance. *Management Science*, *68*(11), 8430-8448.
Löbner, S. (2014). Evidence for frames from human language. *Frames and concept types: Applications in language and philosophy*, 23-67.
Loewenstein, J., Ocasio, W., & Jones, C. (2012). Vocabularies and vocabulary structure: A new approach linking categories, practices, and institutions. *Academy of Management Annals*, *6*(1), 41-86.
Lounsbury, M., & Glynn, M. A. (2001). Cultural entrepreneurship: Stories, legitimacy, and the acquisition of resources. *Strategic management journal*, *22*(6-7), 545-564.
Ludwig, S., De Ruyter, K., Mahr, D., Wetzels, M., Brüggen, E., & De Ruyck, T. (2014). Take their word for it. *Mis Quarterly*, *38*(4), 1201-1218.
Manning, C. D. (2008). *An introduction to information retrieval*.
Mantere, S. (2013). What is organizational strategy? A language-based view. *Journal of Management Studies*, *50*(8), 1408-1426.
Matias, J. N. (2019). The civic labor of volunteer moderators online. *Social Media+ Society*, *5*(2), 2056305119836778.
Miller, K. D., Fabian, F., & Lin, S. J. (2009). Strategies for online communities. *Strategic management journal*, *30*(3), 305-322.
Miric, M., Ozalp, H., & Yilmaz, E. D. (2023). Trade-offs to using standardized tools: Innovation enablers or creativity constraints? *Strategic management journal*, *44*(4), 909-942.
Mohan, S., Guha, A., Harris, M., Popowich, F., Schuster, A., & Priebe, C. (2017). The impact of toxic language on the health of reddit communities. Advances in Artificial Intelligence:

30th Canadian Conference on Artificial Intelligence, Canadian AI 2017, Edmonton, AB, Canada, May 16-19, 2017, Proceedings 30,
Nagaraj, A., & Piezunka, H. (2020). *How competition affects contributions to open source platforms: Evidence from OpenStreetMap and Google maps*.
Navis, C., & Glynn, M. A. (2010). How new market categories emerge: Temporal dynamics of legitimacy, identity, and entrepreneurship in satellite radio, 1990–2005. *Administrative Science Quarterly*, *55*(3), 439-471.
Neeley, T. (2017). *The language of global success: How a common tongue transforms multinational organizations*. Princeton University Press.
Neeley, T. B., Hinds, P. J., & Cramton, C. D. (2012). The (un) hidden turmoil of language in global collaboration. *Organizational Dynamics*, *41*(3), 236-244.
Nigam, A., & Ocasio, W. (2010). Event attention, environmental sensemaking, and change in institutional logics: An inductive analysis of the effects of public attention to Clinton's health care reform initiative. *Organization Science*, *21*(4), 823-841.
O'mahony, S., & Ferraro, F. (2007). The emergence of governance in an open source community. *Academy of Management Journal*, *50*(5), 1079-1106.
Ocasio, W. (2011). Attention to attention. *Organization Science*, *22*(5), 1286-1296.
Ocasio, W., & Joseph, J. (2008). Rise and fall-or transformation?: The evolution of strategic planning at the General Electric Company, 1940–2006. *Long Range Planning*, *41*(3), 248-272.
Ocasio, W., Laamanen, T., & Vaara, E. (2018). Communication and attention dynamics: An attention-based view of strategic change. *Strategic management journal*, *39*(1), 155-167.
Ogbonna, E., & Wilkinson, B. (2003). The false promise of organizational culture change: A case study of middle managers in grocery retailing. *Journal of Management Studies*, *40*(5), 1151-1178.
Okhuysen, G. A., & Bechky, B. A. (2009). 10 coordination in organizations: An integrative perspective. *Academy of Management Annals*, *3*(1), 463-502.
Parhankangas, A., & Renko, M. (2017). Linguistic style and crowdfunding success among social and commercial entrepreneurs. *Journal of Business Venturing*, *32*(2), 215-236.
Park, S., Piezunka, H., & Dahlander, L. (2023). Coevolutionary lock-in in external search. *Academy of Management Journal*(ja).
Phillips, N., Lawrence, T. B., & Hardy, C. (2004). Discourse and institutions. *Academy of management Review*, *29*(4), 635-652.
Piezunka, H., & Dahlander, L. (2015). Distant search, narrow attention: How crowding alters organizations' filtering of suggestions in crowdsourcing. *Academy of Management Journal*, *58*(3), 856-880.
Piezunka, H., & Dahlander, L. (2019). Idea rejected, tie formed: Organizations' feedback on crowdsourced ideas. *Academy of Management Journal*, *62*(2), 503-530.
Ploog, J. N., & Rietveld, J. (2024). Rolling the dice: Resolving demand uncertainty in markets with partial network effects. *Academy of Management Journal*(ja), amj. 2023.0133.
Ren, Y., Kraut, R., & Kiesler, S. (2007). Applying common identity and bond theory to design of online communities. *Organization Studies*, *28*(3), 377-408.
Safadi, H., Johnson, S. L., & Faraj, S. (2021). Who contributes knowledge? Core-periphery tension in online innovation communities. *Organization Science*, *32*(3), 752-775.
Scherer, K. R., & Moors, A. (2019). The emotion process: Event appraisal and component differentiation. *Annual review of psychology*, *70*(1), 719-745.

Schweisfurth, T. G., Schöttl, C. P., Raasch, C., & Zaggl, M. A. (2023). Distributed decision-making in the shadow of hierarchy: How hierarchical similarity biases idea evaluation. *Strategic management journal*, *44*(9), 2255-2282.

Seering, J., Wang, T., Yoon, J., & Kaufman, G. (2019). Moderator engagement and community development in the age of algorithms. *New Media & Society*, *21*(7), 1417-1443.

Seidel, V. P., Hannigan, T. R., & Phillips, N. (2020). Rumor communities, social media, and forthcoming innovations: The shaping of technological frames in product market evolution. *Academy of management Review*, *45*(2), 304-324.

Seo, E., Nagle, F., & Shah, S. (2021). Does Who Helps You Impact Your Behavior? Examining the Effects of Social Interactions on Knowledge Sharing in Online Communities. *Examining the Effects of Social Interactions on Knowledge Sharing in Online Communities (July 22, 2021). Harvard Business School Strategy Unit Working Paper*(21-026).

Shah, S., & Nagle, F. (2019). Why do user communities matter for strategy? *Harvard Business School Strategy Unit Working Paper*(19-126).

Shah, S. K., & Tripsas, M. (2007). The accidental entrepreneur: The emergent and collective process of user entrepreneurship. *Strategic entrepreneurship journal*, *1*(1-2), 123-140.

Shepherd, D. (2015). Party On! A call for entrepreneurship research that is more interactive, activity based, cognitively hot, compassionate, and prosocial. *Journal of Business Venturing*, *30*(4), 489-507.

Snihur, Y., Thomas, L. D., Garud, R., & Phillips, N. (2022). Entrepreneurial framing: A literature review and future research directions. *Entrepreneurship Theory and Practice*, *46*(3), 578-606.

Sood, S. O., Churchill, E. F., & Antin, J. (2012). Automatic identification of personal insults on social news sites. *Journal of the American Society for Information Science and Technology*, *63*(2), 270-285.

Statista. (2023). *Steam gaming platform - Statistics & Facts*. https://www.statista.com/topics/4282/steam/

Steam. (2011). *Steam Workshop Coming for Skyrim*. https://store.steampowered.com/oldnews/6906#:~:text=Announcement%20%2D%20Valve,Gift%20Cards%20%7C%20Steam%20%7C%20@steam

Steam. (2020). *Chat Filtering Now Available on Steam*. https://store.steampowered.com/news/app/593110/view/2855803154584367415

Sutton, L. A. (2001). The principle of vicarious interaction in computer-mediated communications. *International Journal of Educational Telecommunications*, *7*(3), 223-242.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive science*, *12*(2), 257-285.

Tuggle, C. S., Borgholthaus, C. J., Harms, P. D., & O'Brien, J. P. (2024). Setting the tone to get their way: An attention-based approach to how narcissistic CEOs influence the board of directors to take more risk. *Strategic management journal*, *45*(10), 2095-2121.

Urban, G. L., & Von Hippel, E. (1988). Lead user analyses for the development of new industrial products. *Management Science*, *34*(5), 569-582.

Von Hippel, E. (2005). Democratizing innovation: The evolving phenomenon of user innovation. *Journal für Betriebswirtschaft*, *55*, 63-78.

Von Hippel, E., & Von Krogh, G. (2003). Open source software and the "private-collective" innovation model: Issues for organization science. *Organization Science*, *14*(2), 209-223.
Von Krogh, G., & Von Hippel, E. (2006). The promise of research on open source software. *Management Science*, *52*(7), 975-983.
Vuori, N., Vuori, T. O., & Huy, Q. N. (2018). Emotional practices: How masking negative emotions impacts the post-acquisition integration process. *Strategic management journal*, *39*(3), 859-893.
Vuori, T. O. (2024). Emotions and attentional engagement in the attention-based view of the firm. *Strategic Organization*, *22*(1), 189-210.
Vuori, T. O., & Huy, Q. N. (2016). Distributed attention and shared emotions in the innovation process: How Nokia lost the smartphone battle. *Administrative Science Quarterly*, *61*(1), 9-51.
Vuori, T. O., & Huy, Q. N. (2022). Regulating top managers' emotions during strategy making: Nokia's socially distributed approach enabling radical change from mobile phones to networks in 2007–2013. *Academy of Management Journal*, *65*(1), 331-361.
Weber, M. (2008). The business case for corporate social responsibility: A company-level measurement approach for CSR. *European Management Journal*, *26*(4), 247-261.
Weick, K., E., Sutcliffe, K. M., & Obstfeld, D. (1999). *Organizing for high reliability: Processes of collective in mindfulness*. Elsevier Science/JAI Press.
Weick, K. E., & Sutcliffe, K. M. (2006). Mindfulness and the quality of organizational attention. *Organization Science*, *17*(4), 514-524.
Whorf, B. L. (1956). Language, thought, and reality: selected writings of….(Edited by John B. Carroll.).
Wired. (2020). Toxicity in Gaming Is Dangerous. Here's How to Stand Up to It. *Wired*. https://www.wired.com/story/toxicity-in-gaming-is-dangerous-heres-how-to-stand-up-to-it/
WSJ. (2010). Firm Bans Naughty Words in Emails; An "Unlearnable Lesson" on Wall Street? *WSJ*.
Zhang, X., & Zhu, F. (2011). Group size and incentives to contribute: A natural experiment at Chinese Wikipedia. *American Economic Review*, *101*(4), 1601-1615.
Zhou, J. (1998). Feedback valence, feedback style, task autonomy, and achievement orientation: Interactive effects on creative performance. *Journal of Applied Psychology*, *83*(2), 261.
Zhou, J. (2003). When the presence of creative coworkers is related to creativity: role of supervisor close monitoring, developmental feedback, and creative personality. *Journal of Applied Psychology*, *88*(3), 413.

TABLE 1: Descriptive Statistics and Correlation Matrix

	Mean	SD	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Cosine Distance	.65	.41									
Treatment	.24	.43	-0.17								
Post	.37	.48	0.09	-0.04							
User Retention	.91	.13	0.02	0.01	-0.21						
Community Size	4236.3	1464.7	-0.21	0.09	-0.13	0.37					
Community Age	30.00	27.9	0.02	-0.15	0.12	0.11	-0.09				
Fog Index	6.42	5.5	0.03	0.00	-0.01	-0.02	-0.04	-0.02			
Word Count	8.46	8.7	0.05	0.06	-0.01	-0.02	-0.05	-0.01	0.63		
Emotionality	.24	2.14	-0.01	0.00	0.00	0.00	-0.01	0.01	-0.02	-0.01	
Elapsed # of hours	63.46	67.42	0.08	-0.08	0.15	0.16	0.08	0.29	-0.08	-0.08	0.01

All correlations ≥ |0.01| are significant at p < .05

TABLE 2. Fixed Effects OLS Regression Analysis for Expression of Novelty

DV: Cosine Dissimilarity	**Model 1**	**Model 2**
	β (*SE*)	β (*SE*)
Control variables		
User Retention	0.387*** (0.012)	0.386*** (0.012)
Community Size	-0.000 (0.000)	-0.000 (0.000)
Community Age	0.007*** (0.000)	0.007*** (0.000)
Treatment		0.160 (0.161)
Post		0.043*** (0.010)
Treatment × Post (H1)		-0.064*** (0.014)
Constant	0.021 (0.097)	-0.115 (0.125)
Year FE	Yes	Yes
Month FE	Yes	Yes
Game FE	Yes	Yes
Creator FE	Yes	Yes
Observations	7,115	7,115

TABLE 2A. Relationship between vocabulary control and novelty (CS21)

	Model 1	Model 2	Model 3	Model 4
Treat X Post	-0.059**	-0.051**	-0.046**	-0.034**
	[-0.111, -0.008]	[-0.098, -0.005]	[-0.090, -0.002]	[-0.074, 0.006]
Obs.	7,115	7,115	7,115	7,115
Workshop ids	1,074	1,074	1,074	1,074
Operationalization	Cosine Distance (Recent, TF-IDF)	Cosine Distance (First, TF-IDF)	Cosine Distance (Recent, BERT)	Cosine Distance (First, BERT)
Time	[-4, 3]	[-4, 3]	[-4, 3]	[-4, 3]

TABLE 3. Fixed Effects OLS Regression Analysis for Richness in Feedback

Variable	**Model 1**	**Model 2**
	DV: Fog Index	DV: Number of Unique Words
	β (*SE*)	β (*SE*)
Control variables		
User Retention	0.369*** (0.083)	0.776*** (0.132)
Community Size	-0.000 (0.000)	-0.000 (0.000)

Community Age	-0.001** (0.000)	0.003*** (0.001)
Independent variables		
Treatment	-0.979*** (0.125)	-1.098 (0.199)
Post	-0.121** (0.048)	-0.184** (0.077)
Treatment × Post	-0.141** (0.074)	-0.253** (0.117)
Constant	7.586*** (0.089)	10.846*** (0.141)
Year FE	Yes	Yes
Month FE	Yes	Yes
Game FE	Yes	Yes
Observations	61,175	61,175

TABLE 4. Effect of Treatment on Salience of Feedback and Delays in Feedback

	Model 1	Model 2
	Salience of Feedback	Delays in Feedback
	Emotionality	*Time Between Feedback*
Treat X Post	-0.090*	15.198**
	[-0.187, 0.121]	[1.18, 29.21]
Obs.	61,175	61,175
Workshop ids	1,194	1,012
Method	TWFE	TWFE
Time	[-4, 3]	[-4, 3]

TABLE 5. Results of bootstrapping mediation regression analysis for relationships between Treatment × Post, Fog Index, and Novelty

	DV = **Novelty**				
			95% CI		Robust in
Independent variable	*B*	SE	Lower	Upper	Sobel test
Indirect effects mediated by Fog Index					
Treatment × Post	-0.0008***	0.0004	-0.0001	-0.0016	Yes
Direct effects					
Treatment × Post	-0.1215***	0.0073	-0.1359	-0.1072	
Total Effects					
Treatment × Post	-0.1207***	0.0073	-0.1351	-0.1063	

TABLE 6. Results of bootstrapping mediation regression analysis for relationships between Treatment × Post, Number of Unique Words, and Expression of Novelty

	DV = **Novelty**				
			95% CI		Robust in
Independent variable	*B*	SE	Lower	Upper	Sobel test
Indirect effects mediated by Number of Unique Words					
Treatment × Post	-0.0044***	0.0011	-0.0023	-0.0065	Yes
Direct effects					
Treatment × Post	-0.1255***	0.0069	-0.1391	-0.1119	

Total Effects					
Treatment × Post	-0.1211***	0.0068	-0.1345	-0.1077	

TABLE 7. Results of bootstrapping mediation regression analysis for relationships between Treatment × Post, Emotionality, and Novelty

	DV = **Novelty**				
			95% CI		Robust in
Independent variable	*B*	SE	Lower	Upper	Sobel test
Indirect effects mediated by Emotionality					
Treatment × Post	-0.0067	0.0043	-0.015	0.0018	No
Direct effects					
Treatment × Post	-0.0921***	0.0116	-0.0692	-1.1150	
Total Effects					
Treatment × Post	-0.08544***	0.0120	-0.0618	-0.1090	

TABLE 8. Results of bootstrapping mediation regression analysis for relationships between Treatment × Post, Times between Feedback, and Novelty

	DV = **Novelty**				
			95% CI		Robust in
Independent variable	*B*	SE	Lower	Upper	Sobel test
Indirect effects mediated by Times between feedback					
Treatment × Post	-0.0277***	0.0041	-0.0357	-0.0196	Yes
Direct effects					
Treatment × Post	-0.1122***	0.0095	-0.0934	-0.1310	
Total Effects					
Treatment × Post	-0.0845***	0.0100	-0.0647	-0.1043	

FIGURE 1: Implementation Timeline of Profanity Filter in Steam Workshop

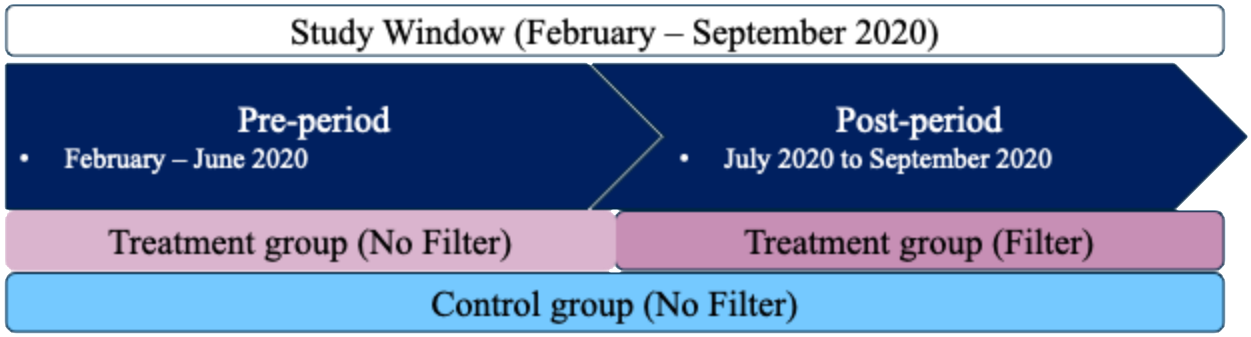


FIGURE 2: Effect of Vocabulary Control on Novelty

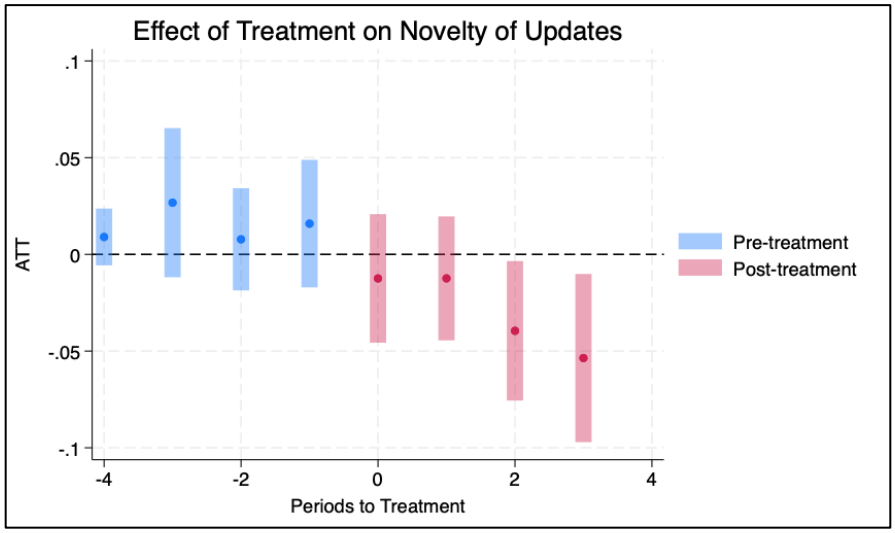


FIGURE 2: Effect of Vocabulary Control on Vocabulary Complexity

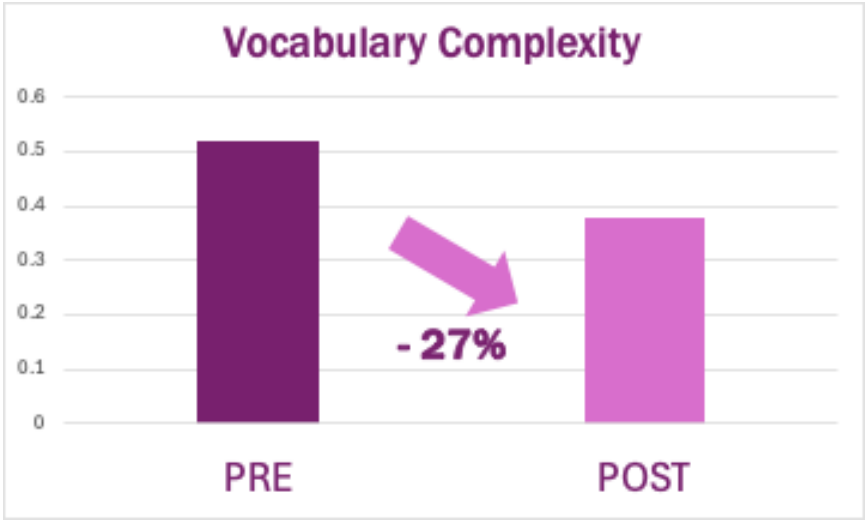


FIGURE 4: Effect of Vocabulary Control on Vocabulary Diversity

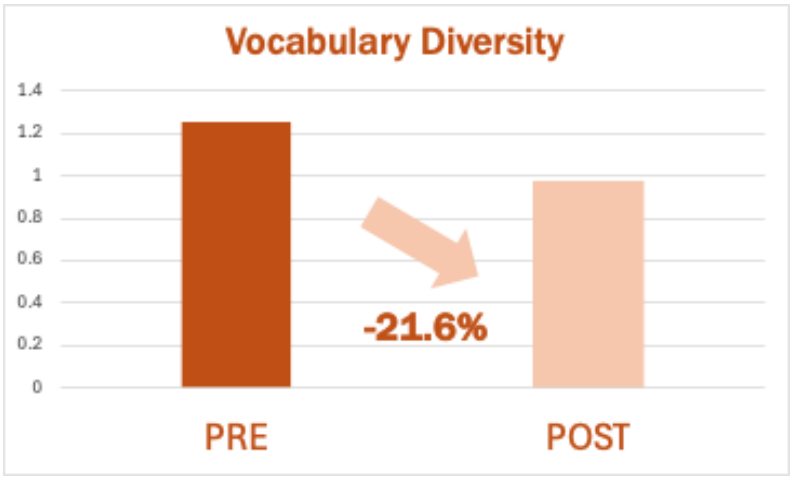


FIGURE 5: Effect of Vocabulary Control on Emotionality

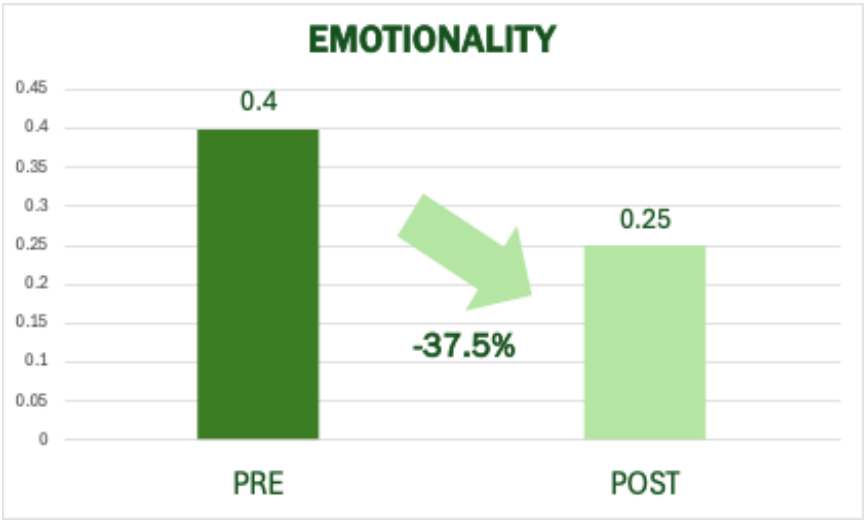
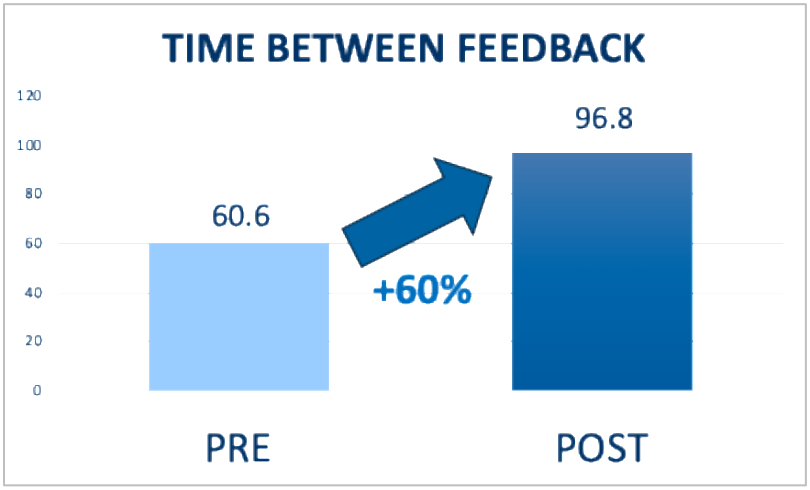


FIGURE 6: Effect of Vocabulary Control on Times Between Feedback



APPENDIX

1. Alternative Dependent Variable: Frequency of Updates

As an additional outcome, I examine whether vocabulary control influences **the** *frequency of community updates***,** measured as the number of updates submitted per month. Similar to where I predict the decrease in novelty of updates, I predict a decrease in frequency of updates.

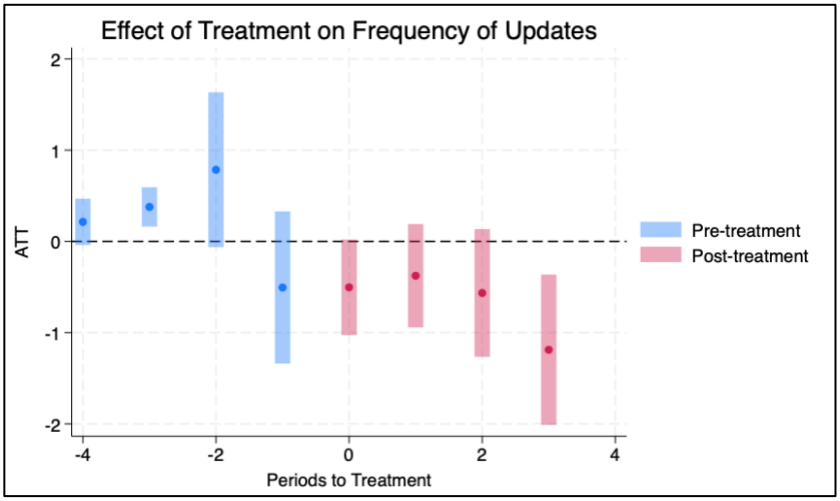
Results from CS21 DiD estimation (Table B1) confirm this prediction. Using a least-squares outcome model with inverse probability weighting, I find that the treatment reduces the frequency of updates by approximately 0.63 updates per month (approximately 40% decrease). This effect is statistically significant and indicates that vocabulary control not only dampens the novelty of updates but also reduces the overall pace of community-driven contributions.

FIGURE B1 shows a clear shift in update activity following treatment. After treatment, the frequency of updates diminishes, showing significantly fewer updates compared to baseline.

TABLE A1. CS21 results of the relationship between vocabulary control and frequency of updates

	Frequency of Updates
Treat X Post	-0.634**
	[-1.19, -0.077]
Obs.	8,729
Workshop ids	1,074
Control Group	Not yet treated
Operationalization	# of updates per month
Time	[-4, 3]

FIGURE A1. Effect of Treatment on Frequency of Updates



2. Boundary Conditions

To show how the impact of control varies depending on the underlying community characteristics, I examine boundary conditions that shape *when* vocabulary control most strongly affects novelty. In particular, I focus on three dimensions—community age, pre-treatment swear usage, and native speakers—that capture structural variation in how communities communicate.

Community Age

The results indicate that the negative effect of vocabulary control on novelty is more pronounced in younger communities, whereas older communities appear more resilient. The event-study plots (Figure D1) show that, for low community age, novelty declines steadily after treatment, while high community age communities exhibit a much weaker drop. The regression results confirm this difference: using cosine distance with TF-IDF, treatment significantly reduces novelty in younger communities (–0.105, p = 0.031), but the effect is small and not significant in older ones (–0.020, p = 0.264)[5]. This finding suggests that vocabulary control disproportionately suppresses novelty in newer communities, while more established communities are better able to absorb or adapt to imposed algorithmic vocabulary control.

TABLE B1. CS21 results of the relationship between vocabulary control and novelty by Community Age

DV: Cosine Distance	Low Community Age	High Community Age
Treat X Post	-0.105**	-0.020
	[-0.200, -0.009]	[-0.056, -0.015]
Obs.	3,791	4,364
Control Group	Not yet treated	Not yet treated
Time	[-4, 3]	[-4, 3]

NOTE: *Cosine Distance* measured with TF-IDF and by comparing the focal change note to the most recent change note.

Prior Swear Words

The results suggest that pre-treatment levels of swear word usage moderate the effect of vocabulary control on novelty. The event-study plot shows that both low and high pre-swear usage groups experience declines in novelty immediately following treatment, but the average drop is more salient among low pre-swear users. Regression estimates align with this: the treatment effect is negative and significant for the low pre-swear group (–0.039, p = 0.097), while the effect for the high pre-swear group is smaller and statistically insignificant (–0.023, p =

[5] A similar though weaker pattern appears when novelty is operationalized using BERT embeddings, with younger communities showing a negative, significant effect (–0.075, p = 0.084).

0.666). This pattern suggests that vocabulary control negatively impacts novelty more strongly in communities where swearing was less common before treatment, whereas communities with higher baseline levels of swearing appear less affected. This may be due to differences in how different communities adapt to moderation. In particular, as high-pre-swear communities were accustomed to informal strategies for circumventing restrictions, when moderation was imposed, these communities may readily draw on adaptive repertoires – such as misspellings or substitutions – that counter the disruptive effect of vocabulary control. I further investigate the types of deviant behaviors in the following post-hoc analyses.

TABLE B2. CS21 results of the relationship between vocabulary control and novelty by pre-swear usage

	Low Pre Swear	High Pre Swear
Treat X Post	-0.039*	-0.023 (p=0.666)
	[-0.084, 0.006]	[-0.126, 0.080]
Obs.	5,096	584
Control Group	Not yet treated	Not yet treated
Time	[-4, 3]	[-4, 3]

Native vs Non-Native English Speakers

For communities with a majority of non-native speakers, vocabulary control significantly reduces novelty, with the treatment effect estimated at –0.123 (p = 0.035). This indicates that because non-native speakers operate with an already narrower shared vocabulary, imposed constraints are more significant. By contrast, in communities with a majority of native speakers, the effect is negligible (0.008, p = 0.666). This suggests that native-majority communities are more resilient, as native speakers can flexibly adapt their expression and maintain novelty despite restrictions.

TABLE B3. CS21 results of the relationship between vocabulary control and novelty by Native vs Non-native English Speakers

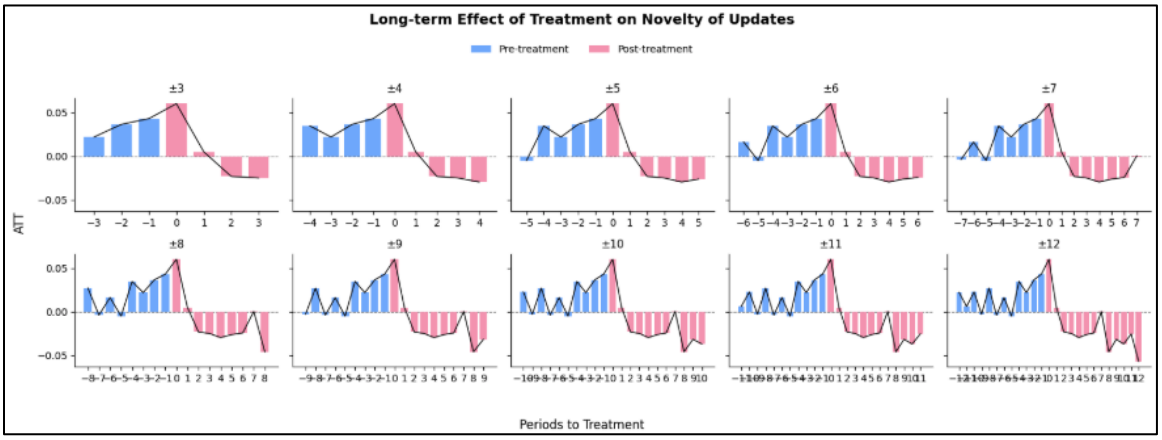
	Non-native	Native
Treat X Post	-0.123*	0.008
	[-0.238, -0.008]	[-0.036, 0.053]
Obs.	1,180	843
Control Group	Not yet treated	Not yet treated
Time	[-4, 3]	[-4, 3]

NOTE: I code a community as *1* if the majority of its members are native English speakers—defined as individuals who acquired English as their first language in early childhood, typically in countries where English is the dominant language (e.g., United States, United Kingdom, Canada, Australia, New Zealand, Ireland)—and as *0* otherwise.

3. Examining Long-term Effects

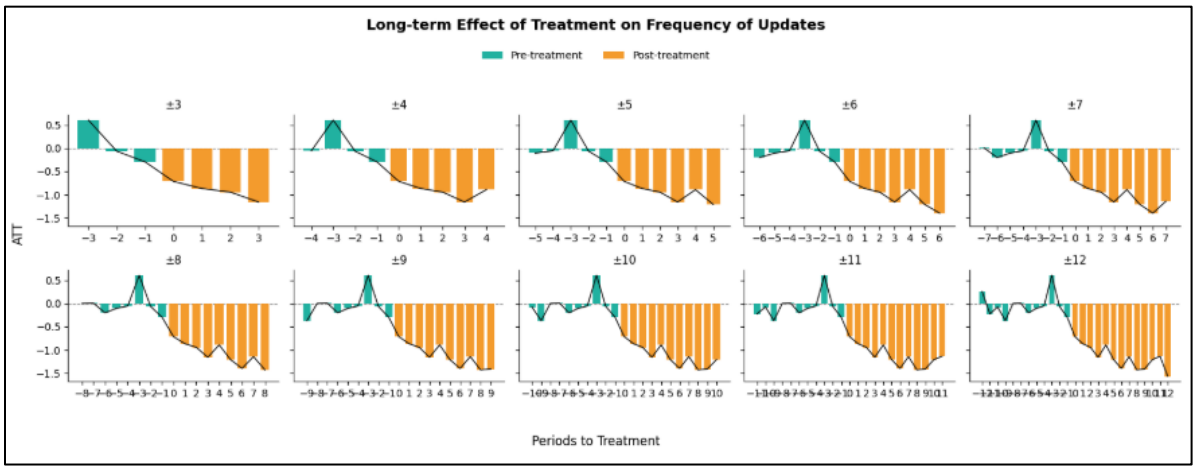
Novelty of Updates

The long-term effects on the novelty of updates reveal a consistent pattern of decline following treatment, with little evidence of recovery. In the immediate post-treatment periods (t = 0 to t = 2), novelty drops sharply below zero across nearly all event windows, indicating that updates become less original once the treatment is introduced. This negative effect persists throughout the longer horizons: for instance, in the ±9, ±11, and ±12 specifications, novelty remains well below baseline for as many as 8–12 periods after treatment. The estimates here show continued negative values without a clear return to pre-treatment levels. This suggests that the treatment induces a long-lasting reduction in novelty, with communities failing to fully recover their prior level of creative variation in updates even over extended timeframes.



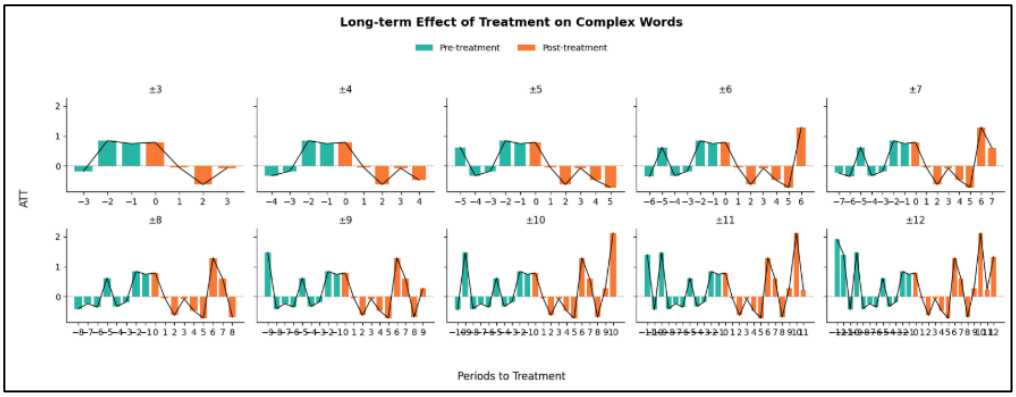
Frequency of Updates

The figures indicate that the treatment substantially reduces the frequency of updates over time. In the first panel, update frequency remains relatively stable before treatment, but declines sharply in the periods immediately following treatment, turning persistently negative. The following panels reveal a consistent downward trajectory: post-treatment effects are strongly negative and grow in magnitude with time. These results suggest that the treatment leads to fewer updates overall, with the reduction in change activity becoming more pronounced and enduring as the observation window lengthens. In short, the treatment not only dampens short-term activity but also has lasting effects on the volume of changes well beyond the initial intervention.



Complex Words

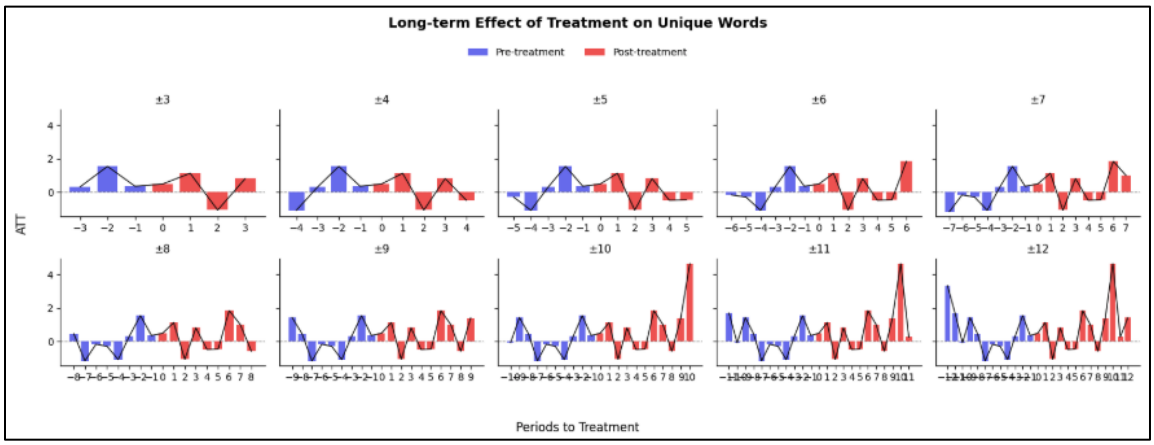
The long-term effects of the use of complex words show an immediate decline right after treatment, followed by evidence of bouncing back in later periods. For example, in the first two post-treatment periods (t = 0 and t = 1), complexity drops sharply below zero across most panels, indicating a simplification of language. Yet by around t = 3–5, the decline begins to moderate, and in several specifications (e.g., ±7 and ±9 windows) the estimates rise back toward baseline or even become slightly positive. By t = 6–8, the recovery is more evident, with the gap relative to pre-treatment shrinking considerably. In the longest horizons (e.g., ±11 and ±12), the pattern alternates between small negative and positive effects, suggesting that teams partially restore their use of complex words over time rather than remaining in a persistently simplified state. This temporal progression highlights a bounce-back dynamic: sharp declines immediately after treatment, but gradual adaptation and recovery within 3–8 periods.



Unique Words

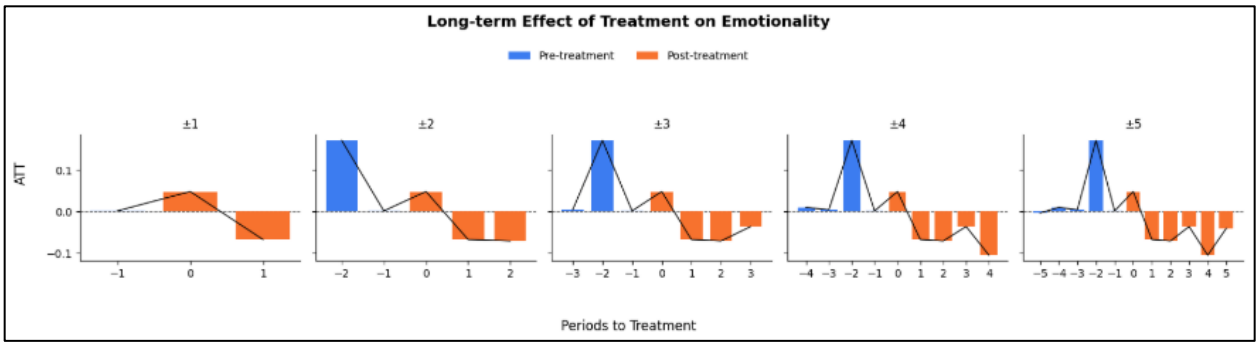
The results for unique words show a short-term disruption but clearer signs of bouncing back over longer horizons. Immediately after treatment (t = 0 to t = 2), the number of unique words tends to dip below zero in several specifications, suggesting a temporary reduction in the diversity of words. However, these declines are not persistent. By around t = 3–6, the estimates often rebound toward baseline, and in the wider windows (e.g., ±9, ±11, ±12), post-treatment periods fluctuate around zero or even turn slightly positive at times. This indicates that while

treatment initially constrains word variety, communities adapt and recover their use of diverse vocabulary, with long-term effects appearing far less negative and suggesting a modest rebound.



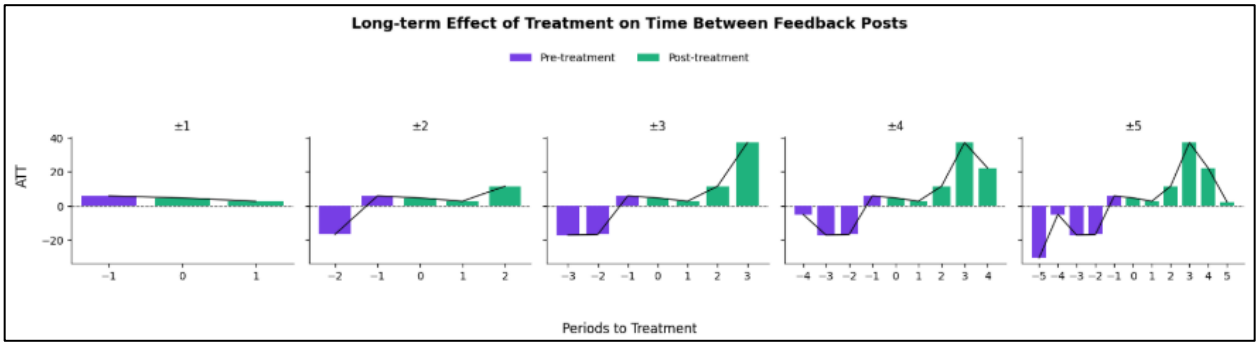
Emotionality

The results show that the treatment had an immediate but short-lived effect on emotionality, followed by a partial bounce-back. Initially, the effect of treatment on emotionality is that it is negative: It decreases emotionality. However, in the longer term, this effect diminishes: subsequent periods show smaller and even rebounds towards the baseline, with estimates converging back toward zero. The pattern indicates that while treatment temporarily decreases emotionality, this response fades, and communication begins to normalize, reflecting a bounce-back toward pre-treatment emotional levels.



Times between Feedback Posts

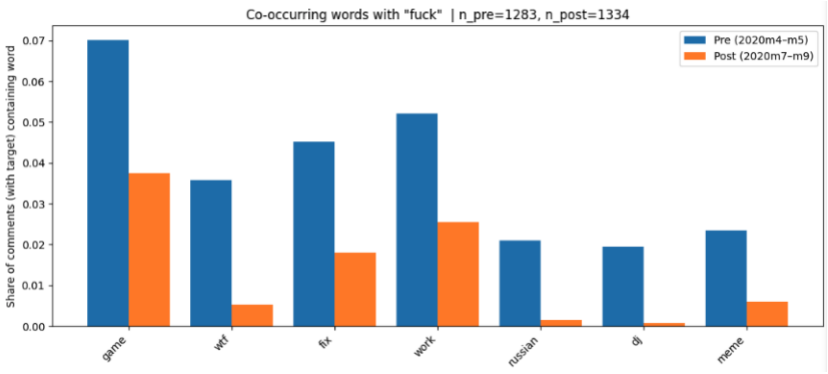
The results suggest that, while the treatment initially increases self-censorship, there are signs of a partial bounce-back over time. Specifically, the elapsed time between feedback rises sharply in the immediate post-treatment periods, indicating slower feedback at first. However, by three to six months out, the upward trajectory moderates: although feedback cycles remain longer than pre-treatment, the estimates show a gradual rebound toward baseline. This pattern implies that the treatment disrupts times in the feedback in the short run, but organizations begin to adjust and partially recover their pace of interaction in the longer term.

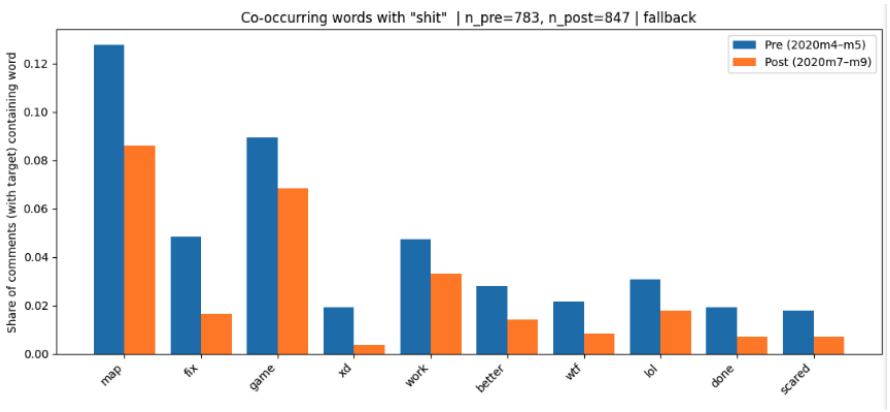


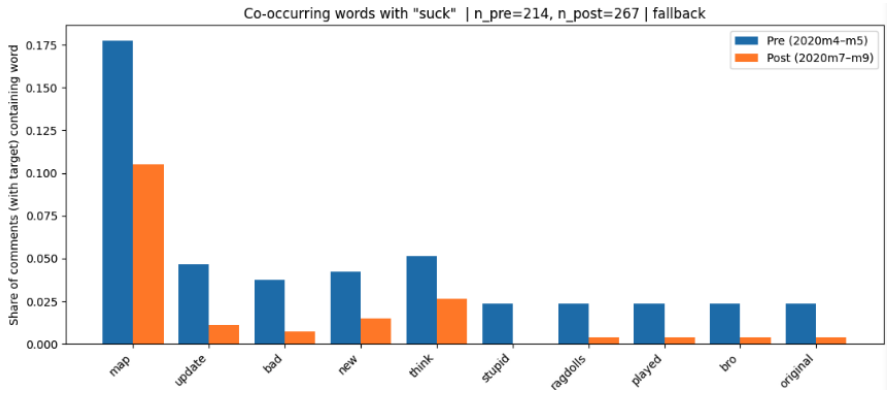
4. Changes in specific terms after the treatment

Changes in the co-occurred terms

The co-occurrence analysis shows that dropping profanity altered the content of communication. For *f***, dropping the term also reduced the presence of words like *game*, *fix*, and *work*, indicating that complaints or technical issues previously expressed through profanity became less prominent. A similar shift is evident for *sh***, where links to *fix*, *map*, *game*, and *work* declined substantially, and for *su***, where co-occurrences with *fix*, *map*, *update*, *new*, *think*, and *original* fell. Taken together, these results suggest that exchanged feedback became less content-rich and less problem-focused.

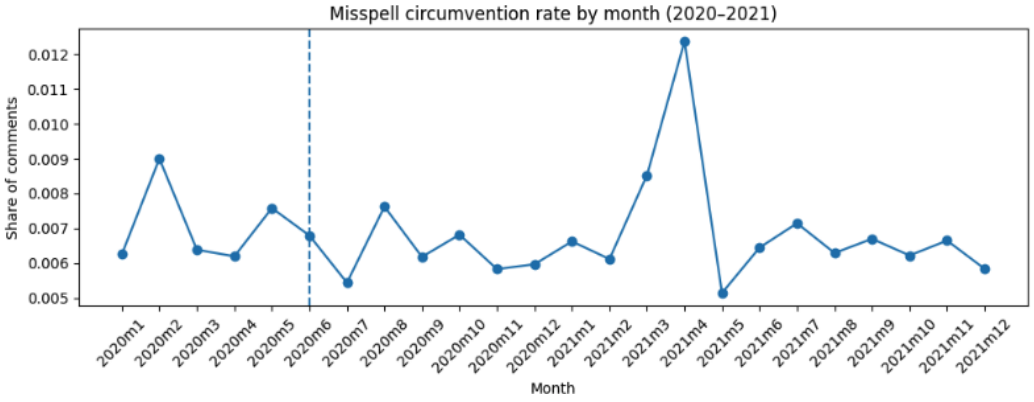


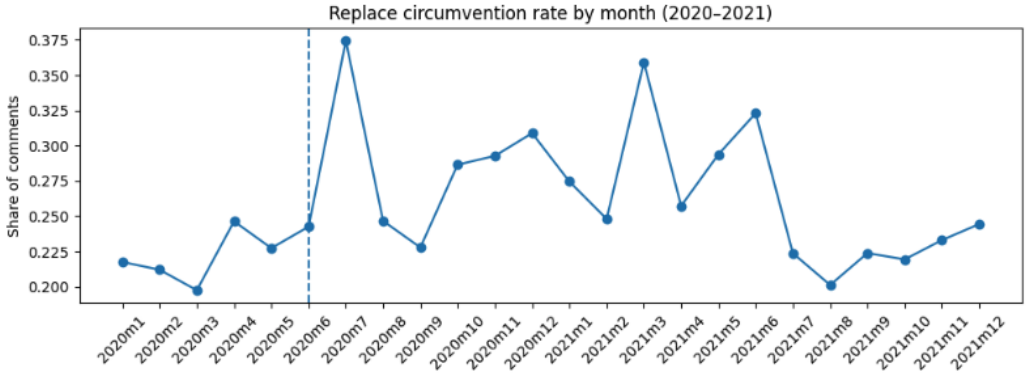




Deviating the filter

Further analysis revealed that users responded to vocabulary control by developing two distinct strategies of circumvention: misspelling and replacement. The misspell circumvention rate remained relatively low but stable over time, suggesting that users sought to bypass filters through misspelling. In contrast, replacement circumvention was much more common, with rates consistently above 20% of comments and surging to 30–35% in multiple months. Analysis of replacement tokens confirms that users substituted profanity with semantically unrelated but contextually adaptive words, many of which became increasingly frequent after treatment. This strategy was complemented by the proliferation of inventive profanity variants.





Root	Variant	Pre	Post	% Change
fuck	fck	9	36	+300%
fuck	fuckng	2	5	+150%
shit	sht	5	391	+7720%
shit	shiet	1	11	+1000%
shit	shity	6	16	+167%
shit	shiit	1	4	+300%
shit	shiiit	1	4	+300%
suck	sux	3	18	+500%
suck	sucka	1	6	+500%
suck	suks	1	2	+100%

5. Alternative mechanism: Changes in the proportion of users

To examine whether algorithmic vocabulary control led members to leave the community, I compared user retention rates before and after the treatment. The results show a sharp decline of approximately 9% in retention immediately following implementation (pre-treatment mean = 0.964; post-treatment = 0.877; $t = 42.70$, $p < 0.001$). This drop indicates that some members exited the community after the intervention. However, further analysis revealed that these departing members were largely inactive users who had not participated in idea generation. Thus, while overall retention declined, the core base of active contributors remained stable. To account for this shift in membership composition, I include retention as a control variable in the main analyses.