Predictability Over Discretion: How Rule Design Shapes Trust, Behavior, and Enforcement

A survey-based policy analysis with implications for Trust & Safety, platform governance, and product design.

Relevant to: Trust & Safety, Platform Policy, Governance, Risk, Product Operations

Cameron Liljedahl

Executive Summary

This project examines how people evaluate predictable, rule-based policies versus discretionary, case-by-case review when they are personally subject to platform rules. Using an original survey of 185 respondents, I compare user preferences between an automatic, time-based refund policy and a discretionary refund process reviewed by product support.

The results show a strong preference for rule-based outcomes. 153 of 185 respondents (82.7%) preferred an automatic refund within a fixed time window, while 32 respondents (17.3%) preferred discretionary review. This preference held despite discretionary review often being framed as fairer or more flexible, suggesting that certainty and speed outweighed the potential benefits of individualized judgment in this context.

The findings indicate that users value knowing outcomes in advance and receiving immediate resolution over the possibility of exception-based consideration. When people imagine themselves subject to enforcement, predictability appears to reduce uncertainty, waiting, and perceived arbitrariness, even when outcomes may be unfavorable.

These results have implications for platform policy design across Trust & Safety, governance, product operations, and risk functions. While discretion remains important for handling true edge cases, the evidence suggests that users prefer systems in which predictable rules are the default and discretionary judgment is narrowly scoped and explicitly bounded.

Project Overview

Platform policies determine not just what outcomes occur, but how those outcomes are experienced by users before, during, and after enforcement. Rule design shapes whether users can anticipate decisions, how long they wait for resolution, and whether outcomes feel consistent or arbitrary once applied at scale.

Policy teams often face a choice between systems that rely on clearly defined rules and systems that rely on discretionary review. While this tradeoff is usually discussed in terms of fairness, flexibility, or abuse prevention, its effects on user expectations during enforcement are less

explicit. Users experience policies through timing, clarity, and consistency, not through policy intent alone.

This project examines how those factors influence acceptance of enforcement outcomes and how user-facing preferences interact with broader questions of platform governance, enforcement systems, and operational design. The goal is to understand how rule design choices translate into user experience, and what that means for Trust & Safety, platform policy, product operations, and risk functions.

Core Question

When people are subject to platform rules, do they prefer predictable policies or discretionary review, and what does that preference imply for platform design?

Problem Framing: Predictability vs Discretion

Platforms design rules that determine how decisions are made and enforced across large numbers of user interactions. In many cases, these design choices come down to a tradeoff between predictable rules and discretionary review. This tradeoff is most visible when users imagine themselves subject to the policy.

Predictable rules are clear, uniform, and known in advance. Users can read the rule and determine the outcome without waiting for further judgment. In the context of refunds, a predictable rule might guarantee an automatic refund within a fixed time window. A user who submits a request within that window knows the refund will be issued, while a user outside the window knows it will not. Even when the outcome is unfavorable, the decision can feel acceptable because the rules are explicit and applied the same way each time. The user does not need to explain their situation or wait to see whether an exception will be made.

Discretionary review, by contrast, allows decisions to be made case by case rather than by a fixed rule. In a refund setting, this means a request is evaluated by a reviewer who considers factors such as timing, usage, or the reason for the request before deciding whether to approve it. This approach can feel fairer in situations where a strict rule would seem unreasonable, such as when a user narrowly misses a deadline or encounters an unexpected issue. However, discretionary review also introduces uncertainty. The user does not know the outcome in advance and may wait days for a decision, and different reviewers may reach different conclusions in similar situations.

Because both approaches have clear advantages, users may reasonably prefer one over the other depending on how they expect the interaction to play out. Predictable rules may be preferred when users value speed, clarity, and certainty, especially when they want to know the outcome immediately. Discretionary review may be preferred when users believe their situation deserves special consideration or when they expect flexibility to work in their favor. These preferences reflect how people expect the policy to work when it is applied to them.

Despite this, discussions of rule design are often driven primarily by platform concerns such as risk management, abuse prevention, or operational capacity. Less attention is paid to how users themselves evaluate predictable rules versus discretionary review when they are personally affected. A policy that appears fair in theory may feel frustrating in practice if its outcome is

uncertain or slow, while a rigid rule may feel acceptable if its consequences are clearly understood in advance.

This project focuses on that gap. Rather than asking which approach platforms prefer or which is easier to administer, it asks which approach users themselves prefer when faced with predictable rules versus discretionary review, and what that preference implies for platform design.

Survey Design & Methodology

To measure user preference between two enforcement models, I conducted an original survey with 185 Colgate University students. Responses were collected via an anonymous poll on Yik Yak over several days, allowing respondents to evaluate a familiar policy scenario without identifying information.

Each respondent was presented with the following question:

"If you bought a $120 item online and wanted a refund, which policy would you rather have?"

Respondents were required to select one of two options, with no neutral choice provided:

- Automatic refund within 14 days
- Case-by-case refund reviewed by product support

The two options were designed to differ only in how the decision was made. The automatic refund option represented a fixed, rule-based policy in which eligibility and outcomes were determined by a time window. The case-by-case option represented discretionary review, where refund decisions depended on individual evaluation rather than a predefined rule. Aside from decision structure, the scenario was held constant across respondents.

Refund policies are common across platforms and marketplaces, and the $120 price point was selected to feel meaningful without being extreme. The survey measured stated preference rather than objective efficiency or fairness. Respondents were not given additional context about platform costs, abuse prevention, or operational constraints.

This approach has limitations. The sample reflects a single university population and a hypothetical scenario. The survey tests one price point and one type of policy decision, so the findings should be interpreted as directional rather than universal. The goal is not to make broad population claims, but to capture how users evaluate predictable rules versus discretionary review in a concrete and familiar setting.

Results

All 185 respondents provided a valid preference between the two policy options.

A strong majority preferred the predictable, rule-based policy:

- 153 of 185 respondents (82.7%) preferred an automatic refund within 14 days

- 32 of 185 respondents (17.3%) preferred case-by-case refund review by product support

The preference distribution shows a clear skew toward the predictable option. Support for the automatic refund policy exceeded discretionary review by a wide margin, indicating that respondents did not view the two options as equally attractive.

These findings reflect stated preferences within a specific population and scenario. The survey does not measure observed behavior or establish causality, and the results should be interpreted as directional rather than universal.

Why Discretion Feels Fair but Is Often Disliked in Practice

Discretionary review often appeals intuitively because it promises judgment. The idea that a real person will consider context, intent, and circumstances aligns with common notions of fairness. Discretion suggests flexibility and the ability to correct outcomes that rigid rules might mishandle. Policies that rely on discretionary review are often defended as more humane or more just.

Discretionary review introduces delays and uncertainty. Requests must be reviewed before a decision is made, leaving users without a clear answer in the meantime. Even short delays can feel costly when the outcome affects money, access, or account standing. By contrast, predictable rules resolve outcomes immediately, removing the need to wait for a decision.

A second issue is uncertainty. Under discretionary review, users do not know in advance whether their request will be approved. Two requests that appear similar may receive different outcomes depending on how they are interpreted or who reviews them. From the user's perspective, this uncertainty can be more frustrating than an unfavorable but predictable result. Not knowing what will happen often feels worse than knowing the answer will be no.

A third issue is the lack of clear expectations. Discretionary systems make it difficult for users to understand what outcome to expect or how to evaluate their chances. Users may feel compelled to explain their situation in detail, justify their behavior, or emphasize particular aspects of their case in hopes of influencing the decision. This can turn policy enforcement into a negotiation rather than a rule, increasing stress and perceived arbitrariness.

When users think about how a discretionary policy would actually work for them, practical concerns take priority. Predictable rules provide clarity. Users can assess eligibility on their own, know the outcome immediately, and see that the same rule applies across cases. Even when the outcome is unfavorable, predictable rules can feel acceptable because the process is clear and does not depend on persuasion or interpretation.

This helps explain the gap between how people talk about fairness in general and how they evaluate policies when they are personally affected. In theory, discretion sounds fair because it allows judgment. In practice, predictability often feels better because it reduces waiting, uncertainty, and the sense that outcomes are arbitrary. What users appear to value is not judgment for its own sake, but confidence that the system will behave in a known and consistent way.

The survey results align with this pattern. Most respondents preferred the predictable option, suggesting that certainty and immediacy matter more than individualized judgment when users imagine themselves subject to a policy decision.

Importantly, this does not mean that discretionary review is always undesirable. Discretion can play an important role in handling true edge cases or correcting clear mistakes. However, broad reliance on discretionary enforcement may conflict with user preferences when policies apply at scale. Users appear to prefer systems in which discretion is limited, and outcomes are largely determined by clear rules, rather than systems where outcomes depend primarily on review.

The key insight is that policies designed around discretionary judgment may not feel fair when users are subject to them. While discretion allows flexibility, users often prefer predictable rules that produce clear and timely outcomes. Understanding this gap matters for platforms designing policies that users are more likely to trust and accept in practice.

Rule Design Tradeoff Matrix

Table 1. Tradeoffs between predictable rule-based enforcement and discretionary review.

Dimension	Predictable Rules	Discretionary Review
Predictability	High	Low
Speed	High	Variable
User Expectations	Clear	Uncertain
Edge-Case Handling	Moderate	Higher
Outcome Variation	Low	Higher
Reliance on Judgment	Low	High

Interpretation of Dimensions:
- Predictability: Fixed rules determine outcomes in advance, allowing users to know exactly whether they qualify, while discretionary review leaves outcomes unknown until a decision is made.
- Speed of decision: Automatic rules resolve decisions immediately, whereas discretionary review depends on review queues and response time.
- User expectations: Clear rules allow users to assess eligibility on their own, while discretionary systems make outcomes harder to anticipate.
- Edge-case handling: Fixed rules are more predictable once boundaries are known, while discretionary review allows denial of suspicious or borderline requests.
- Outcome variation: Rule-based systems apply outcomes uniformly, while discretionary review can produce different results for similar requests.
- Reliance on judgment: Predictable rules minimize the need for human judgment, while discretionary review centers decisions on individual interpretation.

Implications for Platform Teams

The findings from this survey are not prescriptions and do not suggest that any single rule design is always optimal. Instead, they clarify how users tend to evaluate predictable rules versus

discretionary review when they imagine themselves subject to enforcement. For platform teams, these preferences matter because they shape trust, behavior, and operational outcomes once policies are deployed at scale.

Trust & Safety

For trust and safety teams, user preference for predictable rules has direct implications for how enforcement systems are experienced. When outcomes are determined by clear rules, users are more likely to understand decisions without escalation. Predictability reduces the need for users to contest outcomes, explain intent, or seek exceptions, which can lower frustration and support volume.

This does not imply that discretion should be eliminated. Rather, it suggests that discretionary exceptions should be narrow, explicit, and clearly bounded. When discretion is used sparingly and transparently, it can address true edge cases without undermining the overall predictability of the system. Broad or loosely defined discretion can increase uncertainty and perceived arbitrariness, even when decisions are well intentioned.

As a result, enforcement systems that rely primarily on clear, predictable rules may reduce user friction and escalation more effectively than systems that default to discretionary review.

Policy & Governance

From a policy and governance perspective, predictable rules shape how fairness and legitimacy are perceived once enforcement occurs. Policies that specify eligibility and outcomes in advance allow users to evaluate decisions against known rules. Even unfavorable outcomes can feel acceptable when users believe the rule was applied consistently and without interpretation.

Vague or open-ended policies shift responsibility away from the rule itself and toward judgment calls made during enforcement. This can create inconsistency across similar cases and make it harder for users to understand why decisions differ. Over time, reliance on discretion may weaken confidence in the policy framework, even if individual decisions are defensible in isolation.

Product & Platform Operations

From a product and operations standpoint, rule design shapes user expectations long before enforcement occurs. Predictable rules allow users to assess eligibility on their own and make informed choices about whether to act, request a refund, or pursue an appeal. This reduces confusion during moments of enforcement.

Discretionary systems, while flexible, often push uncertainty downstream. Users may not know what outcome to expect or how long a decision will take, which can create anxiety and repeated follow-ups. From an operational standpoint, predictable policies simplify user flows and reduce the need for escalation and repeated clarification.

Risk & Compliance

For risk and compliance teams, clarity improves defensibility. Rules that specify conditions and outcomes make it easier to explain decisions, audit enforcement, and demonstrate consistency over time. When discretion is minimized, variability across cases is reduced, lowering exposure to claims of unfair treatment or inconsistent application.

Discretionary review increases reliance on individual judgment, which can introduce variability even when reviewers act in good faith. While discretion can be valuable for handling unusual cases, heavy dependence on it may increase compliance risk as policies scale across users, geographies, and enforcement teams.

Summary

Taken together, these implications suggest that user preference for predictability aligns with operational clarity across multiple platform functions. Predictable rules reduce uncertainty for users while supporting consistency, scalability, and defensibility for platforms. Discretion remains important, but users appear to prefer it as a limited exception rather than the default mode of enforcement.

Understanding how users evaluate these tradeoffs can help platform teams design policies that are not only enforceable but also trusted and accepted in practice.

Connection to Enforcement Systems

The survey in this paper measures user preference, not system performance. It captures how people evaluate predictable rules versus discretionary review when considering how a policy would be applied to them. Whether those preferences are realized in practice, however, depends on how rules are enforced once a policy is deployed.

Predictable outcomes require more than clearly written policies. They depend on enforcement systems that apply rules consistently and automatically, without introducing new interpretation at each decision point. When rules are executed the same way every time, users can rely on outcomes matching what the policy describes. This consistency is what allows predictability to be experienced rather than merely stated.

Discretionary enforcement places greater weight on human judgment. While judgment can be valuable in handling unusual situations, it also introduces variability across cases. Differences in interpretation, context, or reviewer decision-making can lead to inconsistent outcomes even when policies are well intentioned. As discretionary systems scale across users, teams, and geographies, maintaining uniform application becomes more difficult.

Users experience outcomes, not enforcement intent. A predictable policy that is applied inconsistently can feel just as uncertain as a discretionary one. Conversely, even simple rules can build trust when enforcement is reliable and consistent over time.

This distinction clarifies the relationship between user preference and implementation. The survey results suggest that users favor predictable rules, but delivering that predictability depends on enforcement systems that limit ad hoc judgment and apply rules consistently. Discretion still plays a role, but it appears most effective when used as a clearly defined exception rather than the primary mode of enforcement.

Understanding this link between rule design and enforcement execution is essential for platforms deciding not only what policies to write, but how those policies are carried out at scale.

Practical Policy Recommendation

The findings suggest that platforms are better served by defaulting to predictable, rule-based policies rather than broad discretionary review. Clear rules allow users to understand outcomes in advance, reduce uncertainty during enforcement, and increase acceptance of decisions, even when those decisions are unfavorable.

Discretion remains important, but it appears most effective when it is explicitly limited, clearly defined, and used as an exception rather than a default. Narrow exception windows, written eligibility criteria, and transparent escalation paths can preserve flexibility for true edge cases without undermining the overall predictability of the system. When discretion is bounded, users are less likely to perceive outcomes as arbitrary, and enforcement teams are less exposed to inconsistent decision-making.

Designing policies in this way aligns user preference for clarity with operational needs around abuse control, consistency, and defensibility. Rather than choosing between rigid rules and open-ended judgment, platforms can structure policies so that predictability is the default and discretion is the exception.

This approach treats rule design as a core element of platform trust, not just an enforcement detail, and offers a practical framework for scaling policies that users are more likely to understand, accept, and trust.

Limitations and Future Work

This analysis reflects stated preferences within a specific population and a single policy scenario. The sample is not representative of all platform users, and preferences may vary across user groups, platforms, and enforcement contexts.

The survey measured hypothetical choice rather than observed behavior and tested only one price point and one type of policy decision. Preferences may differ in higher-stakes settings or for policies involving access, moderation, or account restrictions.

Future work could expand this approach to broader populations, additional policy types, and observed enforcement outcomes to better understand how user preferences interact with real-world system design.