

Designing for Critical Thinking on Social Media

Réalisé par :

KUSCH Olivia

Professeur :

CHERUBINI Mauro

Expert :

HUMBERT Mathias

Lausanne, le 24 août, 2026

—

Master en Systèmes d'Information

Faculté des Hautes Etudes Commerciales

Université de Lausanne

Abstract [EN]

Research on misinformation has found that people do not fall for false content because they reason badly, but because they do not reason at all in the moment the judgement is made. This points to the environment in which information is met, and for most people that environment is a social media feed. Most mainstream feeds are engineered for frictionless engagement, and withhold both the occasion to reason and the material to reason with: much of the interface is arranged simply to carry the user to the next post, and rarely encourages them to slow down and think. This is exactly what misinformation needs to travel: a claim that is never evaluated can be believed on familiarity and shared before accuracy is ever considered.

This study therefore asks: how can a feed be designed to restore reasoning, and how do users respond to such a feed? Misinformation is the motivation for this question rather than its measure: the study examines whether users take up the occasion to reason when a feed offers it, not whether they believe or share fewer false claims. The study began with a structured walkthrough of six mainstream platforms, examining which cognitive mechanisms their interface features afford, and whether each favors quick, intuitive judgement or deliberate reasoning. The walkthrough found that mainstream feeds favor the former along almost every dimension, and this yielded a set of requirements for a feed built the other way around. These requirements resulted in REFLECT, a full working social media feed in which each automaticity-affording feature identified in the walkthrough is inverted, including clickable sources on every claim, a panel explaining why each post was shown, user controls over the ranking algorithm, and neutralized engagement counts. In a within-subjects study, fifteen participants then used both REFLECT and their own most-used platform in a single session, followed by a comparative interview.

Each of the interventions set out by the requirements was found and used, but each also met a form of resistance from participants. The means of verifying a claim were taken up: fourteen of the fifteen participants opened at least one source. However, when asked how they would decide what to trust, nine described shortcuts: treating the presence of a source, how many people endorsed a source, or the comments as a sufficient substitute for or preferred method of checking. Eight participants said that the feed felt slower and healthier, but eight also said, sometimes in the same breath, that they would not adopt it in place of what they used now, citing the slowness and lack of dopamine. The means of seeing why posts appeared and controlling the algorithm was also noticed by everyone, but proved redundant, because participants already felt that they understood their algorithm and how to influence it. What they wanted instead was not an explanation or fine-grained control, but a single reset of their algorithm, named by nine of fifteen, though five saw no need to govern their feed at all, content to let the algorithm do the work for them.

The thread between findings is that when REFLECT made claims easier to check, the feed easier to control, and the pace slower, participants preferred a quicker route wherever possible, and the slowness that made the feed feel healthier is what made it unlikely to be adopted. Making a deliberative route easier does not make it the one that users take; the frictionless feed is not only something that the interface imposes on users, but something that users are reluctant to give up.

As misinformation becomes more abundant and more convincing, the moments in which users should pause and think are multiplying, and countermeasures, whether community notes or fact check labels, are currently installed into mainstream frictionless feeds. The findings suggest that making such measures

"easy to reach" does not necessarily make users perceive them as "easy to do", especially when a shortcut remains. The study therefore suggests that a countermeasure cannot rely on a user choosing to deliberate: it must sit inside the feed, and cost no more than the shortcuts users would otherwise take. Countering disinformation on social media is therefore not as simple as adding an intervention, or even rebuilding the feed; the open question this thesis raises is how such measures can be designed so that they are accepted by the users who receive them.

Keywords: dual-system theory; affordances; social media design; deliberation; friction; research through design; prototype evaluation

Abstract [FR]

Les recherches sur la mésinformation ont montré que les gens ne se laissent pas prendre par des contenus faux parce qu'ils raisonnent mal, mais parce qu'ils ne raisonnent pas du tout au moment où le jugement se forme. Cela renvoie à l'environnement dans lequel l'information est rencontrée, et pour la plupart des gens, cet environnement est un fil d'actualité de réseau social. La plupart des fils grand public sont conçus pour un engagement sans friction, et privent l'utilisateur à la fois de l'occasion de raisonner et de la matière pour le faire : une grande partie de l'interface est simplement agencée pour conduire l'utilisateur au post suivant, et l'incite rarement à ralentir et à réfléchir. C'est exactement ce dont la mésinformation a besoin pour circuler : une affirmation qui n'est jamais évaluée peut être crue par simple familiarité et partagée avant même que la question de son exactitude ne se pose.

Cette étude pose donc la question suivante : comment concevoir un fil qui restaure le raisonnement, et comment les utilisateurs réagissent-ils à un tel fil ? La mésinformation est ici la motivation de cette question plutôt que sa mesure : l'étude examine si les utilisateurs saisissent l'occasion de raisonner lorsqu'un fil la leur offre, et non s'ils croient ou partagent moins d'affirmations fausses. L'étude a commencé par une analyse structurée (walkthrough) de six plateformes grand public, examinant quels mécanismes cognitifs les fonctionnalités de leur interface favorisent, et si chacune privilégie le jugement rapide et intuitif ou le raisonnement délibéré. Cette analyse a montré que les fils grand public privilégient le premier sur presque toutes les dimensions, et en a tiré un ensemble d'exigences pour un fil construit à l'inverse. Ces exigences ont donné naissance à REFLECT, un fil de réseau social pleinement fonctionnel dans lequel chaque fonctionnalité favorisant l'automatisme identifiée lors de l'analyse est inversée, notamment des sources cliquables sur chaque affirmation, un panneau expliquant pourquoi chaque post a été affiché, des contrôles utilisateur sur l'algorithme de classement, et des compteurs d'engagement neutralisés. Dans une étude intra-sujets, quinze participants ont ensuite utilisé à la fois REFLECT et leur plateforme la plus utilisée au cours d'une même session, suivie d'un entretien comparatif.

Chacune des interventions issues des exigences a été trouvée et utilisée, mais chacune a aussi rencontré une forme de résistance de la part des participants. Les moyens de vérifier une affirmation ont été adoptés : quatorze des quinze participants ont ouvert au moins une source. Toutefois, lorsqu'on leur a demandé comment ils décideraient de ce qu'il faut croire, neuf ont décrit des raccourcis : la présence d'une source, le nombre de personnes l'ayant approuvée, ou les commentaires, considérés comme un substitut suffisant à la vérification ou comme une méthode préférable à celle-ci. Huit participants ont déclaré que le fil leur semblait plus lent et plus sain, mais huit ont également dit, parfois dans la même phrase, qu'ils ne l'adopteraient pas à la place de ce qu'ils utilisent actuellement, invoquant la lenteur et l'absence de dopamine. Les moyens de voir pourquoi les posts apparaissaient et de contrôler l'algorithme ont eux aussi

été remarqués par tous, mais se sont révélés redondants, car les participants avaient déjà le sentiment de comprendre leur algorithme et de savoir l'influencer. Ce qu'ils souhaitent, ce n'était ni une explication ni un contrôle fin, mais une simple réinitialisation de leur algorithme, mentionnée par neuf participants sur quinze, tandis que cinq ne voyaient aucun besoin de gouverner leur fil, se satisfaisant de laisser l'algorithme faire le travail à leur place.

Le fil conducteur entre ces résultats est que lorsque REFLECT a rendu les affirmations plus faciles à vérifier, le fil plus facile à contrôler et le rythme plus lent, les participants ont préféré la voie la plus rapide chaque fois que c'était possible, et la lenteur qui rendait le fil plus sain est précisément ce qui le rendait peu susceptible d'être adopté. Rendre la voie délibérative plus accessible n'en fait pas celle que les utilisateurs empruntent ; le fil sans friction n'est pas seulement quelque chose que l'interface impose aux utilisateurs, mais quelque chose que les utilisateurs hésitent à abandonner.

À mesure que la mésinformation devient plus abondante et plus convaincante, les moments où les utilisateurs devraient s'arrêter et réfléchir se multiplient, et des contre-mesures, qu'il s'agisse de notes communautaires ou d'étiquettes de vérification des faits, sont aujourd'hui installées dans des fils grand public sans friction. Les résultats suggèrent que rendre ces mesures « faciles à atteindre » ne suffit pas nécessairement à ce que les utilisateurs les perçoivent comme « faciles à faire », surtout lorsqu'un raccourci subsiste. L'étude suggère donc qu'une contre-mesure ne peut pas reposer sur le choix de l'utilisateur de délibérer : elle doit s'inscrire au cœur du fil, et ne pas coûter davantage que les raccourcis que les utilisateurs emprunteraient autrement. Lutter contre la désinformation sur les réseaux sociaux n'est donc pas aussi simple qu'ajouter une intervention, ni même que reconstruire le fil ; la question ouverte que soulève ce mémoire est de savoir comment concevoir de telles mesures pour qu'elles soient acceptées par les utilisateurs qui les reçoivent.

Mots-clés : théorie des systèmes duaux ; affordances ; conception des réseaux sociaux ; délibération ; friction ; recherche par le design ; évaluation de prototype

Table of Contents

Abstract [EN]	1
Abstract [FR]	1
Table of Contents	3
1. Introduction	5
2. Theoretical Framework	6
2.1 Dual System Theory.....	7
2.2 Affordance Theory.....	7
2.3 Misinformation.....	8
2.5 The Six Dimensions.....	9
2.6 Synthesis of Dimensions, Properties, Processing Modes, and Mechanisms.....	15
3. Methodology	16
3.1 Phase 1 – Comparative Platform Analysis.....	16
3.2 Phase 2 – Prototype Development.....	22
3.3 Phase 3 – Concept Study.....	24
3.4 Coding and Analysis.....	27
3.5 Ethics.....	28
4. Comparative Platform Analysis	29
4.1 Algorithmic Amplification.....	29
4.2 Immersive Design.....	30
4.3 Engagement Metrics.....	32
4.4 Sharing and Reporting Friction.....	33
4.5 Content Format.....	34
4.6 Deliberation Affordances.....	35
4.7 Synthesis.....	37
4.8 Design goals (DG) and requirements (R).....	39
5. Design	40
5.1 REFLECT Environment.....	41
5.2 Departures and Design Boundaries.....	45
6. Results	46
6.1 DG1 – Protect Cognitive Capacity.....	46
6.2 DG2 – Transparent, Governable Feed Curation.....	48
6.3 DG3 – Restore Evaluation at Point of Judgment.....	50
6.4 DG4 – Intentional Calibrated Action.....	51
6.5 UEQ-S Comparative Ratings.....	52
6.6 Boundary Conditions.....	53
7. Discussion	54
7.1 Affordances Shifted Processing, but Shortcuts are Displaced.....	54
7.2 REFLECT Was Accepted, But Use Was Conditional.....	56
7.3 Awareness Was Never the Deficit, Levers Were.....	57

7.4 Synthesis.....	59
7.5 Design Implications.....	60
8. Limitations.....	61
8.1 Stimulus and Artifact.....	61
8.2 Evaluation Setting and Measures.....	61
8.3 Sample, Data Quality, Analysis.....	62
8.4 The Comparative Analysis and Scope of Claims.....	62
9. Future Work.....	63
9.1 Evaluate a Personalized Prototype at Scale.....	63
9.2 Isolating Legibility from Prominence.....	64
9.3 Designing for the Slot Users Assigned.....	64
10. Conclusion.....	64
11. Bibliography.....	66
12. Appendices.....	70
Appendix A – Original Comparative Analysis Tables.....	70
Appendix B – AI Prompt Examples.....	77
Appendix C – REFLECT Prototype Changelog.....	78
Appendix D – Prototype Action Logs.....	79
Appendix E – Screening Questionnaire.....	81
Appendix F – Moderator Guide.....	84
Appendix G – Interview Coding & Summary Sheet.....	88
Appendix H – Supplemental UEQ Tables.....	91

1. Introduction

In this era of social media and twenty-four hour news cycles, many people get news and information through a feed. It is also increasingly true that social media platforms decide, to some extent, in what order, in what format, and how much effort it takes to react to it. In many cases, especially on the most dominant platforms, these feeds play video automatically, a user can scroll infinitely, virality is encouraged through likes, shares, and follows, and these interactions can complete in one tap. A singular tap requires little of a user, and this paper sets out to argue that requiring so little of users when exposed to news and information is the problem. If an interface removes every opportunity for a user to pause, consider their thoughts before acting, or follow an argument, it also removes opportunities for careful thinking to begin.

Dual System Theory offers a description for how this happens psychologically. System 1 is fast, automatic, and relies on cues that are familiar and emotional, whereas System 2 is slow, effortful, and only engages when there is spare capacity and something prompts it (Kahneman 2011). Research on misinformation has found that people don't fall for false content because they reason badly, rather because they do not reason at all when the judgment is made (Pennycook & Rand, 2021). This paper argues that this points to the environment. If lack of reasoning is the primary driver of this susceptibility, the design of the environment in which they find information could be a promising opportunity for intervention. Affordance theory is the other lens used, arguing that design choices do not dictate user behavior, but they can raise or lower the friction encountered in executing it (Gibson 1979). These lenses are used together to read a platform feature as "affording" System 1 or System 2 processing.

Existing work has tested individual interventions, like accuracy prompts before sharing or content dispute labels, but these are usually tested in an otherwise unchanged platform. This paper argues that System 1 affording features compound each other, and more than one feature likely needs to be changed in order to fully support a deliberative process. Less is known about the dynamics of an environment fundamentally designed for System 2. While some platforms may afford more critical thinking than others, mainstream platforms overwhelmingly and increasingly prioritize low friction engagement.

This work was organized into three phases that use an output from a prior phase as inputs to the next. Phase 1 is a comparative analysis of six mainstream platforms (TikTok, Instagram, Facebook, X, Reddit, and Bluesky) across six dimensions derived from cognitive literature. The dimensions are algorithmic amplification, immersive design, engagement metrics, friction, content format, and deliberation affordances. Each dimension is first defined in methodological terms alongside a related mechanism. Then, specific interface properties of the six platforms are analyzed to determine whether they afford System 1 or System 2 processing, and therefore which corresponding cognitive mechanism is triggered. The analysis is an assessment of which interface properties most strongly afford System 1, determining how their functionality can be inverted to instead afford System 2. The output is a set of design goals and requirements which are the foundation for Phase 2 which builds a prototype feed, REFLECT, against those requirements. Phase 3 evaluates the prototype in a within-subject study with fifteen participants. Each participant used their own preferred platform and REFLECT in the same session, thinking aloud while they completed the same four tasks and a UEQ-S (Short-User Experience Questionnaire), and took part in a comparative interview. In literature this is called research through design, in which building an artifact is part of how the question is studied.

This is a formative concept study, and does not measure truth discernment or claim a causal shift from System 1 to System 2 or vice versa. It simply asks whether a deliberation-supporting feed is perceived as legible, tolerable, and appealing enough at first user, whether people would consider using it, and which affordances they welcomed, ignored, or worked around. The following research questions structure the evaluation:

RQ1 - Design

What interface changes shift a feed environment from System 1 dominant to System 2 permitting, while still appealing to users regarding their use of the app?

RQ2 - Acceptance

Do users accept and uptake the deliberation-supporting affordances? Do they feel it would change their behavior?

RQ3 - Control

Does visible, governable, control make users feel more in control and perceive that it supports their reflection?

This paper makes three primary contributions. The first is an analytical framework that links six interface dimensions to dual system mechanisms with a comparative analysis of how platforms and their properties surface these mechanisms and processing affordances. The second is the REFLECT prototype, in which every feature can be traced back through a requirement and design goal to the property in the comparative analysis it responds to. The third is the evaluation, which found that taken together, the interventions succeeded in making the environment System 2 permitting, but that permitting deliberation does not necessarily cause it – when a shortcut remained, it was taken. Participants mostly accepted the affordances, but with conditions, and placed REFLECT beside their current platforms rather than in place of them.

The rest of the paper is structured as follows: Chapter 2 describes the theoretical framework and the six dimensions. Chapter 3 describes the entire methodology of all phases, including ethics and AI disclosures. It then loops back to Chapter 4 presenting the comparative analysis and the requirements and design goals it produces. Chapter 5 picks these up, presenting REFLECT and the reasoning behind each feature. Chapter 6 reports the results of the evaluation by design goal, and Chapter 7 discusses them against the research questions, also stating the design implications. Chapters 8 and 9 give limitations and future work, and Chapter 10 concludes.

2. Theoretical Framework

This chapter establishes a theoretical foundation that the rest of the thesis builds on. Dual System theory describes the two modes in which people process information, and affordance theory describes how an interface encourages or discourages certain actions. Together they allow a platform feature to be read as affording either automatic (System 1) or deliberate (System 2) thinking. This reading is what the entire study depends on, and every dimension, requirement, and finding in the chapters is traced back to it.

2.1 Dual System Theory

System 1 is fast, automatic, and associative. Kahneman, author of “Thinking Fast and Slow”, characterizes it as the mode that operates “automatically and quickly, with little or no effort and no sense of voluntary control” (Kahneman, 2011). It functions below conscious awareness and is activated by familiar stimuli, emotional arousal, and conditions of low cognitive effort. Rather than evaluating information against evidence, it relies on mental shortcuts that substitute pattern recognition for deliberate assessment. Khatri, Samuel, and Dennis (2018) note that this automatic cognition is shaped by personality characteristics and by pattern matching drawn from past experience, which means System 1 does not approach new information neutrally but through predispositions formed before the content is encountered.

System 1 processing has direct consequences for how information is judged, as it treats the ease and familiarity of an input as a proxy for its truth. Since evaluating facts requires mental effort that System 2 naturally avoids, it substitutes the difficult question of accuracy for a simpler feeling of cognitive ease. As Kahneman observes, “familiarity is not easily distinguished from truth” (Kahneman, 2011). Under System 1, content that feels fluent or emotionally resonant is therefore accepted as credible without being subjected to accuracy evaluation. In an information environment, this is the cognitive default that governs belief unless it is interrupted.

Instead, System 2 processing is slow, deliberate, and resource-intensive. In Kahneman's words, it “allocates attention to the effortful mental activities that demand it” and is associated with “the subjective experience of agency, choice, and concentration” (Kahneman, 2011). It is recruited when System 1 encounters something it cannot resolve automatically, such as novelty, contradiction, or an explicit prompt toward analytical thinking.

Two features of System 2 are central to the argument. First, it is effortful and, in Kahneman's terms, inclined toward “laziness,” engaging only when a situation appears to require it. Second, it depends on available cognitive capacity: when attentional resources are depleted or the environment does not prompt deliberation, System 2 does not engage and System 1 remains the default processing mode. This conditional, default-interventionist relationship is what platform design exploits, because the conditions under which System 2 would otherwise intervene can be structurally withheld.

The relevance of this architecture to misinformation is established by Moravec, Kim, and Dennis (2020), who show that interventions targeting System 1 and System 2 reduce belief in fake news through distinct routes, and that combining them is roughly twice as effective as either alone. Their results show that the way a cue is presented determines which system responds to it. This is why present work focuses on interface affordances rather than correcting content.

It is important to note that this binary framing is best treated as an analytical model rather than a literal description of two separate mechanisms. Hochman (2024) argues that intuitive and deliberate characteristics coexist within processes rather than dividing cleanly between two systems. This paper adopts the System 1 and System 2 distinction for its explanatory clarity while acknowledging this more graded account.

2.2 Affordance Theory

An affordance is an action possibility that an environment offers to an actor. The concept originates with Gibson, whose ecological account holds that the environment presents action possibilities directly to a

perceiver. In his formulation, "the affordances of the environment are what it offers the animal, what it provides or furnishes, either for good or ill" (Gibson, 1979, p. 127). For Gibson the affordance is relational rather than fixed, since the same feature offers different possibilities depending on the capacities of the actor who encounters it.

Non-determinism is an analytically important point in the context of affordance theory. A design feature does not compel a particular behavior, it raises or lowers the cognitive cost of an action and therefore makes some actions more inviting than others. A feature opens or burdens a pathway without dictating an outcome, making an action deliberately easier or harder to perform. Qi and Tan (2024) make the same point when they reject a technological determinism that would treat platform features as a fixed environment, arguing instead that affordances describe potentials that users may or may not take up. This is what allows a platform feature to be linked to a cognitive process. Bhuiyan (2023) formalizes that link by combining dual-process theory with the affordance concept into what he calls "dual process cognitive affordances", separating the automatic affordances that engage heuristic processing from reflective affordances that encourage analytical reasoning. Whereas Dual System Theory specifies which cognitive system operates under which conditions, affordance theory specifies how design organizes those conditions, and this underlies every dimension in the analysis – each of which links a design property to either System 1 or System 2.

Empirical work supports the claim that affordances shape sharing behavior specifically. Apuke and Omar (2021) found that affordances for information sharing, status seeking, and self-expression predicted COVID-19 fake news sharing, since features that make sharing immediate also make verification unlikely, though their sample was drawn from a single national context (Nigeria) and so generalizes with caution. Wu, Wu, and Xiao (2025) extend this by treating misinformation sharing as an unintended outcome of affordances rather than a malfunction, noting that social media affordances carry contradictory effects in which the same properties that enable participation also enable the spread of false content. In their account, an affordance such as information accessibility heightens involvement while making careful evaluation harder, so sharing proceeds without deliberation. Qi and Tan (2024) show the same affordance logic operating on the corrective side, predicting users' willingness to share debunking information on a single platform (Weibo), which indicates that the mechanism is general to sharing rather than specific to false content. Taken together these studies ground the design-to-cognition link at the center of the analysis, while their single-platform and single-country samples mark the limits of how far each result can be generalized.

2.3 Misinformation

For the purposes of this paper, misinformation refers to false or misleading content that is encountered and circulated through a platform feed. The definition is intentionally scoped to the user interface: the paper is concerned with how the feed frames the encounter and circulation of information rather than with the content's origin or the intent behind its production. This should not be confused with disinformation, which depends on whether misinformation is spread intentionally. This analysis targets the design conditions under which any false or misleading content propagates rather than the motives of those who introduce it.

There are two primary findings from the literature that are relevant to this analysis. First, poor truth discernment is due to a lack of careful reasoning and reliance on heuristic shortcuts such as familiarity rather than politically motivated reasoning (Pennycook & Rand, 2021). This is what makes an interface

encouraging System 2 and affording conscious reasoning a relevant intervention. Second, belief and sharing are separate acts. People share content without necessarily believing or having evaluated it, so a share is not evidence of endorsement. Pennycook et al. (2021) concluded that "sharing does not necessarily indicate belief." This separation is the foundation of the inattention account, a concept later developed in the friction dimension. It supports the claim that frictionless design lets a share complete before accuracy is ever considered. Misinformation, from this perspective, circulates less because users are deceived and more so because the conditions for effectively evaluating accuracy are absent at the moment of sharing.

It should be specified that misinformation is considered to be the motivating problem rather than a measured outcome. The scope of the analysis is the interface and conditions under which content would be encountered, whether the environment supports deliberation, and whether a System 2-affording environment could be accepted en masse. The study intends to understand susceptibility to those conditions rather than in whether they believe the information they digest.

2.4 The Six Dimensions

These dimensions were not adopted from an existing taxonomy. Dual System Theory and Affordance Theory were used as a lens through which social media platforms could be analyzed, the motivation being to understand the underlying mechanisms of critiques of social media. Specifically, how do current platform user interfaces encourage or discourage cognitive activity, healthy usage patterns, or facilitate misinformation being spread without mental engagement. While preventing misinformation is the motivating problem behind this study, the framework is not designed to measure whether readers believe false content, rather *in what mental state do users find information*. Dual System Theory and Affordance Theory provide a structural lens to evaluate the interface itself, and allow us to identify design properties that afford System 1 or System 2.

In this framework a dimension refers to a grouping of properties that have associated mechanisms supported by literature. By mapping these design choices to cognitive literature, a lens is created that can evaluate how the interface structurally shifts a user towards automaticity (System 1) or deliberate engagement (System 2). The six dimensions are listed below, the methodological process of how they were derived from properties is in Chapter 3, and they are analyzed in how they surface across mainstream platforms in Chapter 4.

2.4.1 Algorithmic Amplification: Opacity and Illusory Truth

Algorithmic amplification refers to the use of recommendation systems to curate and prioritize content based on prior engagement signals. Where a recommendation system is present, it is the algorithm that decides what content is surfaced, how prominently, and how repeatedly. This thoroughly shapes the information environment each user encounters without making the curation process visible, easily modifiable, or turned off. This ambiguity is referred to as opacity. Illusory truth instead is the idea that if a user sees information once, the second time they will perceive it as true because of their familiarity with the topic. Several properties have meaningful impacts across social media platforms: whether the curation logic is visible, the degree of user control over it, and how reachable that control is from the feed. This dimension most directly feeds these algorithmic opacity and illusory truth mechanisms. Engagement-based curation surfaces the same and similar content repeatedly, and repetition is the input the illusory truth effect runs on. Because the curation is opaque, users can unintentionally mistake

algorithmically curated repetition for reality, and illusory truth is the mechanical term for this cognitive shortcut.

Existing literature demonstrates why this combination of opacity and repetition makes the interface important to analyze. Pennycook, Cannon, and Rand (2018) show that a single prior exposure increases the perceived accuracy of fake news headlines, both within the same session and a week later, even when the headline carries a fact-checker warning or contradicts the reader's politics. The only limit to this effect is extreme implausibility (e.g. the Earth is perfectly square). Only a slight degree of plausibility was required for the claim to take hold on a user. The authors conclude that disputed tags do not solve the problem and that platforms effectively help false stories take root as belief. Algorithmic curation supplies exactly the frictionless, repeated exposure this mechanism requires.

The same curation logic also applies to the kind of content most likely to be shared without scrutiny. Brady et al. (2021), drew on two observational Twitter studies (7,331 users and 12.7 million tweets) and two behavioral experiments, and found that positive social feedback for expressions of moral outrage increases the likelihood of further outrage, and that users additionally conform to the outrage norms of their networks. They conclude that platform design interacts with these human learning mechanisms to shape online discourse. Since the algorithmic preference for emotionally charged content is opaque, users perceive the amplified anger as a reflection of public sentiment which manufactures illusory truth about social norms.

Rhodes (2022) supplies the downstream effect on belief. In a survey experiment, participants who read fake stories that matched their own political worldview rated them as more believable than participants who saw a mixed set of stories, which indicates that when an algorithmic filter bubble surrounds users with content they already agree with, it raises the credibility of false content without needing to increase the amount (of false content) they are exposed to. This environment weaponizes both opacity and illusory truth because the curation mechanism hiding dissenting views is opaque and the repeated exposure to fake news, or simple agreeable news, creates the illusion of truth.

In all of these scenarios, System 1 reads the resulting familiarity as a signal of truth, making it a victim of the illusory truth effect. System 2 could theoretically discount that familiarity if it could be recognized as the result of curation, but this would require the curation logic (making the algorithm less opaque) to be visible and the feed to be something the user could alter. Chapter 4 discusses whether any platform supplies those conditions, and at what cost in effort.

2.4.2 Immersive Design: Cognitive Load and Capacity Depletion

For this section, the properties of infinite scroll and related dark patterns are used to define the “immersive design” dimension. Infinite scroll is the continuous loading of content that removes the natural stopping points a finite or paginated feed would impose. Each boundary in a feed, the end of a page or the edge of a unit, is a moment where the user makes a small decision about whether to continue, and when there is no “end” to the feed, there are no more micro-decisions where deliberation might begin. Dark patterns are interface choices that exploit cognitive biases and decision heuristics to steer behavior toward platform interests rather than user intent. Gray et al. (2018) derived five recurring dark-pattern strategies: nagging, obstruction, sneaking, interface interference, and forced action. Mildner et al. (2023) show these strategies operate in social networking services specifically, generally appear bundled rather than in isolation and are only partially recognized. These properties are described together because, though infinite scroll generates the depleted state, the retention strategies (obstruction: entry and exit cost; interface

interference: autoplay) are what hold the user inside it. By encouraging the user into a continuous flow, the immersive design increases cognitive load and therefore depletes mental capacity for deliberation. The following paragraphs detail this sequence from the interface property, to cognitive load, to capacity depletion, and how this shuts down System 2 processing.

Both the infinite scroll and dark patterns terminate in the same mechanism, cognitive load, but they occupy different positions in the causal chain. Cognitive Load Theory holds that working memory has limited capacity and that load is the demand placed on it at a given moment (Sweller, van Merriënboer, & Paas, 1998). In the case of social media platforms, infinite scroll produces that load. A continuous feed keeps content arriving without a break, keeps demand on working memory high, and the boundaries that would let a user pause are removed by the continuous load of the infinite scroll. Bhowmik (2025) describes the endless scroll as generating information overload that draws down cognitive resources without returning meaningful cognitive engagement. The dark patterns in this dimension do not generate that load so much as sustain it, especially in cases where the environment is easy to enter, and requires a conscious decision to leave. Darda and Gupta (2026) trace a pathway in which dark patterns produce induced cognitive load and decision fatigue, which is consistent with this “sustaining” role.

How this load impacts thinking is the direct link to Dual System Theory. Maranges, Schmeichel, and Baumeister (2017) make the important distinction that cognitive load is not only a gradual draining of willpower, it is as a distraction that pulls attention away from critical processing. Attention occupied by a continuous stream makes this attention not available to evaluate a single item within it. A boundary in a feed would be a place to stop and give you the attention capacity to think, since the small decision of whether to continue is itself a moment of deliberation, and removing that moment is the specific design contribution of this dimension. System 2 engages only when spare capacity is available and something prompts it to act. A boundary-free feed that holds attention at capacity doesn’t supply either, so System 1 continues to run as the default mode. The user does not decide to stop checking accuracy because the design arranges for the question never to receive the attention it would need.

2.4.3 Engagement and Sharing Mechanics: Heuristic Substitution

Engagement metrics are the publicly visible quantification of interactions, such as likes, comments, shares, reposts, and views, displayed at the point of consumption. Since these counts appear next to content before a user has evaluated it, they make social data available in advance of any judgment about the content itself. This dimension examines which social signal properties are visible (likes, comments, views, circulation) each platform surfaces and how it presents them.

These visible signals are the input to a heuristic substitution. Cialdini (2001) describes social proof as the tendency to treat the behavior of others as a guide to one's own, most strongly under uncertainty, which is the standing condition of the fast, immersive feed described in Section 2.4.2. Morris et al. (2012) show what this looks like for credibility specifically. Across their experiments, users were poor judges of truthfulness from content alone and leaned instead on surface cues such as the author's name and image, regardless of the item's actual truth. The visible signal stands in for an evaluation the user does not perform.

The substitution is where this meets Dual System Theory. A credibility cue does not only influence a judgment, it replaces the effort of making one. Meinert and Krämer (2022) suggested in their study that the presence of an expertise cue may reduce the time participants took to judge credibility and choose a

source. This reduced decision effort is the defining signature of a heuristic. The expertise cue lets System 1 return an answer at the very moment System 2 would otherwise have to engage to evaluate the claim.

The idea that a higher like count makes content more believable should be stated carefully, as it is separate from the expertise cue and not well supported. Meinert (2020) found that source and expertise cues raised perceived credibility while raw like and share counts did not reliably do so, and recent evidence likewise indicates that the influence of visible engagement counts on trust is conditional rather than direct (He and Zheng, 2025). The robust version of the heuristic therefore rests on source and expertise cues, and on friend-level endorsement, rather than on raw popularity counts. Chapter 4 records which of these signals social media platforms surface, and with what precision. The effort of *producing* these signals belongs to the friction dimension and is analyzed there.

2.4.4 Friction: The Inattention Account

Friction refers to the degree of effort, the number of steps, or the interruption built into a user action. In platform design this effort is distributed unevenly. Many of the actions that drive engagement, such as sharing, reacting, and continuing to scroll, are built to require minimal deliberate effort. Conversely, actions that involve reflection, such as reading before sharing, checking a source, reporting content, or adjusting settings, carry more steps and more interruption. When both these types of friction exist at once, it is called friction asymmetry. To analyze this dimension, properties like taps to repost, taps to share privately, taps to report or flag, and whether the original source is visible when sharing are considered.

The distribution of the friction is intentional. Garland et al. (2026) argue that efficiency is the dominant design choice on social media, chosen to maximize engagement and ad revenue, at the expense of deliberation and reflective engagement. Tomalin (2023) shows that friction is applied selectively rather than eliminated outright, showing that platforms remove friction from actions they encourage and add it elsewhere. Users tend not to push past friction unless they are highly motivated, the platform's friction choices guide which behaviors actually occur.

This dimension feeds the inattention account, established by Pennycook et al. (2021) which holds that people often share content without having evaluated it, not because they cannot tell true from false but because their attention is absorbed by momentum and fluency of the feed and other cues. They demonstrated that the truth of a headline has little effect on whether people intend to share it even though it strongly affects their judgments of its accuracy, and this gap is driven by inattention rather than by indifference to truth. What matters most for this dimension is their further finding that a small prompt directing attention to accuracy reduced the sharing of false content. A minimal piece of friction, aimed at accuracy, was enough to change the outcome. The friction asymmetry is what removes that prompt. A frictionless share completes before any accuracy check can occur, and nothing in the interface interrupts to ask for one. Tomalin (2023) explains that since users rarely push past friction, placing the effort on verification and none on sharing essentially guarantees that verification will be skipped.

In Dual System Theory terms, the frictionless share is a System 1 action that completes before System 2 has any occasion to engage. A micro-decision, a prompt, or a required step would be an opening for deliberation, and the asymmetry is the systematic removal of those openings from the engagement path while preserving them on the reflective path that users are unlikely to take. This is also where the depleted state from the immersive design dimension converts into behavior. A user whose capacity is already occupied by a continuous feed defaults to the low-effort path, and the asymmetry ensures the low-effort path is the one that shares without checking. The share completes, the social signal propagates, and at no

point was the content evaluated by anyone in the chain. Chapter 4 measures where each platform places the various steps of sharing and reporting.

2.4.5 Content Format: Processing Fluency and Affect

Content format concerns the modality a platform structurally privileges, whether video, image, text, or a mix, and how that modality is presented within each unit. Where the immersive design dimension examines the structure of the feed and how a user moves through it, this dimension examines what is inside each unit. This dimension is set up to assess properties like default video playback, dominant feed format, video position in navigation, if reading is required to consume a unit, and whether text is truncated by default.

The central argument is that content format is not cognitively neutral, because video and image are taken in with less effort than text. This connects to two System 1 related mechanisms: processing fluency and affect. Processing fluency is the felt ease of taking content in. A robust finding in cognitive psychology is that people misread this ease as a signal of truth (Kahneman, 2011; Pennycook et al., 2018). Video and image are more fluent than text, since they are absorbed with less deliberate effort. Therefore, a video-led feed delivers more content in the modality that feels truest regardless of its veracity. Sundar et al. (2021) demonstrate this in the context of misinformation as well. Comparing false stories presented as text, audio, or video, they found that video is processed more superficially and, as a result, is believed more readily and shared more often. They refer to this mental shortcut as a realism heuristic, because a video looks like a direct recording of reality, so the user treats seeing as evidence rather than evaluating the claim.

Affect is the immediate feeling response a stimulus evokes, positive or negative, which can itself serve as an input to judgment in place of appraisal (Slovic et al., 2007). Visual and especially video content carries emotional charge more immediately than text, and emotion functions as its own heuristic. Martel et al. (2020) found that relying on emotion increases belief in fake news, while deliberation increases the ability to tell true from false. An emotionally arousing video produces a felt response that System 1 can read as a signal of significance or truth before any analytic appraisal occurs (Kahneman, 2011). Format therefore supplies both the fluency and the effect that these heuristics run on. Newman et al. (2012) found that this applies to images as well. Even a decorative, nonprobative photograph raises the perceived truth of an associated claim, which bears directly on the image-heavy feeds in this set.

However, the limit of these effects should be noted. Lau (2025), in a meta-analysis, found that fluency raised narrative transportation, the experience of being absorbed into a story, for text and text-plus-image but not for video, and transportation is related to but distinct from credibility. This matters for two reasons. It reports no effect for video, the format this dimension leans on most, and it measures absorption rather than belief. The claim does not rest on absorption, it rests on superficial processing raising credibility, which Sundar et al. (2021) measured on credibility and sharing directly and for video specifically. This distinction keeps the mechanism precise: the dimension claims that video is processed shallowly and believed more readily, not that video transports its viewers.

In terms of Dual System Theory the other dimensions largely remove the conditions under which System 2 would engage. Content format can change what System 1 receives while it runs. A fluent, affectively charged unit of content is one that System 1 can accept with ease, whereas text is harder. A video-led feed instead delivers the feed directly to System 1, and any text-based check becomes effortful and therefore optional rather than necessary. Format does not deplete capacity or block deliberation directly, it selects

the processing mode, leaving the choice up to the designers of a platform. Chapter 4 will assess which platforms afford which types of content format, and therefore make inferences about the modality in which it is processed.

2.4.6 Deliberation Affordances: Suppressed Collective Deliberation

While other dimensions consider how a user's capacity for reflective processing is reduced or neglected, this dimension will instead focus on the user's interface and how it impacts a claim's traceability, validity, and informed engagement. Quality deliberation is the social expression of System 2. The platform can stifle critical thinking not only by hindering the individual user's reflection, but also by dismantling the communal safety net that would have caught a mistake. Deliberative theory provides the standard through the Discourse Quality Index (DQI). The DQI (Steenbergen et al., 2003), rooted in deliberative theory, evaluates discussions based on participation, mutual respect, constructiveness, and how well claims are justified. This is referenced as a standard for what quality discourse implies: justification, reciprocity, respect, and will not be used as a measurable instrument. The dimension asks whether an interface makes this criteria feasible at all. For these criteria to be met, participants must be able to justify their argument, reciprocate, and respect others, rather than stubbornly asserting opinions or dropping evidence without explanation (Steenbergen et al., 2003). Other research is largely pessimistic about mainstream social media as an ideal location for fruitful discourse and locates the cause in platform affordances rather than in users alone (Halpern & Gibbs, 2013; Shugars & Ha, 2025).

Suppression of collective deliberation essentially disables a communal System 2. In theory, social media could be a space where people could freely share ideas, challenge claims, provide evidence, and engage in democratic discourse. However, this would only be feasible if the interface supports traceable, evidence-based arguments and encourages deliberative standards. Evidence suggests that mainstream comment environments fail to meet these standards, reasonings being low justification, absent reciprocity, and incivility (Peacock, Scacco, and Stroud, 2019; Halpern and Gibbs, 2013). Without any incentive to justify an argument or reciprocate with other users respectively, there is no incentive for a collective System 2.

Where corrective context exists, it is mostly displaced onto platform warning labels, which compounds the problem rather than repairing it (Pennycook, Cannon, & Rand, 2018). Labeling a subset of content increases the perceived accuracy of whatever remains unlabeled (Pennycook, Bear, Collins, & Rand, 2020). A label also prematurely issues a verdict without explaining the underlying reasoning. Deliberative standards explicitly reject this kind of unexplained evidence (Steenbergen et al., 2003). By doing this, the platform acts as the ultimate judge of truth, rather than a neutral host for debate. A verdict can be accepted or rejected but not followed; collective System 2 is restored not by telling users what is true, but by letting them see why others hold what they hold.

One counterexample proves that this outcome is a deliberate design choice and is not an unavoidable reality. The Polis platform, used in Taiwan's vTaiwan process, produces measurable cross-group agreement through a different arrangement, presenting statements individually, removing a reply button that would invite personal attack, and surfacing the positions that find support across opposing groups rather than the ones that ignite a single side or opinion (Small et al., 2021). Furthermore, it has been deployed at national scale with input into formal legislation. Polis is not a social media platform, rather a civic consultation tool with no feed, no engagement metrics, and no entertainment function. Therefore, this example should not suggest that a mainstream platform could easily adopt its design. Instead, This

example should show that reasoned, across-the-aisle exchange is possible when the interface affords it. The deliberative shortcomings in this dimension come from platforms built for engagement rather than healthy, traceable discourse. The arrangement created by each platform is recorded in Chapter 4.

2.5 Synthesis of Dimensions, Properties, Processing Modes, and Mechanisms

This table provides a visual organization of dimensions and their mechanisms, the processing mode afforded when the mechanisms are used, and the properties that will be observed in real platforms. The process for choosing the dimensions, properties, and mechanisms is explained in Section 3.1, and observed properties under each dimension are analyzed within the context of specific platforms in Chapter 4.

Table 2.1 – Synthesis of the six dimensions: the mechanism each operates through, the processing mode it affords, and the properties observed per platform.

Dimension	Mechanism	Afforded processing mode	Properties observed
Algorithmic amplification	Algorithmic opacity → illusory truth	System 1 affording when algorithm logic is opaque.	- Default feed type - Algorithm logic and control - Feed organization
Immersive design	Cognitive load → capacity depletion	Affords System 1 when there is low entry/high exit cost into immersive feed, when autoplay is present, and when the content unit is a stream rather than card.	- Auto-advance & advance decision - Content unit - Autoplay - Entry/exit cost
Engagement metrics	Heuristic substitution	Read the count as credibility before evaluating the claim	- Likes, comments, views, circulation counts - Count precision
Friction	Inattention account	Act before evaluating	- Taps to repost - Taps to share privately - Accuracy prompt or interruption on share - Taps to report or flag
Content format	Processing fluency and affect	Absorb rather than read	- Default feed format, video playback, text truncation - Required reading - Video position in navigation

Deliberation affordances	Suppressed collective deliberation	Accept a verdict rather than follow an argument	- Clickable links - Discussion structure - Default comment ordering - Location of quoted reply - Attached corrective context
---------------------------------	------------------------------------	---	--

3. Methodology

This study is divided into three phases, each producing deliverables that become inputs to the next phase. Phase 1 is a comparative analysis of six platforms across six dimensions, conducted through a systematic walkthrough of each interface (Light et al., 2018) to establish whether individual design choices afford System 1 and/or System 2 processing. Phase 1 confirms that most mainstream platforms are System 1 dominant, and therefore generates a set of design requirements, consolidated design goals, and analytical language and references. Phase 2 draws on the design requirements and goals to build a feed prototype, coined REFLECT (Reflective Environment with Feed Legibility, Explanation, Control, and Traceability, acronym used for this study). REFLECT is intended to act as an intervention against System 1 processing and satisfy the design goals and requirements laid out in Chapter 4. Phase 3 uses the REFLECT prototype in a within-subject study in which participants use REFLECT alongside the social media platform they use the most. The results and their discussion (Chapters 6 & 7) are organized by design goals and research questions, so any finding in the evaluation can be traced back to its diagnostic claim, making this study a singular holistic argument rather than three separate studies reported in sequence.

This structure makes the work research through design, an established approach in Human Computer Interaction (HCI) in which building an artifact is itself a means of producing knowledge (Zimmerman et al., 2007). No existing platform offers a feed designed to support deliberation on dual-system principles, so the prototype had to be built before any response to it could be studied. The author designed the artifact, moderated the sessions, and coded the interviews alone. The limitations this introduces, and the mitigations adopted, are discussed in Chapter 8.

3.1 Phase 1 – Comparative Platform Analysis

Phase 1 applies six dimensions defined in Chapter 2 to six mainstream platforms through a structured, cold-state walkthrough of each platform, recording the properties under each dimension. This section will explain how the platforms and properties were chosen, the procedure for the walkthroughs, how observations were coded, and the boundaries of data collection. The output of the phase is the coded platform matrices analyzed in Chapter 4, from which the design requirements and design goals behind REFLECT are derived.

3.1.1 Operationalization of the Framework

Dual System Theory and Affordance Theory aforementioned in Chapter 2 are the analytical lens for comparative platform analysis. Dual System Theory specifies which cognitive mode operates under which conditions, and affordance theory specifies how the design arranges those conditions. Therefore, each dimension in the analysis links a design choice to the mode of processing that it affords. The System

1/System 2 distinction is adopted as an explanatory and analytical frame rather than as a pure claim about two separate cognitive mechanisms.

To develop these lenses into a testable framework, the dimensions were derived through an interactive process of feature mapping and grouping. Other frameworks were explored such as Gray et al. (2018)'s account of dark patterns, but most examples are specific to a singular dimension. While they are helpful for context this study, they do not provide sufficient breadth for use as an entire framework. Therefore, an informal inventory of prevalent social media interface properties (e.g. infinite scroll, like count, friend-endorsement, algorithmic feeds) was generated based on general familiarity with mainstream platforms. At this point the application of Dual System Theory and Affordance Theory had been ideated but not yet empirically applied.

A literature review was conducted to understand the psychological impact of these properties in an attempt to determine what cognitive mechanisms were documented in relation to how they shift user processing toward System 1 or System 2. It is through this that many of the mechanisms previously described in each dimension were discovered. When a cluster of interface properties could be anchored to a well-documented mechanism in literature, the mechanism was paired and formalized into an analytical dimension.

When the dimensions were established, the platform walkthroughs (Sections 3.1.3 and 3.1.4) were designed and conducted to measure how the dimensions assigned properties manifested across specific platforms sampled for this study. The platforms chosen and their inclusion and exclusion criteria described in Section 3.1.2.

This framework originally held only five dimensions but was refined in the design phase (Section 3.2.2) to add the sixth dimension, deliberation affordances, as suggested by an expert during design iteration. Since this final dimension added properties not explicitly collected in the first walkthroughs (e.g., discussion structure, clickable links), post-level observations were performed to complete data collection for this Phase 1, always using the same device and neutral accounts, and collecting no other data that was necessary for the sixth dimension.

3.1.2 Platform Sampling

Platforms were selected based on the focus of the study, specifically mainstream, algorithmically curated feeds through which information is frequently encountered. Sufficient design variation is also necessary to support comparison, yet similar enough to remain within the definition of “social media”, and lacking influence from overly political or professional networks.

Inclusion Criteria

- **Scale and mainstream adoption** – must have a broad user base to be socially significant as a misinformation environment.
- **Public-facing, algorithmically curated content feeds** – must present content through a feed or discovery interface driven by algorithmic selection.
- **Asynchronous content sharing as a primary function** – the platform must support the production and redistribution of content by users as a core feature.
- **Sufficient design feature variation to support cross platform comparison** – selection of platforms must support comparing more similar platforms like TikTok/Instagram, as well as more different ones like TikTok/Reddit

- **Politically and demographically balanced user bases** – cannot have a user base overly skewed towards a single political or ideological orientation that would threaten the generalizability of the analysis.
- **Accessible for a systematic walkthrough analysis** – as this is the data collection method of this study.

Exclusion Criteria

- **Politically skewed user base** – platforms that are demonstrably homogenous in ideological user composition (e.g. TruthSocial).
- **Primary function is not content feed consumption** – platforms who are capable of information sharing but the primary feature is not content feed consumption were excluded (e.g. WhatsApp, Telegram).
- **Longer-form content**– platforms whose default consumption component requires sustained, intent/search driven engagement (e.g. YouTube)
- **Professional networks** – platforms whose design positions users in an explicitly professional environment (e.g. LinkedIn, Quora).

3.1.3 Onboarding Protocol

Each platform was onboarded with a neutral account, with the same information, and under a “mandatory only” rule. This means that only the steps required to reach the main feed of each platform were completed, and every optional step was skipped. The purpose of this was to reach a cold-state feed with the least possible signal introduced by the researcher, so that the first observable feed reflected platform defaults rather than elicited preferences or a personalized algorithm. Throughout, the term “neutral” refers to this goal – introducing the least possible influence to whatever is observed. This cold-state baseline also serves a second purpose. Several dimensions, in particular algorithmic personalization and out-of-network exposure, only become observable once an account has accumulated a following and an engagement history. The neutral starting point established here is therefore the controlled baseline from which later phases could build upon.

The “mandatory only” rule meant that whenever a step could be skipped, it was skipped, and where information was required, the most neutral option available was chosen, and any such choice was recorded in the session log with its rationale. Interest and follow prompts were also followed with a strict protocol. When the platform allowed it, interest selection was skipped, and if it was not allowed to skip, broad, neutral categories were chosen (cooking, travel). Political, news, and celebrity options were avoided to prevent an overly skewed signal, and the accounts were subsequently unfollowed once the account was created. If content or accounts were required to be selected during account creation, then the minimum number was selected which avoided news, politics, and celebrities. After creation, these accounts were promptly unfollowed upon reaching the cold-state feed.

When permission prompts arose, notifications, contact access, and location were declined where possible. If there were any prompts regarding data use, ad personalization, or cookies, these were allowed in a deliberate choice as ad and cookie-based personalization are part of the cognitive environment that this study sets out to describe. Disabling these would have removed a feature of the object of analysis. Notifications were deemed outside the scope of analysis, and contact and location analyses may have impacted the desired neutrality.

The only deviation was X, which required the selection of three “packs” of accounts to follow during onboarding. The required selection was included in the screen recording, accounts were unfollowed immediately afterwards, and the episode was considered a dark-pattern observation in the analysis.

3.1.4 Cold-state Observation Procedure

Once onboarding was completed, a cold-state observation was performed to analyze existing features without the influence of a personalized algorithm or an established network. This was completed via screen recording by following a standardized sequence initialized by preparation, ensuring that the application remains closed between onboarding and cold-state observation, initiating the screen recording before tapping the app icon.

1. **Initial State Capture:** Upon opening the app on a mobile phone (methodological detail in 3.1.6), record the initial screen to document the default feed type, dominant content format, default video playback (pausing immediately to avoid feeding the algorithm), content unit size, format hierarchy, and transition type.
2. **Navigation Survey:** After capturing the initial state, a brief search around the platform was completed to understand what else is available to capture things like position of video in navigation, alternative feeds or surfaces, entry and exit cost of immersive feed.
3. **Metric Inventory:** Document which engagement metrics are visible, whether counts are exact or rounded, the text-to-visual ratio, and any text truncation or optional reading requirements.
4. **UI Flow Testing:** Test the interface friction by counting the taps required to share, repost, or flag a post (without completing the action), while noting any prompts to add a comment or view the original source.
5. **Settings Exploration:** Navigate platform settings to capture any user controls over the algorithm or feed, default content preferences, and privacy settings.
6. **Scroll Probe:** Conduct a brief scroll of the main feed to observe transition types, presence or absence of natural stopping cues, and any entry/exit friction into immersive environments.

Afterwards, review the screen recording and record findings in a table that is organized by each psychological dimension. The table content is paraphrased and used for analysis in the Chapter 4.

3.1.5 Variables and Coding

This table explains how the procedure maps to the dimensions and the properties within them, as well as how they were coded. The properties are listed under the names used in the Chapter 4 matrices; the original walkthrough tables, with their original row names and descriptive detail, are reproduced in Appendix A.

Some of the naming and specificity conventions of the properties were changed from the first walkthrough in order to provide more clarity in the tables in which they are analyzed. This was only applied in cases where a single row answered more than one question. For example, "visible metrics" was split into a row for each individual metric, and the navigation half of "format hierarchy in navigation" was separated out as "video position in navigation." All such deviations are noted alongside the original tables in Appendix A.

Deliberation affordances was added as a sixth dimension during the design phase, following expert feedback. It is mentioned here because the properties and coding were formatted the same way. Post-level observations that populate it were made after the initial walkthrough on the same accounts and devices. Because these variables are structural properties of the post interface rather than properties of feed content, they do not depend on the account's state at the time of observation.

Table 3.1 – Mapping of properties to how they were coded

Property recorded	Coding
1. Initial state capture – algorithmic amplification, immersive design, content format	
Default feed type	Algorithmic / Chronological / Choice
Dominant unit format	Card / Stream
Content unit size	Small / Large / Full screen; whether the unit enlarges when a video is opened
Default video playback	Autoplay / tap to play, in which service
2. Navigation survey – immersive design, content format	
Video position in navigation	Primary / prominent / secondary / none
Escalation to stream	None / # of taps
Immersion exit cost	Low / Moderate / High ; specification if relevant
3. Metric inventory – engagement metrics, content format	
Like count	Yes/ No; net score noted where separate totals are not shown
Comment count	Yes / No
View count	Yes / No; recorded as profile-only where it appears there alone
Save count	Yes / No
Circulation count	Type listed (reposts, shares, saves); None
Count precision	Exact / rounded, with the rounding threshold
Required reading to consume a unit	Yes / No
Text truncated by default	Yes / No, with the scope stated where it varies
4. UI flow testing – sharing and reporting friction	
Taps to share privately	Count ; Detail if relevant
Taps to repost (post publicly)	Count, Detail if relevant
Original source shown when sharing	Yes / No / Partial

Taps to report/flag	Count from feed to submission, report not completed
5. Settings exploration - algorithmic amplification	
Algorithm logic visible	Yes / No
User control over algorithm	Low / Moderate / High
Steps to reach algorithm control	Count of taps from the feed; presence of subsections noted
Chronological feed availability	Yes / No / Partial, with the scope of the partial case stated
6. Scroll probes – immersive design	
Autoplay	Yes / No, with default state and toggleability noted
Feed termination / segmentation	Yes / No, with default state and toggleability noted
7. Post-level inspection – deliberation affordances	
Clickable links	Clickable / restricted / not rendered, and where
Discussion structure	Threaded, with depth / minimal
Default comment ordering	Named as the platform labels it
Argument-traceability controls	Listed per platform
Attached corrective context	None / platform / community-contributed / platform-adjacent, described

3.1.6 Scope and Provenance

To ensure standardization and reproducibility, this section explains some boundaries and specificities regarding devices used, app versions, dates of data recording.

Modality and Device Environment. All walkthroughs were restricted to native mobile applications on iOS. Desktop, tablet, or web browser interfaces were excluded, as mobile applications represent the primary vector for immersive, continuous feed consumption, leverage distinct touch-based affordances (e.g., swipe gestures, taps instead of clicks), and are more consistently available for user consumption. All observations were conducted during May, 2026 using an iPhone 15 Pro running iOS 26.5.

Primary Object of Analysis. Data collection was strictly bound to each platform’s default feed, whichever the app automatically opened to (generally “For You” tabs). Secondary surfaces (e.g., chronological tabs, other feed types reachable via navigation bar or “For You” tabs) were inspected during settings and navigation exploration to assess user control, and the friction and transitions required to enter and exit these secondary screens were recorded. However the default feed was the one treated as the baseline consumption screen.

Versioning, Dynamic Rollouts, and A/B Testing. Observations were conducted using the latest public client builds available on the Apple App Store at the time of data collection in May, 2026. It is noted that major platforms (particularly Meta of Instagram and ByteDance of TikTok) continuously deploy server-side updates, regional feature flagging, and interface A/B testing. This means that the location

interface elements may exhibit slight variation across and exact replication of interface properties may not be perfectly possible. The observations recorded here capture the stable UI state rendered for a clean, non-personalized mobile client during the observation window.

3.2 Phase 2 – Prototype Development

Phase 2 uses the requirements from Section 4.8 to scaffold a testable artifact which will become the prototype named REFLECT. This is the research section of a research-through-design framework defined by Zimmerman et al. (2007). Chapter 5 will discuss in depth the resulting design, this section is reserved for the production process. AI was used for parts of the prototype and wireframe development, and all use cases are in Section 3.5 which covers all ethics considerations, including AI disclosures. Prompt examples in Appendix B, and the changelog of the prototype is in Appendix C.

The complete source code of the prototype, together with a README documenting its structure, is publicly available at <https://github.com/oliviakusch/REFLECT>.

3.2.1 Development Process

The initial ideation phase of the prototype began with paper prototypes, consisting of hand drawings with various ideas for potential screens for the feed, algorithm controls, and sources. Even from the first prototype, the requirements resulting from the initial comparative analysis were considered and designed against. Once a few primaries had been landed on, AI was used to bring them to life in digital wireframes that would be cleaner and more distributable to experts for review.

The prototype itself was built by building a document that explicitly outlined the design choices, locations, requirements and submitting it to Claude Code for the prototype to be built. REFLECT was implemented as a client-side HTML, CSS, and JavaScript application served from a local machine, and was opened by participants in the mobile browser of their own phone so that they would use the same device they would use for their baseline apps. There were multiple review rounds and brainstorming sessions with the supervisor, other PhD and Master students in HCI and across the Information Systems department at the University of Lausanne where the researcher would provide the current state of the prototype and ask for feedback. However, it should be noted that these expert reviewers (with the exception of the supervisor) were asked to comment on the usability of the prototype rather than on the design rationale, as they had not been involved in the comparative analysis or in the development of the requirements and design goals.

It was within these sessions that it was suggested to build out a deliberative space on the platform, similar to a comment section on existing platforms. This is what triggered the addition of the sixth dimension: deliberation affordances. Other suggestions included a moderator dashboard and action logs in order to record participants actions and provide some quantitative data (though it is limited at N=15), and most other suggestions were simply design choices like button placement and accessibility features. Once the prototype was thought to be done, a gap analysis was performed to ensure that every requirement had a feature, and every feature had an action log that was triggered upon participant interaction. A changelog was kept throughout and can be found in Appendix C. The build was frozen the day before this study began and not edited again, though a final document push was committed to the github repository after this study.

3.2.2 Content Pool

The content pool was created to fill the design and make a simulateable feed. The final set consisted of 35 items: 31 organic posts and 4 sponsored posts. Format was deliberately spread across the pool, with 14 text posts, 18 image posts and 3 short videos, so that no single modality dominated the browsing experience. The content here is intentionally civil and non-political, avoiding making any real claims that could surface any politics or points of controversy. The choice of neutral content comes with a tradeoff. Neutral content lacks the emotional, provocative, or controversial content that exists in real social media platforms. Remaining neutral was thought to be necessary for ethics clearance and ensure that no participant is more reactive to a set of content than another. The simulated community layer was written in a similar way, comments and replies were written to read casually, yet still neutrally, or at least within the context of neutral content. The feeds topic filtered were made fully functional rather than nominal, but any feed adjustments would only affect posts the participant had not seen yet. This means that though every participant is given the same feed, it can diverge once they choose to adjust it.

3.2.3 Scope Decisions

This section defines the claim ceiling and defines what the prototype is intended to test, as the development of the REFLECT prototype should include boundaries to ensure the study is feasible and coherent.

This prototype was designed to be a single-session lab artifact. There was no data persistence, onboarding, tracking of long term use, nor any backend or network architecture. As a result, everything is rendered client side and the prototype does not have an actual recommender system. Any of the transparency features that may suggest the presence of one are nominal and content is based on a fixed rule set rather than a dynamic algorithm model.

The content pool was fixed and identical across participants to maintain the possibility of direct comparison across participant experiences and study control. The community layer where there are comments, votes, and context cards, was fixed and identical as well. All of this content was pre-populated and authored by the researcher using AI to generate the content under specific neutral-but-natural instructions, and was reviewed to ensure alignment. This allowed us to evaluate the deliberative affordances, but consequently it does not perfectly reflect organic user interactions.

Any conclusions about the design are based only on the direct examination of the prototype and participants' reactions to it. Long term susceptibility effects or causal efficacy of the platform are projected at most. Further limitations are reserved for Section 8.

3.2.4 Instrumentation

To generate behavioral data analyzed later in Section 3.3.5 and Chapter 6, the REFLECT prototype included an event logging feature that would run during a participant session to collect data on their actions, such as feed views, context-panel opens source clicks, and votes. The complete tracking schema is provided in Appendix D.

All logged interaction data was compiled and downloaded as a CSV after the closure of each individual participant session, which was managed through a hidden moderator console built into the prototype

which allowed the researcher to assign session IDs, trigger CSV data exports, and reset the prototype environment between participants. Log reliability is assessed in Section 3.3.5 and 8.3.

3.3 Phase 3 – Concept Study

Unlike evaluations that introduce a single intervention into an otherwise familiar platform, the interventions here were combined in a single environment. The features were combined so that what was evaluated was the acceptability of a System 2-affording environment as a whole, rather than the feasibility of any single intervention.

3.3.1 Study Design and Rationale

The evaluation study was performed via a single-session, within-subject, mixed-methods concept study combining moderated usability testing with a comparative user-experience evaluation. It is within-subject because each participant experiences two “conditions” in one session: whichever social media application they use most frequently on their own device. Participants were pre-screened in order to limit the options to the baseline apps considered in the comparative analysis). The REFLECT prototype is considered the other condition, also used on their own device, but in-browser. An identical questionnaire was administered after each block, and a comparative interview after all else was completed.

This study is designed to be a formative concept evaluation. It should not be interpreted to measure efficacy as truth discernment is not measured, nor does this claim to prove a causal shift from System 1 to System 2 processing. This study is intended to evaluate whether users accept this redesigned feed and consider it to be usable, comprehensible, and appealing enough to adopt. In other words, the objective is to understand whether the added deliberation registers as acceptable and if users feel it would change their behavior (RQ2).

3.3.2 Participant Selection

Screening. Screening used a Qualtrics prescreener fielded neutrally as a “study registration questionnaire” through HEC-LABEX from 21 July 2026. The instrument was unpaid, under five minutes, and opened with a consent page carrying an explicit agree/decline branch, and it was noted that only the lab session provided a 25CHF compensation. Seven substantive items were presented on a five-point agreement scale with statements regarding compulsive use, information-seeking behavior, feed curation, confidence in adopting new apps, an attention check question, and an attitudinal item on whether they believe social media causes societal harm. The screener also asked for their most-used platform, an occupation and prior-participation multi-select, availability, and contact fields. The details of the questionnaire including the questions themselves are held in Appendix E. Participants were excluded if they did not consent, reported non-use of social media, failed the attention check in the survey, and availability. Of 51 responses, 37 were eligible, with seven excluded for failing the attention check, three as non-users of social media, and four for unavailability. Insider knowledge and recent participation in similar studies were initially excluded but reclassified as flags and taken on a case by case basis. Availability across the ten listed session dates was a selection criterion, with breadth of availability used as a tie-breaker. LABEX handled all invitations and booking, and contact data were separated end-to-end from screening answers.

Personas. To assess a diversity of thought across the sample, the screener attempted to group respondents into three preliminary personas: Reflexive Scrollers (heavy, habitual users), Skeptics (users who are skeptical of what they see online and like to fact check), and Technologically Constrained (users who report difficulty in learning a new app). In practice, the questionnaire yielded less variance in responses than anticipated. Participants were categorized to a persona as faithfully to their survey data as possible, but the limited variation meant the boundaries between the groups were less distinct than originally intended. Given the small target size of the groups being only four to five per persona, the cohorts were inherently too small to support meaningful comparative analysis regardless. Therefore, these personas functioned exclusively as a heuristic to maximize diversity during recruitment.

3.3.3 Baseline vs REFLECT conditions

In Condition A, participants used a social application they use regularly, on their own device and account, completing the task set while thinking aloud. They were informed that any content on their feed need not be shared with us as the intention is to only study their mode of usage, not the content they consume. A study phone with blank accounts was provisioned as a fallback for participants without a phone or unwilling to use their own account; it was not required.

In Condition B, the participant used REFLECT on a fixed content pool of 35 posts identical for every participant; topic filters act on unseen posts only, so the early feed environment remained shared across participants. The prototype was served locally in the session room, with the participant's device connected to the moderator's machine; event logging was confirmed as part of the pre-session checklist and logs were downloaded and archived per participant at session close.

3.3.4 Procedure

The procedure for each participant session was followed as such:

Welcome, consent, icebreakers	4 mins
Block 1 – four tasks on one condition, think aloud	11 mins
UEQ-S	3 mins
Block 2 – same four tasks on other condition, think aloud	11 mins
UEQ-S	3 mins
Questionnaire	3 mins
Comparative Interview	10 mins
Design exercise (if prompted & time permitting)	(3 mins)
Debrief, cash form, close	2 mins

Both blocks used the same four hand-carded tasks:

Task (goal)	Hand card, as used	Observed
-------------	--------------------	----------

T1 Browse (discover)	“Browse the feed the way you normally would. Stop whenever you’d naturally stop.”	Posts to stopping; stop reason; immersion remarks. Stopping was observed, never instructed.
T2 Judge (judge / understand recommendation)	“Find a post that makes a claim. Decide whether you’d trust it, and walk me through how you’d check it.”	Verification behaviour; stated trust judgment and rationale; whether a reporting path is noticed unprompted (one neutral probe if not).
T3 Interact (share + interact, merged)	“Show me how you normally interact with your feed, this can be liking, commenting, sharing”	First interaction chosen; the share flow including any hesitation; likes, votes, comments within the same task.
T4 Adjust (manage visibility / control)	“Imagine you’ve been seeing too much of one topic lately. Show me how you would change what your feed shows you.”	Whether controls are found without pointing; the path taken; adjustments made; comments on feeling in or out of control.

Immediately after each block, the participant completed the UEQ (User Experience Questionnaire) paper questionnaire for the application just used. After both blocks, a semi-structured comparative interview of approximately ten minutes was conducted. Some questions in the interview were fixed, but the researcher kept question banks from so that the interview could drift depending on what the participant mentioned in the blocks. The moderator guide with question banks is provided in full in Appendix F.

When time allowed and the participant had spontaneously offered design suggestions, a short design exercise followed, the participant would be encouraged to draw a two-minute sketch or description of one change they would make to their own social application. Because administration was conditional, sketches are treated as opportunistic qualitative material, not per-participant. Every session closed with a scripted full debrief disclosing the research purpose, the curated content, and the fact that no participant action left the room.

3.3.5 Measures

Two types of measures were collected. First, the UEQ-S (Schrepp et al., 2017), a short questionnaire of eight paired-adjective items that produce pragmatic and hedonic quality scales. These were administered on paper immediately after each block with identical wording, scored with the official scoring tool. The UEQ-S’s primary scores were the comparisons of REFLECT versus participants’ own apps, but a published benchmark serves as a secondary, external anchor for the prototype’s absolute scores.

The UEQ-S was the sole questionnaire instrument and therefore speaks directly to RQ2; the constructs specific to RQ1 and RQ3 — comprehension of design intent and felt control — carry no questionnaire measure and are evidenced through the comparative interview, the behavioural logs, and the qualitative coding described in Section 3.4.

Second, action logs in the prototype condition only: context-panel opens, source clicks, the share funnel (initiated → accuracy prompt → completed or abandoned) with hesitation timestamps, votes, comments

and likes, control use, posts-to-stop, and stopping-cue behaviour, exported per participant in CSV and JSON.

Interaction logs exist only for the prototype. Since data on participant's baseline apps is only reported or observed, interaction logs are not used in comparison to the baseline, only for quantifying the frequency certain actions where relevant. Log analyses also proved to not be entirely reflective of whether a task was understood to completion. For example, participants frequently initiated but did not complete the sharing process. This hesitation could stem from their awareness that the environment was simulated, rather than a lack of understanding of how to execute the action.

3.4 Coding and Analysis

The analysis of the REFLECT prototype relied on a mixed-methods approach, triangulating qualitative interview data, quantitative self-reports, and descriptive interaction logs to generate comprehensive design insights. Because the sample size (N=15) inherently limits statistical power, the analysis was qualitatively driven, with quantitative metrics serving to contextualize and cross-verify participant statements.

3.4.1 Qualitative Coding and Thematic Analysis

Once all participant sessions were completed, the audio recordings were submitted to UniL's Corv, a private transcription tool, to ensure data security. To facilitate the coding process, an Excel spreadsheet template was developed with the assistance of AI (Claude). This template features designated dropdown cells for participant quotes, memos, primary and secondary codes that map to the six design dimensions, participant stance, and evidence type. It also included tags to flag specific usability issues, feature suggestions, or mentions of design novelty. To protect participant privacy and prevent feeding sensitive data into external AI models, all transcriptions were manually copied into this local spreadsheet, and all coding was completed entirely by hand.

The transcripts were analysed using a codebook developed prior to coding that was based on the six design dimensions of REFLECT, while remaining open to emergent themes regarding usability and design novelty. The researcher was the sole coder, conducting the analysis by applying high-level dimension codes and usability tags, and then extracting specific quotes, synthesizing memos, and evaluating the participant's stance. A summary tab within the spreadsheet automatically aggregated these inputs, allowing the data to be easily sorted by dimension, stance, and evidence type, alongside the corresponding participant ID and quote. The schemas for both the coding and summary spreadsheets, as well as filled examples, are provided in Appendix G.

3.4.2 Quantitative and Log Analysis

To capture standardized usability and user experience metrics, the Short User Experience Questionnaire (UEQ-S) was administered for both the participants' baseline apps and the REFLECT prototype. The data was processed using the official UEQ data analysis templates to compare REFLECT against the benchmark and directly against the participants' baseline sessions.

Additionally, system interaction logs were summarized to calculate the count of participants who completed specific actions (e.g., opening the context panel, clicking on sources, or accessing the algorithm controls). Given the single-session lab environment, these logs were analysed at a descriptive, generalized level. As noted in Section 3.2.5, the prototype logged certain tasks such as initiating a share or a report that participants deliberately abandoned because they knew the environment was simulated.

Therefore, the analysis erred on the side of over-collecting log data but specifically filtered for intent-driven actions (e.g., verifying that all participants successfully opened the control panels).

3.5 Ethics

This study was reviewed and approved by the University Commission for Research Ethics of the University of Lausanne (submission HEC-6a4f968737e03) which covered a data collection period of 20 July to 7 August 2026. Approval was required to begin any aspect of this study, including screeners or recruiting. Participants were compensated 25CHF in cash for a 45 minute session.

Recruitment was run through HEC LABEX participant pool, and consent to participate was performed once in the screening questionnaire and again before the beginning of a session, where the paper consent form was kept and all forms were returned to LABEX at the end of the study. Consent covered the tasks to be performed, the data to be recorded, the voluntary nature of participation, and the right to withdraw at any time without consequence. A separate cash form was signed before the compensation was distributed as well and also returned to LABEX at the end of the study.

Participants were not informed of any specific research objective in advance. The session was introduced as a usability evaluation of a prototype feed so that behavior during the task was not encouraged by the reflective design framing. No false information or inflammatory content was given at any point, so the protocol was not deceptive, nor did it pose any type of physical, mental, or health risk. The only foreseeable discomfort was self-consciousness during the thinkaloud, mitigated by keeping sessions short and reminding participants that they could pause, skip a task, or stop at any time.

All data came directly from participants, no demographic or special-category data was collected. Participants were screened on self-reported behavior and the only information retained by the researcher is the participant code. The project involves no personal or sensitive data within the meaning of the Loi sur la protection des données personnelles du canton de Vaud, no health-related data under the Federal Act on Research involving Human Beings, no human biological material, and no images or video of recognisable persons. Data collection took place in Switzerland, no data were shared with institutions outside the project, and all results are reported in anonymised or aggregate form.

The baseline task block was added after submission, and it was conducted on the participant's own device and account in order to establish a baseline against the prototype. This did not extend the approved scope nor require an amendment. No screen recording was made and no feed content, account details, contacts or third-party posts were captured; observation was limited to field notes, and the material collected was confined to the participant's own account of their ordinary use, which they were free to give or withhold. Questions in this block concerned routine habits rather than sensitive disclosure.

3.5.1 AI Disclosure

Generative AI tools were used at several points in the production of this research. Specifically, Claude Cowork was used, primarily Opus, but Fable was used for coding the prototype.

In the initial comparative analysis phase, it was used for suggesting relevant key words to aid in searching Google Scholar for relevant literature.

In the design phase, they supported the iteration of wireframes and the implementation of the prototype, for which a complete change log is provided in Appendix C. They were also used in curating the content pool from which the stimulus posts were assembled.

In preparing the evaluation, they were used to draft a document of the study protocol from the author's own inputs, for organisation and scripting, and to ideate candidate questions to put to participants. The final protocol, task set and interview guide were determined by the author.

In the analysis phase, their use was confined to infrastructure. The spreadsheets used to summarise the action logs and to hold the interview coding were built with AI assistance, but no participant data was supplied to or processed by an AI tool in their construction, and every coding decision was made by the author alone. AI was further used to check the organisation of the analysis chapters against reporting conventions in HCI, for formatting, and to render existing material in table form, such as the mapping between variables and codes.

Throughout the paper, AI would be occasionally consulted for phrasing or thought-organization suggestions, but in these cases the author would consistently rewrite to ensure their own voice was what was represented in the prose and to achieve fluency with the rest of the work.

No finding, interpretation, analytical claim, or direct prose presented in this thesis was generated by an AI system. The author takes full responsibility for the content of the work.

4. Comparative Platform Analysis

This chapter presents the comparative analysis of the chosen platforms across the identified dimensions as a diagnosis of the environment and what interface properties withhold System 2 processing.

The properties must be identified across existing platforms in order to design an alternative that can credibly claim to achieve a different end. In other words, a design that claims to support critical thinking is more meaningful if designed against a specified and research-backed account of what suppresses it. The analysis therefore examines the six chosen platforms across the dimensions described in Chapter 2 and each is read against the mechanism it was paired with, as well as a statement about the mode of processing it supports.

Each section is organized the same way. A property matrix records how the properties of the headline dimension are surfaced across each platform, followed by an account of what varies and what platforms cluster together. Then, an interpretation of what the variation does to the corresponding mechanism, and finally a connection back to Dual System Theory and the design requirements implied by each analysis. The synthesis in Section 4.7 draws the dimensions together, and the consolidated requirements and design goals are the liaison to Section 5, which draws on them to develop the final design that was used in the qualitative study.

4.1 Algorithmic Amplification

Table 4.1 – analysis of properties of algorithmic amplification dimension across the six platforms

Variable / Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Default feed type	Algorithmic	Algorithmic	Algorithmic	Algorithmic	Algorithmic	Algorithmic
User control over algorithm	Low	Low	Moderate.	Low	High	High
Taps to reach algorithm control	3	3	3, many subsections	4	4, many subsections	4, many subsections
Chronological feed availability	No	No	Partial, replies only	Partial, replies only	No	No

In this matrix, it is shown that on every platform, the default feed type is algorithmic, with no algorithm logic visible. TikTok and Instagram cluster together as having higher user control over the algorithm, but these controls are also buried much deeper in settings (3-4 taps, often with multiple subsections for specific controls) which carries high friction. Furthermore, across all of the platforms there is no clear incentive to adjust this in a way that would challenge a user's thinking about the content they are ingesting. The control that does exist is a double-edged sword in the sense that users that are adding their own preferences will often reinforce their own beliefs, narrow their feed further, and intensify the illusory truth mechanism rather than interrupting it. This is strongest where the control is highest. TikTok and Instagram give the user the most say over what the algorithm shows them, so they also give the most room to narrow the feed. Even if a user does reach these controls, using them does not necessarily protect them from illusory truth.

As explained in 2.4.1, algorithmic curation is what provides the repeated exposure that affords illusory truth, and this is compounded by curation logic being opaque. Any feed controls that could invert this are buried under friction, and support reinforcing curation, so a corrective shift may never happen. This logic directly aligns with the illusory truth mechanism and this familiarity reads as truth. This directly supports System 1 processing.

This dimension therefore yields two requirements:

Transparent, adjustable feed – explanations for why a post appears and offers in-place controls to diversify or refresh the feed and reduce repetition.

Cross-viewpoint transparency – the user should have reason to believe that the platform learns topics rather than opinions, and warns against repeated exposure.

4.2 Immersive Design

Table 4.2 – analysis of properties of immersive design dimension across the six platforms

Variable / → Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Dominant unit format	Card	Card	Card	Card	Stream	Card / Stream
Feed termination / stopping cue	No	No	No	No, but in/out tap between units	No	No
Escalation to stream (entry cost)	One tap → video stream	One tap → video stream	Stream in separate location	No stream	Stream is default	One tap → video stream
Autoplay	Yes, in-feed	Yes, in-feed	Yes, in-feed	Yes, by default, but toggleable	Yes, core mode	Yes, in-feed
Immersion exit cost	Low: swipe out	Low: swipe out	Low-moderate: two swipes	N/A	High: must close app	Moderate: swipe out, but back into video heavy feed

All platforms except TikTok carry a card feed, but X, Facebook, and Instagram still offer a singular tap escalation to stream, and Bluesky still keeps the stream, just in a separate part of the platform. Reddit is the outlier that does not carry any stream at all. Autoplay is present everywhere, and exit cost is only high or moderate for TikTok and Instagram, which are respectively stream by default and or still exit into heavily video-forward unit format.

Through the lens of cognitive load and capacity depletion, the immersiveness of the default format and escalation to stream feed should not be understated. This decision boundary for users' entry into the immersive infinite scroll environment is tiny, and especially in the cases of TikTok and Instagram, the exit costs to leave are higher. This clears the path for the continuous stream, which is what generates the cognitive load, and the asymmetry between the entry and exit sustains it, especially in the case of TikTok where you must close the entire application to get out. As explained in Section 2.4.2., processing power occupied by the continuous stream is therefore processing power not available to evaluate any single item within it. This depletion is not a side effect, it is the primary mechanism and it directly affords System 1. Content format is another dimension that severely compounds the cognitive load, capacity depletion, and System 1 affordance, and will be discussed further in Section 4.8.

This dimension therefore yields two requirements:

A bounded, segmented feed with a terminal state – returning decision to continue to the user at every boundary rather than holding it through continuous cognitive load and the feed offers no escalation into a full-screen stream.

No escalation to stream– videos do not autoplay and the feed does not default to full screen; captions, claims, and context are shown by default rather than hidden or truncated.

4.3 Engagement Metrics

Table 4.3 – analysis of properties of engagement metric dimension across the six platforms

Variable / → Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Like count	Yes	Yes	Yes	Net score up/downvote	Yes	Yes
Comment count	Yes	Yes	Yes	Yes	Yes	Yes
View count	Yes	No	No	No	Profile only	Profile only
Save count	No	No	No	No	Yes	Yes
Circulation count	Reposts	Shares, saves	Reposts	None	Shares, saves	Reposts, shares
Count precision	Rounded >1k	Rounded >1k	Rounded >1k	Rounded >1k	Exact to 10k	Exact to 10k

Every platform surfaces aggregate interaction counts next to content, so the social layer is visible before the user evaluates anything. The variation is in which counts are visible and how they are framed. Reddit is the clearest departure, showing a net score rather than separate totals and hiding the downvote count entirely, so the displayed number gives a directional verdict rather than raw volume. TikTok and Instagram foreground saves and shares alongside likes, which read as signals of virality rather than simple approval. X gives prominence to view counts, which is a reach signal, though this is also visible on TikTok and Instagram on profile view, where it could be read as a virality or credibility signal. Count precision varies as well, with X, Bluesky, Facebook, and Reddit rounding above 1k while TikTok and Instagram are exact to 10k.

As described in Section 2.4.3, these metrics are signals that are inputs to the heuristic substitution. Across these platforms, this is particularly pronounced in the case of Instagram and TikTok where almost every metric is visible and with more specificity, and views counts are also visible on users profiles, which can encourage a judgment about the source of a video or claim rather than simply the video or claim itself. Section 2.4.3 established that the more significant heuristic is the source and expertise cue rather than popularity count. What TikTok and Instagram make visible at the account level is still a popularity count, so though this is still a reach signal and not necessarily a credibility one, it still introduces another opportunity for substitution.

It should be noted that evidence on raw engagement counts is mixed. Meinert (2020) suggested the possibility that higher like and share counts lowered perceived credibility. He and Zheng (2025) found that their influence is conditional rather than direct. Morris et al.'s (2012) participants reported relying on

author-level cues such as follower counts. The requirement is therefore kept high-level, and it is robust to any of these readings. If counts are irrelevant or harmful, de-emphasizing them costs nothing. If they substitute for evaluation, de-emphasis is protective.

Engagement signals presented neutrally and not without possibility of evaluation – engagement counts are de-emphasized and if they exist, it is next to an actual evaluation opportunity

4.4 Sharing and Reporting Friction

Table 4.4 – analysis of properties of sharing and reporting friction dimension across the six platforms

Variable / → Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Taps to repos (post publicly)	2	2	2	2+, must choose community	1	1
Taps to share privately	2	3, poster settings toggle whether others can share	3	2	1 (slide send to recents) or 3	1 (slide send to recents) or 3
Accuracy prompt or interruption on sharing	No	No	No	No	No	No
Taps to report or flag ¹	3+	3+	3+	5+	2+	2+

Friction is the effort, steps, or interruption built into an action, and this dimension analyzes the placement of it. In this case, the analysis covers the number of taps to repost publicly, share privately, whether an information cue is shown when sharing, and the number of taps to report or flag a post. Following, unfollowing, and muting creators are excluded because the neutral account and absent network effect make them unobservable.

Across all applications, the direction is the same: there is essentially minimal friction to sharing publicly, little friction to share privately, no accuracy prompt or source shown when sharing, and much more friction when reporting or flagging. More specifically however, Reddit is an outlier as the user must choose which community to share it to, which requires much more deliberation and is essentially an invite for discussion as opposed to a “share to all followers” or with a specific person that would be found on any of the other platforms. Reddit is frictionally expensive in both directions, therefore doesn’t hold the

¹ “+” implies at least one more tap would have to be done to complete the report or that the number of steps depended on the reason for reporting. Reports were not sent in order to maintain neutrality on the accounts used for the study.

asymmetry the same way because it is the most expensive reflective path in the set. The amount of friction and the direction of it are two different things. This is seen in the other direction in the cases of Instagram and TikTok with one tap to repost, publicly or privately depending on who you send it to, showing the opposite case of zero friction on a share.

These placements feed the inattention account which holds that people share content without having evaluated it, not because they cannot discern the truth but because their inattention is elsewhere when they share it. None of the platforms introduce any suggestion on the share path that would move a user's attention. Reddit introduces more share friction than the other platforms, but the friction is still related to the destination, not the accuracy of the content itself. As described in Section 2.4.4, Pennycook et al. (2021) found that even a small prompt directing attention to accuracy curbed the share of false content. Though veracity is not discussed in this paper, this maintains relevance as it shows that although there are existing interventions that would make sharing more intentional, they are not a part of any of these mainstream platforms.

In Dual System Theory, System 2 is off by default and engages only when spare capacity is available or something prompts it to act. It is seen here that neither of these are introduced in sharing, so sharing becomes a System 1 action that is already completed before System 2 has the opportunity to insert itself. It should also be noted that friction is a cross-cutting feature across the other dimensions as well, and its compounding effects will be discussed in the synthesis. This dimension therefore yields the following requirement:

Intentional sharing – does not allow single-tap sharing

Encouraged verification – where a share exists, prompt directs attention to checking accuracy before a claim is accepted or acted on.

4.5 Content Format

Table 4.5 – analysis of properties of content format dimension across the six platforms

Variable / → Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Dominant unit format	Mixed photo/text or video	Mixed photo/text or video	Mixed photo/text or video	Text	Video	Photo/video
Default video playback	Autoplay	Autoplay	Autoplay	Autoplay	Autoplay	Autoplay
Required reading to consume a unit	Sometimes	Sometimes	Sometimes	Yes	No	No
Text truncated by default	Yes	Yes	Yes	Only for long posts	Yes	Yes

Variable / → Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Video position in navigation	Prominent	Prominent	Secondary	None	Primary	Prominent

Content format should not be confused with immersive design, which examines the structure of the feed and how users move through it. Content format as it is analyzed here considers what sits inside each unit of content in a feed. Across these properties, autoplay is the default on every platform. Reddit is the only platform for which there is a larger text limit before truncation, and in this case the truncation is still seen as a preview and is accompanied by the fact that a user is required to read in order to consume the post. On the other end of the spectrum, TikTok and Instagram are much more video dominant in terms of unit format and position in navigation. X, Facebook, and Bluesky are somewhere in between: X and Facebook are labeled as video prominent due to the infinite video scroll starting once you tap into one video, whereas Bluesky holds this feature in a completely separate part of the platform.

These properties feed processing fluency and affect. Fluency refers to the ease of consuming content, but the problem is that this ease can subconsciously feel like truth. Since autoplay is the default across every platform, the unit begins before the user has even decided to attend to it, making this a design choice by the platform rather than a choice made by the reader. Video is the most fluent format as it does not require reading, in which case the reader would be able to set the pace of consumption, making it a more intentional action rather than unintentional consumption. This is why Reddit sits so far away from TikTok and Instagram on the Dual System Theory spectrum in this case, and why the prominence in X, Facebook, and Bluesky to an extent, are not innocent either.

In a context where the content decides the pace of the consumption of a unit (video-dominant feeds and formats), there is no natural pause where a System 2 intervention could fit. Affect compounds this because video can create an emotional charge much more quickly, and elicit feelings that stand in for evaluation. The more dominant processing mode afforded by the content format across most of these platforms is to absorb rather than to read, and it happens before a user has the cognitive space to make a decision. The next requirement is defined as such:

Format-balanced, pre-curated content pool – the feed is populated from a fixed set of neutral, non-partisan posts balanced across text, image, and video without autoplay.

4.6 Deliberation Affordances

Table 4.6 – analysis of properties of deliberation affordances across the six platforms

Variable / Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Clickable links	Yes	Yes	Yes	Yes	No	No
Discussion structure	Threaded	Threaded to 2 levels	Threaded	Deep threaded	Threaded to 1 level	Threaded to 1 level

Variable / Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Default comment ordering	Engagement	Engagement	Engagement	Engagement	Engagement	Engagement
Location of quoted reply	New feed post	New feed post or story	New feed post	In thread	New video	New video or story
Attached corrective context	Community	Platform	Platform adjacent	Assigned moderators or content filters for some subreddits	Platform	Platform

This dimension questions whether the interface allows users to trace claims, check where it came from, and participate in traceable discussion with other users. The variables analysed here are whether links are clickable, how deeply discussion threads, how comments are ordered by default, where a quoted reply is placed, and what corrective context is attached to a post. The biggest and possibly most consequential diversion across the platforms is the attachment of links, which is not a supported feature on TikTok or Instagram. Conveniently, these platforms are also the ones that afford the least number of threads supported in the comment section. The combination of lack of context or sources, no threading for traceable arguments, not to mention engagement based comment sorting, is fatal to community deliberation. On the other hand, Reddit allows for the deepest threading on a post, and has assigned moderators and/or content filters for some subreddits, which help to ensure that discourse is relevant and not abusive.

When wielded incorrectly, properties suppress collective deliberation which is essentially the social expression of System 2. In System 1 environments the steps of following an argument are removed one at a time: the source cannot be reached, the exchange cannot be followed, the rebuttal cannot sit beside the initial claim or a particular response, and what remains is a verdict with no reasoning attached. Comment ordering compounds this, because ranking by engagement is either algorithm or engagement heuristics re-entering at the level of discussion, so a thread is a ranked list rather than a record of an argument. Factual corrections may attract less engagement than the claim it corrects. If it is ranked beneath the claim, the reply that would have prompted a second look exists but is not visible due to the design choices. Suppressing collective deliberation disables a communal System 2, so the platform stifles critical thinking not only by impeding individual reflection but by dismantling the safety net that would have caught the mistake. The final requirements are defined as such:

Source, viewpoint, summary on each post – each post opens to one screen carrying the claim, the range of views, contributor-added clickable sources each paired with a short argument, and the discussion.

Deliberative discussion surface – discussion is structured for reciprocity and justification, is not ordered purely by engagement, and its tone is a controlled property of the curated content.

4.7 Synthesis

The six dimensions described are not independent, but an *interdependent* system that works on each of the platforms analyzed in this paper. The platforms' combination of these dimensions are what create a cognitive environment from the second you open the application, and in many cases, the dimensions amplify each other, creating a deeper environment than if any one of them appeared on their own.

4.7.1 Compounding and Interdependencies of the Dimensions

Depletion of cognitive resources is the precondition for System 1. Immersive design features, especially infinite scroll and dark patterns that obstruct exit, occupy working memory first and fastest. By continuously capturing attention, they exploit the "laziness" of System 2, which engages only when something prompts it to, and here nothing does, because the capacity to deliberate is already drained. As Maranges, Schmeichel, and Baumeister (2017) show, cognitive load works by distracting attention away from deliberate processing.

With System 2 effectively sidelined, the platform's other mechanisms operate unopposed. The opaque algorithm repeatedly surfaces familiar content, and without the capacity to evaluate it, users fall back on the familiarity-as-truth heuristic, so prior exposure actively raises how accurate a post feels (Pennycook, Cannon, and Rand, 2018). Visible engagement metrics add an immediate credibility shortcut, and video-heavy formats that require no reading bypass evaluation through processing fluency.

Friction asymmetry then converts this depleted state into action, because when sharing takes little effort, it can be completed before any accuracy check occurs, and the user shares through inattention rather than belief.

The illusory truth effect, the social-proof shortcut, and one-tap sharing do not carry the same cognitive weight on a rested, deliberating user. The initial, structurally induced depletion is what gives the rest of the environment its power, which is why this dimension functions as the base of the system rather than as one mechanism among five.

The dependencies also run between the mechanisms that do not force depletion. Algorithm repetition only converts to belief because the curation is invisible, so familiarity is not attributed to the repetition and is therefore read as evidence of truth. (Pennycook, Cannon, Rand, 2018). Similarly, friction asymmetry only matters because attention is already elsewhere, since a frictionless share converts to an action exactly when users are not inclined to consider accuracy (Pennycook et al. 2021). Compounding therefore runs in two directions at once: depletion enables some, and those enforce each other within the depleted state.

It should be made clear that this sequence is explained as a way to describe how these effects can compound. Not every dimension or feature is present in every single platform, but it remains true that various combinations of the System 1 affording side of the dimensions can contribute to the creation of an overloaded cognitive environment that does not allow for System 2 thinking. The subsequent section will explain these clusters and perform a more thorough cross-platform analysis.

4.7.2 How Platforms Cluster

If the six dimensions were completely independent of one another, the platforms in this study would scatter across them. Each platform would land high on some dimensions and low on others in no particular pattern and there would be no particular trend. Our observations show the opposite: many of the mechanisms move together at the platform level: TikTok and Instagram afford System 1 most consistently (heavier video format, easier sharing, autoplay, more friction asymmetry) while Reddit is at or near the low end across nearly every dimension (text heavy, less sharing options, no autoplay).

Table 4.7 – Overall ranking of intensity of System 1 processing affordance across all dimensions and platforms.²

Dimension	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Algorithmic Amplification ³	Moderate	High	Moderate	High	Moderate	Moderate
Immersive design	Moderate	Moderate	Moderate	Low	High	High
Engagement and sharing mechanics	Moderate-to-high	Moderate	Moderate	Low-to-moderate	High	High
Friction	Moderate-to-high	Moderate	Moderate	Low	High	High
Content format	Moderate	Moderate-to-high	Moderate	Low	High	High
Deliberation affordances	Moderate	Moderate	Moderate	Low	High	High

The summary platform shows that there are not necessarily two rigid categories, and the platforms sit along a continuum from environments that structurally permit System 2 to those that actively attempt to remove the conditions for it. There are some extremes, such as TikTok, which remains near the highest end of this continuum, with video-only formats, frictionless sharing, complete autoplay, and high exit friction. At the opposite end sits Reddit, who has extremely text heavy formats, limited sharing capabilities, and requires more intense deliberation to completely comprehend and reply to a post.

² Ratings indicate the degree to which each dimension's design affords System 1 processing (High = strongly affords heuristic processing). E.x., for the friction row, "High" denotes the absence of friction affording System 1, not to be confused with having "High" friction.

³ It should be noted again that in the context of this study, algorithmic amplification is limited to the controls visible within the platform – there is no information as to what data is used and how in order to build algorithms, so the values here could be over or understated and are only intended to be reflective of the data collected in this study.

4.8 Design Goals and Requirements

DG1 – Protect cognitive capacity

Return natural boundaries to the feed and keep text/context available rather than defaulting to autoplay and full-screen immersion. The comparative analysis identifies that many platforms are immersive, lack boundaries, and are frequently driven by dark patterns and autoplay which depletes System 2 capacity. The new design should limit or exclude these features in order to restore System 2 capacity.

A bounded, segmented feed with a terminal state – returning decision to continue to the user at every boundary rather than holding it through continuous cognitive load and the feed offers no escalation into a full-screen stream.

No escalation to stream– videos do not autoplay and the feed does not default to full screen; captions, claims, and context are shown by default rather than hidden or truncated.

Format-balanced, pre-curated content pool – the feed is populated from a fixed set of neutral, non-partisan posts balanced across text, image, and video without autoplay.

DG2 – Support transparent, governable curation

Make feed logic visible and give users real, low-effort control over what they see while guarding against illusory truth (i.e. warn against self-selected repetition). The comparative analysis found on some platforms, curation logic was invisible and any controls were hidden behind several layers of preferences and settings. Research showed that algorithmic transparency protected against misinformation susceptibility and should be designed accordingly.

Transparent, adjustable feed – explanations for why a post appears and offers in-place controls to diversify or refresh the feed and reduce repetition.

Cross-viewpoint transparency – the user should have reason to believe that the platform learns topics rather than opinions, and warns against repeated exposure.

DG3 – Restore evaluation at point of judgment

Make the origin of a claim visible and clickable. Make the verification of a claim easy and encouraged. Present engagement signals neutrally or at least have them refer to something that they are encouraged to consider. Comparative analysis found that social proof counts next to content before it's evaluated and is difficult or impossible to directly attach a source on many platforms. The design should rethink engagement metrics and support sources and argument summaries.

Engagement signals presented neutrally and not without possibility of evaluation – engagement counts are de-emphasized and if they exist, it is next to an actual evaluation opportunity.

Source, viewpoint, summary on each post – each post opens to one screen carrying the claim, the range of views, contributor-added clickable sources each paired with a short argument, and the discussion.

Deliberative discussion surface – discussion is structured for reciprocity and justification, is not ordered purely by engagement, and its tone is a controlled property of the curated content.

DG4 – Action should be intentional and calibrated

Deliberation should be encouraged within the context of sharing and scrolling. The comparative analysis found that sharing was sometimes as little as 1 tap, which does not create enough of a decision boundary to encourage System 2. Though friction asymmetry is relevant, it cannot be accurately assessed within the boundaries of this study.

Intentional sharing – does not allow single-tap sharing.

Encouraged verification – where a share exists, prompt directs attention to checking accuracy before a claim is accepted or acted on.

Non-functional requirements

NFR1. Familiar enough to adopt – the interface departs from mainstream patterns only where a functional requirement demands it.

NFR2. Environment authenticity – posts, comments, and community have to read as a real platform and not study stimuli.

NFR3. Legible at first contact – every mechanism must be discoverable and interpretable without onboarding, instruction, or documentation within a participants first session.

NFR4. Robust for unassisted, single-session use on participant devices – the prototype runs in the browser of a participants own phone with no installation or account, tolerates all browser types and operating systems, and preserves session state across page reload.

5. Design

This chapter presents REFLECT, the feed environment built from the diagnosis in Chapter 4, and gives the reasoning behind each of its features. It is organised by design goal rather than by screen, so that the argument runs based on design goal and requirements.

A property recorded across the platforms analysed becomes a requirement, requirements consolidate into a design goal, and the goal is met by a feature in the prototype. Each design choice can be therefore traced along the chain of dimension finding → requirement → design goal → feature.

The sections that follow take each design goal in turn and explain why a feature takes the form it does, and which alternatives were available and rejected.

Participants’ reaction to the prototype is reserved for Chapter 6, and is not asserted in this chapter. The constraints under which the prototype was built: a single session, no network or persistent account, a fixed content pool identical for every participant are set out in Section 3.2.2, and the features they left nominal are identified where they arise and revisited in Section 5.2.

5.1 REFLECT Environment

5.1.1 DG1– Protect Cognitive Capacity

The first design goal responds to the immersive design and content format dimensions (Sections 4.2, 4.5) where autoplay, truncated text, and easy escalation to stream format were found to occupy mental capacity before evaluation of content can begin. The corresponding requirements are a bounded, segmented feed with a terminal state, no escalation to stream, and neutral, format-balanced, pre-curated content pool. The content pool was created after the scaffolding of the prototype, so this requirement is satisfied by the nature of the methodology. Two primary screens contain the features that were designed in response to the other two requirements, and they are explained below:

Content unit. Figure 5.1 shows that in the prototype, every post presents the claim in full with no truncation. If a video is attached, it does not play automatically when the screen is open to it. Though video content is generally considered to carry high fluency and therefore can be System 1 affording, it would deviate too much from current platforms and was thought to possibly become means of rejection of the platform if no video content was possible at all.

Feed and boundaries. The feed is finite and is delivered in segments – at first offering a decision boundary for whether to keep going (Figure 5.2), and then a termination state at the end of the feed (5.3). It is possible a timer or time spent warning could achieve the same goal, but these address user habits and not feed supply which is easier to manage from a platform level when curating a healthier feed. “That’s the day’s main topics covered” returns the decision to continue without making it about the user’s habits. The continue control is deliberate as a boundary that could not be crossed would obstruct to a point that would frustrate a user.

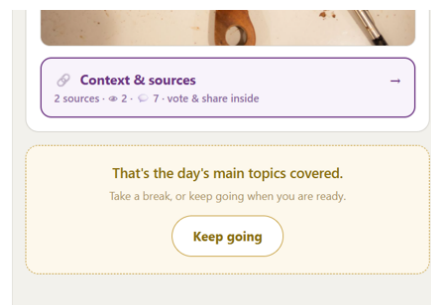


Figure 5.3 – Second segment boundary

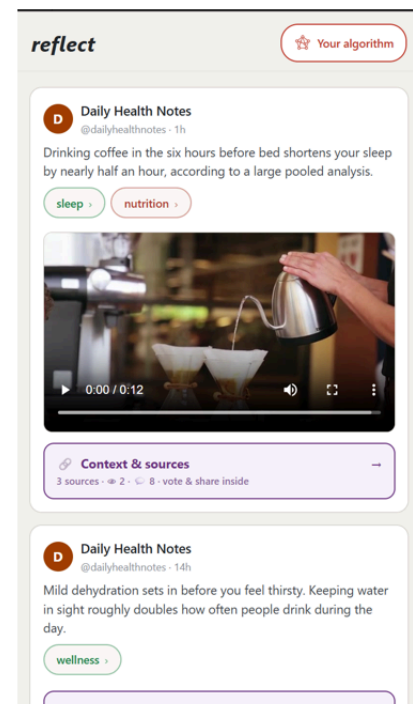


Figure 5.1 – Standard feed appearance

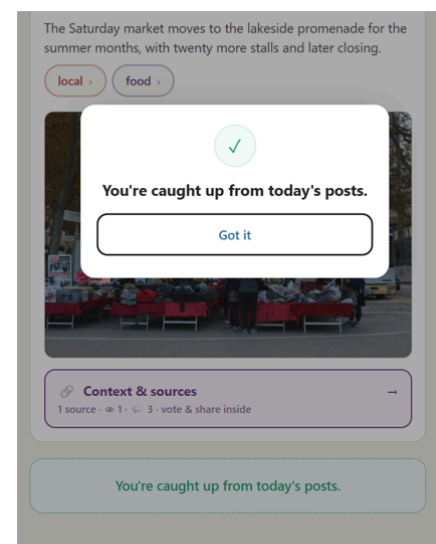


Figure 5.2 – First segment boundary

5.1.2 DG2 – Transparent, governable curation

The second design goal responds to the algorithm amplification dimension (Section 4.1), where curation was algorithmic by default on all platforms, logic was visible on none, and controls were three to four or more taps deep in settings with nothing encouraging their use. Two requirements follow: transparent, adjustable feed, and cross-viewpoint transparency. Two primary screens demonstrate these requirements:

Topic screen. Figure 5.4 shows topics under each post, modeled off of Reddit's “tags”. However, in this case, the tags are used not just for indicating the topic of a post, but surfacing explanations for why a user may be getting a post on their feed, and to provide immediate adjustment opportunities for either the topic, or the whole algorithm. The proximity of the adjustments to the explanation responds directly to the requirement of transparent, adjustable feed.

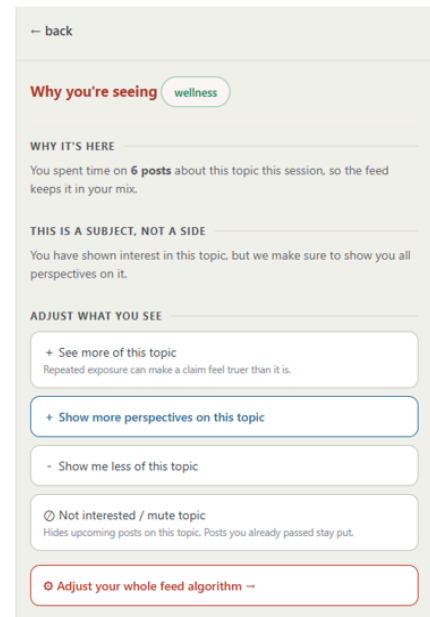


Figure 5.4 – Topic label adjustments

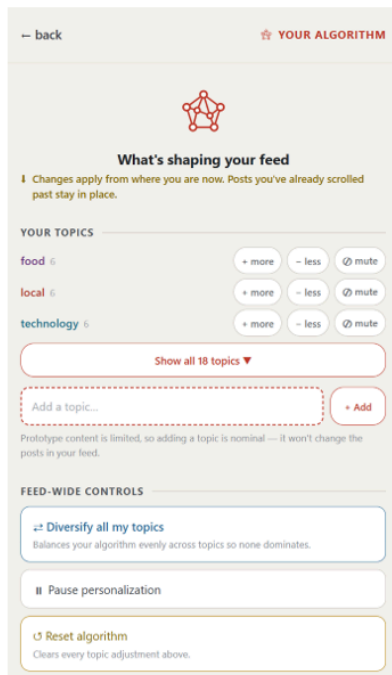


Figure 5.5 – Full algorithm panel

The algorithm panel. Figure 5.5 shows the whole-feed algorithm control view, openable from persistent control in the feed header in a single tap, against the three to four taps and nested subsections recorded in Section 4.1. The panel displays the weight each topic holds as a figure rather than only accepting adjustment: the diagnosed deficit was invisibility, and a control that conceals its own state does not answer it. Three feed-wide actions follow: diversify, which levels the distribution so no topic dominates; pause; and reset, which clears every accumulated adjustment. No platform offered a chronological feed, and reset is the functional equivalent. While it's not an unranked feed, it gives users an opportunity to leave a feed that a profile has accumulated. Where a control cannot do what it appears to, the panel explains that adding a topic is marked nominal, because the fixed pool holds no posts to supply.

5.1.3 DG3 – Restore Evaluation at the Point of judgment

The third design goal responds to the deliberation affordances and engagement metrics dimensions (Sections 4.6 & 4.3) where a claim's origin was frequently unreachable and social proof stood within the post with no way of evaluating it. This section addresses the requirements that engagement signals should be neutral and not without possibility of evaluation, that each post should have sources, viewpoints, and summaries, and that it should have a deliberative discussion surface.

Where counts appear. Limited engagement counts appear in the feed, and they are ordered by number of sources, views, and comments, with the indication that voting and sharing can only occur inside the next screen. Within the context & sources (figure 5.8), counts appear next to the source with the intentions of indicating validity of a source and encouraging users to read them. The ordering of the page also encourages them to slow down and read the claim, documents, and viewpoints first before voting. The idea is that these metrics shouldn't be evaluated as a substitution because it is not placed before the verdict and it is placed next to something that can be evaluated with a single tap.

Context surface. Every post can be opened to a screen that quotes the claim separately from the original post. Extracting the claim makes it the claim itself the content unit rather than the entire post. On other System 1 affording platforms, this would not be possible. It also mentions that everyone is shown the same claim, and that it is sorted by topic, not "side". Personalization is generally a feature of algorithmic amplification, and though separate from this dimension, it aims to support the restoration of evaluation. The screen then provides a primary document if available (in the screenshot example it is a law directly), and views follow with attached sources that are designed to reflect sources and sentiments provided in the comment section below. Rather than deciding whether a claim is true or false, the interface simply provides information without drawing conclusions or stating something singular as factual. The badge in Figure 5.9 marks newer posts whose views have not yet accumulated, so that low activity reads as recency rather than as a judgment of the claim.



Figure 5.6 – Metrics visible from feed card

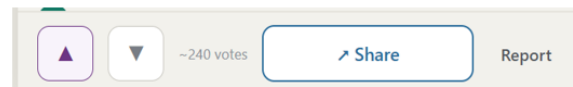


Figure 5.7 – Metrics visible from Context & Sources (after voting)

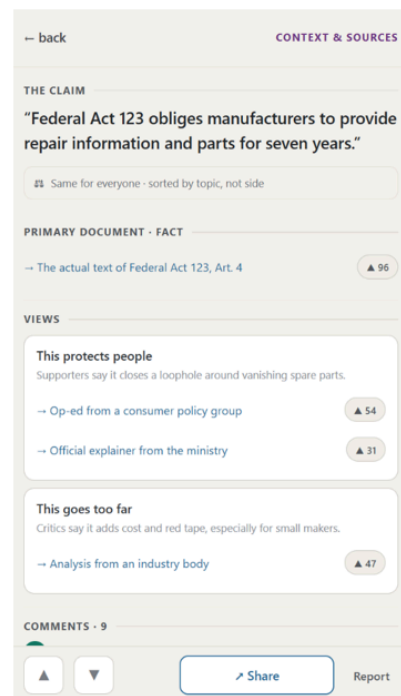


Figure 5.8 – Claim, views, and sources screen

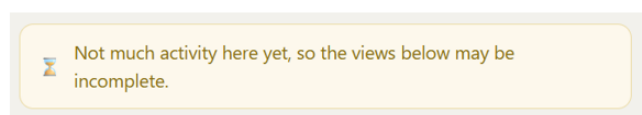


Figure 5.9 – Badge when there is not robust engagement

Deliberative discussion. Comments are organized into threads so replies are directly linked to the messages they address. The text editor also allows users to include sources, encouraging evidence-based contributions rather than mere assertions (Figure 5.10). Notably, comments are not sorted by engagement and engagement metrics are not noted. Ranking by likes tends to promote the toxicity and dogpiling that ruin healthy debate, which would bring back the very popular metrics this design aims to avoid. However, one aspect of this example shouldn't be mistaken for a design success: the thoughtful, respectful tone of the comments. This civility stems from the fact that the content was specifically written for the study, not because the interface naturally generated it. A static set of data cannot prove whether real users would actually maintain this polite tone in practice.

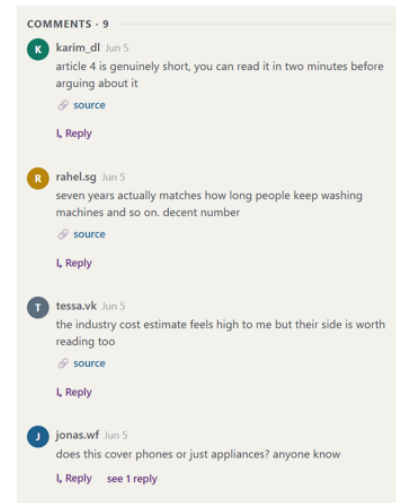


Figure 5.10 – Discussion space

5.1.4 DG4 – Intentional, calibrated action

The fourth design goal responds to the sharing and reporting friction dimension (Section 4.4) where a public repost costs one or two taps, private sharing sometimes only one, and a report three or more. No platform placed an accuracy prompt on the share path. Two requirements follow, intentional sharing and encouraged verifications. Three features answer this design goal and are explained below.

Where actions live. No action is even possible from the feed card directly, so it's nearly impossible to complete an action based on only what you see in the feed card. A user must navigate to context and sources, and only there may they up or downvote, comment, or share a post. This was an intentional choice as it introduces more decision boundaries and actions between taps. A user would need to read the post, decide whether they care enough to know more, tap again into context and sources, and still need to scroll past the claims and sources before they comment. Votes for the overall post are also at the bottom of the context and sources screen, keeping actions completely separate from the scrolling feed and only within the context and direct visibility of sources.

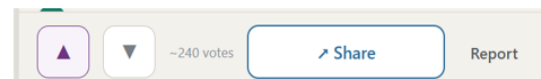


Figure 5.11 – Sharing and voting, only within Context & Sources

Before you share. Even within the context and sources screen, selecting to share opens a screen that asks whether the claim is accurate. It specifically asks whether it is accurate, and directs attention to the source. As noted in literature, a small prompt directing attention to accuracy reduces the sharing of false content, and a prompt recalling a particular action just taken is harder to habituate than a standing warning (Pennycook et al., 2021). There is also a note saying that the source and content travel with the share, a feature that is unique to the platform because the platform also provides the source. This is all shown in one screen so as to not leave users with a feeling of obstruction. It's not that users should not be allowed to share, it's that they should be encouraged to engage System 2 before doing so.

Reporting. “Report” sits beside “Share” in the same bar, at the same depth. Across the platforms analysed a repost cost one or two taps against three to six for a report, so matching them is a question of placement rather than of added friction. This is a property of the artifact that can be counted, not a claim about behaviour. With no audience and no real contributors, neither sharing nor reporting could be observed as action to be accredited to natural reasoning, and Section 5.2 returns to what that leaves unevaluated.

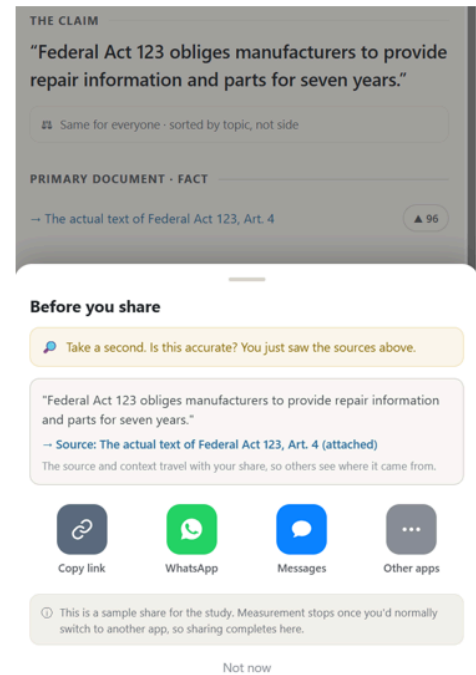


Figure 5.12 – Accuracy prompt at share

5.2 Departures and Design Boundaries

The context layer retains vote counts on community contributed views. Under the requirement for engagement signals being presented neutrally and not without possibility of evaluation, this is not a deliberate departure but a boundary. The requirements aim to de-emphasize engagement counts rather than eliminating them, as eliminating any and all metrics would make the platform appear stale, contradicting the non-functional requirement for environment authenticity. This goes both ways, counts only appear after the user has opened a post, and counts are attached to clickable sources rather than a claim itself. Chapter 4 also found evidence on raw counts mixed, so the retained count was judged a low-risk tradeoff rather than safe. Participants' interpretations of these counts are addressed in Section 6.3.

The design also includes a feature for reporting, though its use is not included in data collection or results. It is included to maintain honesty with the diagnosis that the friction asymmetry between sharing and reporting is friction that affords System 1 (easy to share, hard to report). The content provided in the prototype is intentionally neutral and there are ethical considerations against providing content that would reasonably be reported. Therefore, the use of this button is not explicitly assigned a task and it is not included in the analysis of the results.

Sharing holds a related caveat. The accuracy prompt appears only once a share has been initiated, so it was only encountered by participants who chose to share. This was not assigned a specific task either, they were only told to “engage” with the platform. Without stimulating content, or contacts behind the shared targets, they may not have felt the urge to.

The constraints named in the beginning of this chapter (single session, no network or persistent account, fixed content pool) leave some features purely nominal. The why-labels for topics come from static logic and not a recommender, and the reset algorithm is performed on a feed whose algorithm has not been built.

6. Results

This chapter will report the findings of the comparative concept study, organized in order by the four design goals, comparative ratings based on the UEQ-S survey, and a section on boundary conditions. The design goals correspond to the research questions as follows: DG1 (protect cognitive capacity) and DG4 (afford intentional calibrated action) bear primarily on RQ1 (design) and RQ2 (acceptance and appeal), DG3 (restore evaluation at point of judgment) also leans on RQ1 (design), and DG2 (transparent, governable feed curation) responds to RQ3 (algorithm control). Each of the design goal sections will describe the baseline accounts and compare the REFLECT block against it.

Three conventions are held throughout the results. Counts are reported as absolute numbers of participants, never percentages. The transcriptions were segmented, and denominators are therefore not comparable across participants. A participant counts towards a tally if any of their coded segments supported the finding, directly or indirectly, and any claim sentence is written generally enough to cover everyone coded. Interviews are treated as primary evidence, and action logs and UEQ data are sometimes used to establish exposure to a mechanism.

It should also be noted, as these results are read, that counts report how many participants' coded segments supported a finding, whether spontaneously or in response to a general probe. They describe prevalence in the talk, not prevalence in the sample. Participants who do not appear in a count did not address the point, which is not necessarily evidence that they would disagree.

6.1 DG1 – Protect Cognitive Capacity

This design goal targeted the features in the comparative analysis identified as affording System 1 processing, such as infinite scroll, autoplay, and image-led content. The prototype replaced them with a segmented feed with stopping points, text-first posts, and no autoplay. The intent was not to reduce use, but to reopen cognitive capacity that keeps the decision to consume intentional.

6.1.1 De-escalation and Stimulation Tradeoff

Baseline. 8 of 15 participants (P01, P08, P09, P10, P11, P12, P13, P14) spontaneously described the pull of their own feed (in each case a System 1 dominant platform, most often Instagram or TikTok) as something done *to* them rather than something they choose. They cite losing track of themselves mid-session (P10-010), and feeling “kind of trapped [in their feed]” (P11-097). Four others frame the cost in health terms, from “brain rot” (P14-166) to a platform bad for “humans’ brain or...humans’ health in general” (P13-150). One described the pull as a function of the genre of application rather than any single app, since “now, all the apps have the same things” (P13-152). The fullest statement came from P09:

“On Instagram you just like go see and it’s like reels, reels, reels and then you’re like oh I’m trapped in this. And then it’s gone without even wanting to be.” (P09-025)

Participants also did not expect to retain what they saw in their feed (P10-046, P14-046). Scroll time is not considered to be used with cognition.

REFLECT. 8 of 15 participants described the prototype in de-escalation terms, using words like slower, healthier, simpler, less stimulating. P13 mapped it onto the same vocabulary she had used to describe her own feed: less distractions, “more healthy”, a feed where “you can actually learn something” (P13-023).

However, the same “less stimulating, slower” aspect of the app was seen negatively by 8 of 15 (P02, P05, P06, P09, P10, P12, P14, P15), and the complaint was under-stimulation rather than anything specific missing (P14-077). P05 read the same “if you want to make an addictive app, I would definitely put more things” (P05-094). The clearest statement of the trade reads as:

“I think for getting information and...general culture, I think it’s really good, but...if I had the choice between this and something more...fast-paced, to just stretch my brain while I’m bored on the transport, I think I would choose the other thing that’s...giving me more dopamine.” (P14-083)

This writes an interesting story: the property participants recognized as healthier is the property that led participants to still prefer other platforms over REFLECT. P14 carries the whole trade in a single interview — he called his own feed “brain rot” (P14-166), triaged it in five seconds (P14-031), found REFLECT more boring and slower to read (P14-081; P14-150), and still allocated “60 or 70% of reflect” if both were available (P14-138; P14-136). Two participants reported no difference in tone at all (P03-080; P01-092).

6.1.2 Content Format Dictates Browsing Speed and Effort

Baseline. 6 of 15 (P02, P03, P04, P06, P10, P14) described a visual screening process that runs on the image and resolves within seconds: “the image is the thing that captures my attention first...I don’t even read the title” (P02-104), a decision “done in, like, I think, five seconds” (P14-031). The depth to which participants read was dependent on topic (P02-108), the shallow mode was enjoyed rather than tolerated (P03-008), and control over what was presented was already valued (P14-062; P14-066).

REFLECT. In the prototype, the format was inverted: title and text placed first, image secondary, and no autoplay for videos, and this landed with participants. Participants noted that more content is legible per screen (P10-068; P11-011) and accepted the added clicks where their purpose was visible: “clicking onto things is not necessarily a bad thing...if I have to read context and sources” (P14-144). However, the lack of images comes at a deliberation cost:

“It took me definitely longer...because there were no images and that’s what catches my eye when I look at Instagram. So first it’s the image...if it’s a topic that’s interesting, my brain will usually just react to it, and then I will see it more in detail and then the text.” (P14-150)

There were variations of the same frustration: it’s too verbose (P15-039), too information-dense (P09-043), insufficiently “shiny” (P12-076), the image should come first (P02-126), video would be preferred (P06-071). The format being text-heavy removed the processing fluency that made a fast decision possible, and the cost surfaced for the same reason. The deceleration noted in 2.1 and the friction here are parts of the same mechanism, and it should be made clear that the prototype cannot be credited with effective deliberation affordance without acknowledging frustration due to the extra effort.

6.1.3 REFLECT serves a separate purpose from baseline apps

Baseline. A common sentiment across participants was a division of labor between platforms. Some routed anything worth keeping away from the feed: P08 reserves research for a browser or YouTube, "a better platform to learn" (P08-013); P10 scrolls Instagram "just to pass time" but watches YouTube videos through in bed (P10-044), keeping "important stuff, like news" off her feed (P10-046). Others made the same division by assigning the informational purpose to their primary platform itself: Twitter "to stay informed" (P01-010), Twitter as the field's informational platform (P15-136), a curated Reddit used for research (P09-004; P09-017). Participants already operate a two-slot model: purposes assigned at the level of platforms, not managed inside them, which follows from the non-retention expectation in Section 2.1 and is the model REFLECT would be sorted into.

REFLECT. Where participants filed, the prototype answers the same question from the other direction. 8 of 15 (P02, P04, P05, P06, P08, P10, P14, P15) responded to whether they would use REFLECT by locating it relative to a platform they already use, and seven of the eight placed it in a slot distinct from their existing feed:

“[Instagram] is when you just use it out of your usual habits. When you’re bored, you just go on it and scroll a bit. REFLECT, I would more use it as if I’m following specific topics and I want to have news that are backed up. Kind of like Substack...But I wouldn’t spend, I think, a long time on it.” (P04-082)

P10 anticipated using it “only for news” (P10-133) and asked outright whether the intention of the app was “social media or just to browse information” (P10-137); P08 placed it “between Instagram and YouTube” (P08-116; P08-114). P02 is the exception inside the pattern, concluding the prototype “works the same way” apart from the sources and preferring the platform her friends already use (P02-099; P02-100). Read this way, relocation is not a failure of the design goal: capacity protection was accepted in the only form participants were willing to take. They give it a specific purpose rather than a reform of the scrolling they already do. It is also where the strongest claims stop, since every statement in this subsection is an anticipation made after a first exposure.

The prototype had two stopping mechanisms: a mid-feed segment card that hides the remaining posts behind "Keep going," and an end-of-feed modal, neither of which were mentioned by any participant, in a context where other features drew unprompted comments. The logs show the silence did not mean that they didn’t see it: thirteen of fifteen reached the segment cue, eleven tried to scroll past it, and seven pressed "Keep going" — encountered by almost everyone, halted about half, named by none (task-driven, time-boxed sessions and a finite post pool limit how far this can be taken).

6.2 DG2 – Transparent, Governable Feed Curation

This design goal answers the finding that curation on the platforms examined is opaque and ungovernable. The prototype attaches a topic label and a stated reason to every post, and places an algorithm panel with topic-level controls, reset, diversify and pause one tap from the feed.

6.2.1 Algorithm Awareness Already Existed

Baseline. Curation awareness was consistent with this sample. All thirty baseline segments coded to this dimension come from twelve of the fifteen participants, and every one of the twelve described a working

model of how their feed is assembled: content follows what you search for, watch through, and linger on (P08-031; P08-033; P03-030; P02-089). The fullest account came from P04:

“A few months ago, I broke up with my girlfriend and then I started having a lot of reels about couples and things like this. I just thought it didn’t interest me and then I skipped. And then after like five posts that I had tagged as not interested in, I had to go.” (P04-034)

Participants also named the effects of this and similar to the immersive design, named them as things done to them: an echo chamber (P05-058; P05-065), drift toward undeclared topics (P15-128), selection for shock (P01-040), topic flooding (P12-046). The only stated opacity in the corpus (P07-039) occurs in a REFLECT block; the three participants who contributed no baseline curation talk are silent on the question, not unaware of it.

REFLECT. The per-post labels were noticed quickly (P15-004; P11-004; P06-038) and were endorsed as a good idea (P15-037; P14-119; P08-097), but enthusiasm declined when participants applied them to themselves, and the two positions were held by the same people, in sequence. P06 moved from “kind of cool” (P06-038) through “doesn’t really interest me” (P06-040) to “the why is not important, but to suggest to have more or less of that is pretty cool” (P06-042); P15 ran the same arc from endorsement (P15-037) to “too explicit” (P15-008) to doubting “many people actually care” (P15-010). Overall, the added transparency was endorsed as a public good but rejected as a private need. Only P14 reimagined the explanation into something he did not already know: a periodic report of “what’s my centers of interest” (P14-121) like a personal audit of himself rather than a justification of each post.

6.2.2 Controls are Desired, but Require Undesired Effort

Baseline. Not one participant described using a specific in-platform control with any regularity (P10-036; P06-031; P15-157). Governance ran through behavior instead: 8 of 15 (P02, P03, P04, P06, P08, P09, P10, P14) described steering performed at the algorithm by unfollowing, tapping “not interested,” seeding the feed by searching, and modulating watch time. The most deliberate version is P14’s, who skips content he finds interesting, so the feed does not over-fit to it (P14-055). Boundary-setting performed as self-denial, because the only channel available is the record of his own attention. That channel also cannot express what users mean: one gesture must carry approval, interest and agreement at once (P05-043).

REFLECT. Every participant found the control surface unprompted: all fifteen reached the topic sheet and opened the algorithm panel, the one log figure this section uses. The surface was found in a prototype first seen. Nine of fifteen received the controls positively. P03 explained as they worked through the panel that it’s nice to have “more freedom of choosing what we want to have in our feed” (P03-050).

Resetting the algorithm was one of the most wanted controls, named by 9 of 15 (P08-093; P12-038; P11-041). P06 marked the limit of that enthusiasm, stating she would “never reset the algorithm completely...I’d keep the personalization,” P06-068). P14 echoed the concern and a way back to a previous algorithm as a potential fix (P14-117). Projected use was episodic: at account creation and rarely afterwards (P15-095; P14-109); only P11 described ongoing use (P11-039). Against all of this, a minority of 5 of 15 declined ongoing governance, and what they declined was the labor rather than the control: P10 doubted she could out-curate the algorithm because “my algorithm knows me better than I do” (P10-115), and “if I try to do my own algorithm, it would be worse” (P10-099), unable in session to commit to a choice at all (P10-103). P06 estimated the accumulated work a reset would spend (P06-086). One participant valued the controls in theory, but cited concerns regarding the data collected by social media companies to make the algorithm (P13-067; P13-182).

6.3 DG3 – Restore Evaluation at Point of Judgment

6.3.1 Identity can Settle a Judgment; Counts Rarely Do

Participants described judgment resolving or escalating at three successive points: the poster and the signals attached to a post, the comment thread, and the sources. Each is presented baseline-first. One instrument caveats the section: because participants' own feeds are private, the baseline block did not ask them to evaluate posts; the two blocks are therefore not symmetric at the content level, and no claim below rests on a comparison at that level.

Baseline. When asked how they would decide whether to believe a post, six participants named the identity of the source as the thing that settled their stance on a piece of information (P02, P03, P04, P10, P12; P07 in reverse):

“The first thing that popped up onto my feed when I logged in was a piece of news from the Financial Times, which I know is a reliable source of information. So I would generally trust what they say on there. Then if it’s something related to politics or public opinion, then I will usually go to Safari or Google and try to dig a bit into the subject.” (P04-014)

P07 inverted the signals discounting material in advance by political affiliation (P07-069). Two participants extended the logic to themselves, treating their own competence as the qualifying condition for belief (P15-134). Where the topic fell outside of their expertise, sometimes the outcome was abandonment rather than escalation: “I have no expertise in this, and I cannot use energy to fact check” (P12-050; P12-054). Engagement counts, by contrast, were raised far more often as something participants emit than as something they read (covered in 5.1). The sole exception was P14, for whom a post with no likes or comments was not less true but “less, like, peer-reviewed maybe” (P14-035), making a metric of attention read as a metric of scrutiny.

REFLECT. At the post level the gating worked as designed: no participant described using a post's engagement figure to decide whether to believe it. The reasoning itself survived, resurfacing at the one figure that remained visible, the vote counts on sources. Those counts were designed to carry a different meaning, a community judgment of a source's validity rather than a post's popularity, but participants read them the way the withheld count would have been read: as a number that settles what to believe.

Choosing between two sources, P04 read their votes as a warrant: “both have been voted by the community, so I think that it will be more reliable” (P04-053); P13 weighed a source by its count (P13-055); and P07, otherwise skeptical of links, named the count as the thing that would settle it (P07-019). On the response side, P10, unable to decide her own vote, explained “because there’s no one who voted, right?” (P10-091), implying the missing metric was what would have helped her decide.

6.3.2 Comments can be a Verdict or a Step

Baseline. Comments served two purposes, and only one is a metrics problem. Read as a poll, five participants went to the thread to establish what most people thought and treated that as the answer (P03, P06, P10, P12, P14) “Next, I’ll go to the comments to see what they say to decide whether it’s or not like true” (P03-016). If a participant was looking for information and reason the thread was a step rather than a verdict, depending on if they find an answer. P10 checks the comments for an answer and, when they fail to provide one, goes “out of Instagram” to search elsewhere (P10-016); P14 reads them and then Googles the subject regardless (P14-029). P11 used the thread differently, reading its composition rather than its

majority: "a voice for both sides" could be taken as evidence the claim was not being pushed at him (P11-055). For the two who described the thread not satisfying them, the only option was to leave the platform.

REFLECT. On the prototype's context screen the sources sit above the comment thread, so a participant who consulted sources first may simply have been following the order of the screen rather than changing how they judge. This ordering was not new to everyone: P15's own platform already imposes it, since on X/Twitter he cannot reach the comments without passing the community notes (P15-144), while P14 wanted it reversed, preferring to see people's opinions sooner (P14-140). The sequence that cannot be explained by screen order is P04's, because it contains a return: he read the report first (P04-051), found a comment challenging the survey's premise, and went back to the source (P04-055), which is precisely the escalation this design goal was written to make available. Four participants searched for and scrolled directly to the comments, passing the verification placed directly above them (P02-017; P06-075; P10-077; P15-075). In the deepest version the crowd does not merely come first, it resolves: "if four other comments say the same thing, I would just say, yeah, that's the truth" (P10-081), the account of the one participant whose logs record no source clicks. Two participants questioned whether the threads were real at all (P02-013; P11-023).

6.3.3 Sources Became a Tap, and Sometimes a Badge

Baseline. On their own apps verification was an act of exit. Nine of fifteen treated their platform as insufficient on its own to settle a claim: five named the move: "only by seeing this, I won't trust it. I will go on the internet, maybe do more researches" (P08-021; P03-018; P04-016; P10-016; P14-029), and four described it without naming a destination (P06-055; P09-015; P15-065; P02-124). The refusal is the strongest evidence that exit was the only route: P12 declined to verify precisely because verification meant effort outside the app (P12-050; P12-054).

REFLECT. With sources attached to the post and consolidated at the top of the context screen, users can tab a source instead of exiting the platform to search. The cleanest evidence is within-participants. In the baseline block P03 said:

"Sometimes if I don't know the person, I'm going to Google who is this person and see if I'm going to get more links about this news." (P03-018)

and in the prototype block, on the same "judge a claim" task:

"The thing that is good here is that we have some links... we don't have to go on Google and find the whole thing. We can just go through the links and see." (P03-042)

Seven participants described the sources as an early or first stop (P02, P03, P04, P09, P11, P14, P15); fourteen of fifteen clicked at least one source (P10 the exception); P09 named the substitution directly (P09-035). The finding is that presence can substitute for checking. For three participants a source being there was itself sufficient, without being opened: "even if I just see studies, then I'm probably going to believe it...It's not necessarily you have to tap on it" (P05-022; P08-060; P10-085). The existence of the source seemed to occasionally read as a badge of the property it is meant to establish. Skepticism ran the other way for others, citing that sources could be fabricated (P14-089), links pre-discounted (P07-019).

6.4 DG4 – Intentional Calibrated Action

This section intends to capture how participants interact with their baseline app, and how they project they would use it in REFLECT. However, the limit of this section is structural: the post pool in REFLECT was synthetic and not personalized to users: there was no account, no contacts and no audience, and no participant had reason to act on any post. Action in this block was therefore nominal by construction, and the action logs are used to establish exposure to a mechanism, never preference for it. What varied across the action affordances is whether participants could read what each was for.

6.4.1 Engagement Counts Are Outputs, Not Inputs

Baseline. Participants described engagement counts as something they output more so than something they consume. They tend to moderate how much they participate and whether it can be seen by other users (P05-072; P12-060; P10-028; P10-030). Several described a feature in Instagram where your friends can see if you like or comment on a post, so they withhold interaction due to the fact that it will be publicly visible. Other comments about how they engage with their social media feeds were generally standard and congruent: liking sometimes, commenting rarely, posting rarely. They tend to consume their feed more than they create.

REFLECT. The voting system, initially introduced to allow more than one interaction type, was not recognized by all participants. They looked for the more familiar “like” button. One participant, looking for the like button, didn’t find the voting system or recognize it as a substitution, and instead reached for a curation control: “since there’s no like button or something, I think I’ll just put see more” (P06-057). The rationale of the design was recognized unprompted by one participant, but she still objected to it:

“I like on TikTok that, compared to your app, that I can just, without having to tap on the posts, I can just like it... I understand why you do that, because a lot of what your content is, is more like facts, so then you can check it before you like it.” (P05-054)

6.4.2 Share Privately For Agreement, or Publicly at a Price

Baseline. 8 of 15 participants (P02, P03, P04, P05, P06, P07, P08, P13) described sharing as sending a specific post to someone, usually a friend. “Sometimes I send to a friend because I thought he thought the same as me” (P07-091; P04-022; P03-022). When deciding whether to share something, the question is usually whether a recipient would like or agree with it rather than whether it is truthful. When the action was public however, participants were more guarded: 7 of 15 (P01, P04, P05, P06, P11, P13, P15) described avoiding or being extra selective about actions their network could see, whether posting, commenting, liking, or reposting. The trend is generally that people avoid public actions, where other users can see their activity, and prioritize any private activity.

REFLECT. Consistent with the structural limit above, sharing in REFLECT was nominal. In the context of REFLECT, there is no audience or recipient a share would go to. The action could not carry its baseline meaning, and the logs are read as exposure rather than preference. Exposure was nonetheless substantial. Action logs show nine participants opened the share sheet, the share-with-source prompt appeared eleven times, and one nominal share was completed, the rest abandoned. The affordance was therefore discoverable, but the bigger note is that the prompt attached to it never appeared in any participant’s account of the session. Whether attaching the source at the moment of sharing changes the private-sharing tradeoff described at baseline is not answerable from this data.

6.5 UEQ-S Comparative Ratings

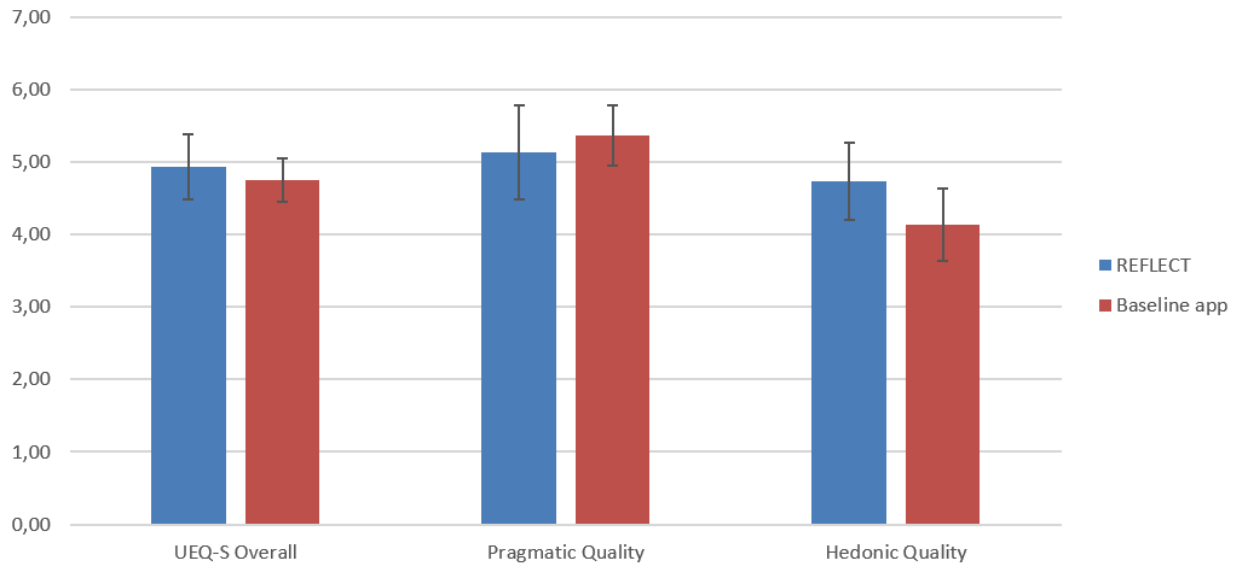


Figure 6.1 UEQ-S Overall, Pragmatic, and Hedonic Quality Scales on REFLECT versus participants' baseline app.

Both the baseline and prototype conditions were rated with the UEQ-S immediately after each task session. REFLECT's scale means were statistically indistinguishable from those of participants' applications. At N=15, these figures are descriptive at best and don't carry inferential weight. The baseline app is whichever social media platform each participant named and used in session (and was limited to the applications analyzed in the comparative analysis). The most frequent baseline application was Instagram or TikTok, but a few referenced X (Twitter) or Reddit. This means that the comparison is within-participant against a personally habituated product. The heterogeneity of these comparisons is discussed in Section 8's limitations.

The item-level profile is aligned with the trends found in interviews and explained below. Boring-exciting was the only item with a negative mean (-0.20 on the -3...+3 scale) while not interesting-interesting scored +1.40. This could read as the prototype being informative but not stimulating, and could be subconsciously related to the neutrality of the content. A first-exposure prototype being rated on par with applications its raters had used for years could be evidence that the added deliberation did not register as alienating; interpretation is reserved for the Discussion.

6.6 Boundary Conditions

The design goals are bound by some conditions. The first is platform contingency: for participants whose platform is already bounded and self-curated such as Reddit, defined in the comparative analysis as more System 2 affording, (P09-004; P09-006; P09-023; P07-041; P15-004) the apparatus is a match to what they have rather than an intervention. The second, purpose contingency. When the purpose of a session is to stop thinking, protecting cognitive capacity is a cost rather than a benefit "if I'm tired and I'm quick home and I'm on the train, I don't want to look at something that takes too much brain power" (P14-111; P05-080; P06-040). For switch-off use the design is not rejected but inapplicable. The third is novelty and network effects, which cap readings in both directions: unfamiliarity can present as boredom (P04-080; P15-093), comfort is trainable (P14-085), and no prototype tested in a single session overcomes the

platform one's friends already use (P02-100). Finally, every use-intention reported in this chapter is an anticipation made after a single moderated exposure, and statements of projected use should not be overvalued.

7. Discussion

Though the application and its acceptance are analyzed as a whole, given the compounding nature of the mechanisms (Section 4.7.1), the parts must be broken back down in order to reach specific conclusions about what does and does not work, and why. From there, the features that may have encouraged or prevented acceptance can be examined in turn.

Each subsection responds to the research questions restated below:

RQ1 - Design

What interface changes shift a feed environment from System 1 dominant to System 2 permitting, while still appealing to users regarding their use of the app?

RQ2 - Acceptance

Do users accept and uptake the deliberation-supporting affordances? Do they feel it would change their behavior?

RQ3 - Control

Does visible, governable, control make users feel more in control and perceive that it supports their reflection?

7.1 Affordances Shifted Processing, but Shortcuts are Displaced

RQ1: What interface changes shift a feed environment from System 1 dominant to System 2 permitting, while still appealing to users regarding their use of the app?

7.1.1 Removing Fluency Is Costly

Section 2.4.5 described how video-led content format directly affords System 1, due to its processing fluency. Section 4.5 found that five out of the six platforms had photo and/or video as the dominant unit format, and four of the six had the video feed position in navigation as either prominent or primary in navigation. Therefore, one of the primary design interventions to shift processing was to remove video-led content, replacing it with text forward content, with occasional photos and videos that ensured the prototype still felt realistic and familiar.

At baseline, six of fifteen described how they make decisions or resolutions based on image or video (Section 6.1.2). In the prototype, several of the same participants, among others, reported that judgment took longer and there were complaints of the format being too verbose, too information-dense, or “not shiny enough”. These complaints are the finding. The cost of the removal of fluent format was surfaced at the same point where participants would have used the image to come to a resolution. This is consistent with the fluency account, that removing fluent format raises the cost of an easier route and slows down the user. This may suggest an environmental move toward System 2 permitting. However this claim is

bounded because we cannot necessarily prove that it opened up cognitive capacity. The distinction between System 2 permitting versus causing is discussed in Section 7.4.

For RQ1, the format inversion is supported as a change that can move the environment towards System 2 permitting. The second clause of the question – “while still appealing to users” is where the claim is limited because slowing the processing fluency is also what participants complained about, and is described further in Section 7.2.1.

7.1.2 Any Shortcut Provided Will Be Taken

Section 2.4.3 described heuristic substitution, which explains that where the cue is easier than the evaluation, it replaces it. On mainstream platforms, the engagement counts are the most easily accessible., even though credibility judgments rest more on source and expertise cues than on counts (Meinert & Krämer, 2022; Morris et al., 2012). Section 2.4.6 added that mainstream surfaces make it difficult for users to trace an argument due to absence of robust deliberation affordances. Section 4 found that engagement metric counts were always placed next to content, discussion structure only had any real depth on one platform, and links to sources were not clickable on two of the six platforms. The design choices made to invert this were to remove like counts unless they were next to something that could be checked, attach sources and arguments on the same surface, and structure the comment section as threaded, with the possibility of deep threading.

In the baseline, six out of 15 participants named the identity of the source as what determined whether they believe something, and two others treated their own personal expertise that would confirm whether or not they even care to fact check it (6.3.1). Participants also referred to the comment section to help settle their opinion (6.3.2), and if they did want to verify something, most said that they had to exit the platform to do it (6.3.3). Engagement counts were generally not reported to have any impact on how participants judged a post.

In REFLECT, some read the vote counts on sources as a veracity flag, and others said explicitly that the simple presence of a source would confirm their belief. The other shortcut was that four participants immediately searched for the comments, scrolling past the the sources, reporting that in their own apps, that’s how they confirm whether they believe a claim.

These observations illustrate a consistent underlying mechanism. The initial design assumed that hiding the vote count would stop users from relying on mental shortcuts. Instead, data reveals that seeking out these cues is a deeply ingrained habit. Rather than the cognitive heuristic dissolving, it was just displaced onto the next available signal. If another number was visible, it was a proxy for social proof (P04). The mere presence of evidence was taken as proof of accuracy (6.3.3), or users would seek consensus in the comments (6.3.2). The verdict is therefore displacement of the heuristic rather than disappearance.

Supplying evidence, withholding counts, and chronologically deferring comments until after sources were visible make the environment System 2 permitting in that they give users more opportunities to look at evidence. However, it does not make System 1 resistant because any legible cue available will be used. For RQ1, the possibility of a shift to System 2 is limited by the fact that users will take whichever shortcut is left available, and the adverse effect is that continuing to remove shortcuts makes the interface less pleasing and limits the acceptance aspect of the research question, as found in Section 7.1.1.

7.2 REFLECT Was Accepted, But Use Was Conditional

RQ2: Do users accept and uptake the deliberation-supporting affordances? Do they feel it would change their behavior?

7.2.1 Cost and Benefit are the Same Mechanism

Chapter 2 identified cognitive load as the mechanism of the immersive design dimension. Continuous supply of fluent content with no stopping point removed the boundaries at which a user would decide to continue, and oftentimes the entry into the video-led immersive feed is intentionally made easy. Compounding this, System 2 is a "default-interventionist", meaning it does not engage unless prompted. Cognitive capacity required for deliberation to begin, and fluency saps this capacity because it uses no stopping points.

In Chapter 4, many of the apps proved to have no stopping point, easily afforded an immersive video stream, and truncated text. The intervention was a general capacity-protecting configuration consisting of multiple requirements: a stopping point, no escalation to a full-screen video-forward stream, no autoplay, and no truncated text. Together these open up cognitive capacity and afford microdecisions. They also remove the fluency and immersion afforded by System 1 dominant platforms, since the features that generate the load are the same features that pull users in.

Section 6.1.1 shows this intervention cutting in both directions. Participants who described the pull of their own feed as negative also described REFLECT's feed as slower, healthier, and a place they could learn something. However, they also mentioned that in some contexts they would still prefer the thing that "gives me more dopamine" (P14-083). The same slower register was equally described as understimulating, and oftentimes were said in the same interview. The UEQ-S item profile records the same split, with boring–exciting the only negative mean against a strongly positive not interesting–interesting (8.5). REFLECT cannot be credited with making the feed healthier without also being charged for making it boring, because both describe the same change.

6.3.1 frames the resolution for this. In participants' baseline environments, they already report using different platforms for different purposes. Some platforms are described as being "better platforms to learn" (P08-005) implying that they go to other platforms for information they want to digest intentionally. Another cites keeping "important stuff, like news" off their feed (P10-046), implying that their baseline app is not the correct context in which to consume informational content. The microdecisions and frequency of sources made the app feel more like a chore (P05-080), or too demanding for a tired commute (P14-111). REFLECT supplies less of a pull because the slower, less fluent environment gives no reason to switch to it other than occasions that already call for it. 8 of 15 projected their use of REFLECT in addition to, not instead of platforms that they already use. The aforementioned slower and "boring" nature that participants described is thought to be why.

In terms of the research question, participants suggested that they accept the deliberation affordances, but conditionally. Whether it would change their behavior is mixed, since many described REFLECT as fulfilling a separate purpose from their baseline apps rather than replacing them. Uptake is contextual and therefore may depend more on the purpose the user assigns it to rather than on the design alone. In dual system terms, participants allocate System 2 effort by platform more so than by post, and REFLECT was accepted in contexts where participants would already assume that effort. The verdict is therefore

confirmed because the cognitive load mechanism behaved as predicted in response to the interventions, but the benefit of a slower feed comes at the cost of some adoption and any design choice must consider both at once.

7.2.2 Friction is Accepted Where Purpose is Evident

Friction asymmetry is when the actions that drive engagement cost less than the actions that involve reflection. Section 2.2 supplies the constraint on any fix for this. An affordance raises or lowers the cognitive cost of an action without dictating an outcome, so redistributing friction can open a pathway but cannot compel anyone down it.

Section 4.4 found the asymmetry running the same direction across all six platforms. It would take one to two taps to share, three or more to report, and no accuracy prompt or source shown on sharing anywhere in the set. The intervention balanced these out and attached and consolidated sources at the top of the context screen so that verification became a tap rather than an exit. Reporting was built to cost the same two taps as sharing, and voting required opening the post rather than reacting from the feed.

The cost of deliberation did fall. At baseline nine of fifteen treated their own platform as insufficient to settle a claim, and verification meant leaving it — "I will go on the internet, maybe do more researches" (P08-021). In REFLECT, fourteen of fifteen clicked at least one source, and P03 shows the change within a single participant, moving from "I'm going to Google who is this person" on the baseline (P03-018) to "we don't have to go on Google and find the whole thing" on REFLECT (P03-042). Where the added cost had a visible reason, it was accepted: "clicking onto things is not necessarily a bad thing...if I have to read context and sources" (P14-144).

Visibility of purpose was not sufficient on its own. P05 reconstructed the rationale for the extra clicks unprompted: "I understand why you do that [add more steps in between the feed and voting], because a lot of what your content is, is more like facts, so then you can check it before you like it". However, they objected to it anyway, preferring to like a post without opening it (P05-054). Understanding why more friction exists does not necessarily mean it will be accepted, especially when there is an easier alternative elsewhere.

Making deliberation cheaper was appreciated, but added friction for the feed navigation via extra taps or scrolling was not, even when the reason was understood. Cheaper deliberation is not the same as more deliberation, as was seen in Section 7.1.2. The friction asymmetry mechanism behavior worked as expected in response to the interventions, but only where the added cost produced evidence. Cheapening the route to System 2 by removing friction to sources worked, but routine System 1 actions being taxed did not help recruit System 2. Adding friction for routine actions was generally refused even when the purpose was understood. For RQ2 this means that the added sources and removed friction to read them were accepted, but any other type of friction may not be.

7.3 Awareness Was Never the Deficit, Levers Were

Does visible, governable, control make users feel more in control and perceive that it supports their reflection?

7.3.1 Revised Opacity Diagnosis

Section 2 identified opacity as the mechanism of the algorithm curation dimension, and explained how it feeds illusory truth. When users cannot see why content is placed in front of them, it is easier for them to consume content as being true only because it is familiar. Section 4.1 found that no platform offered an explanation for why an item appears in their feed, all platforms presented algorithmic amplification.

The intervention was to keep curation legible at the point of delivery. A clickable topic label was on every post, and with one tap, there would be an explanation for why a post within this topic was seen, warn them against familiarity-as-truth, explain that the feed was topic sorted (not opinion sorted).

Section 6.2.1 shows that according to participants, algorithms are not completely opaque. It's true that no explicit logic is visible, but participants seemed to clearly understand that their feed was algorithm based and that they were shown specific things based on their interactions. They explain making specific interaction decisions to influence the algorithm to show them what they want. Though it's true the controls nor curation is explicit, they have a grasp on the controls and why their feed shows what it does.

In REFLECT, participants noted these topic explanations and algorithm controls, and endorsed them, but not all wanted it for themselves. They already know why they see the topics that they do – they already control it themselves, just with much more subtle actions that they already perform to some level anyways.

For the third research question, the verdict is revised rather than confirmed. The design correctly targeted the mechanism but misunderstood the problem. The problem is not that they didn't understand the algorithm behind their feed, it was that knowing how it worked didn't change anything for them – it would not alone change their behavior.

7.3.2 Control is an Occasion, Not A Routine

Though an affordance does not change an action directly, it can change how much effort it costs. Section 4.1 found algorithmic amplification on every platform in the set, with controls over it limited to overly specific keywords or per-item signals, and also buried several pages deep in settings. This led to the development of the algorithm panel control feature, accessible both by the main feed as well as the topic button. This panel allowed users to reset, diversify, or pause their algorithm.

The baseline finding in 6.2.2 found that participants did attempt to control their feed, but only through cheaper action that they mostly performed anyway: likes, skipping quickly, selections on the explore page to diversify, etc.

The algorithmic control feature in REFLECT got generally positive reactions, especially for the reset feature, which nine of the fifteen participants said they would use. The topic-level tuners were praised in session but not often mentioned in future use. Projected use overall was episodic rather than continuous, and a minority declined ongoing governance altogether because of the effort it would take. They already interact with the content anyway, they like their algorithms, and rebuilding or curating one themselves is effort without a payoff.

As it relates to the third research question, the governable features do give users a sense of control. The control affordance was accepted and used, and the baseline shows that motivation was there before this design. Whether the control was perceived as supporting reflection is less answerable because most participants talk about algorithm controls was more about comfort and fit of the feed, not about how it changed their judgment. The direct verdict on the research question is confirmed in the sense that it gives

users more control, but the same effort-cost mechanism appeared in response to the intervention. Governance is System 2 work and participants responded to it either by minimizing the work (effort rarely and high level) or by rejecting it altogether and sticking to the actions they already take. The implication is that high-level governance should be cheap and easy to find, and this design made it easy to find, but it was overbuilt for something users don't care to use as consistently.

7.4 Synthesis

The most common thread across these sections is that the same type of mechanism goes in both directions. Removing fluency creates space for System 2 processing, but the feed is boring. Withholding engagement metrics and options forces the user to open evidence, but the presence of it or any legible cue is substituted. Cheap verification is accepted, but only expected to be used in a context where it is already expected to cost effort. Governance is wanted, but only at a high level. The nudges towards deliberation are registered as acceptable, but usually with some type of "but" clause and frequently with a deference to their existing platforms.

In Dual System Theory terms, this reflects Kahneman's description of System 2 being defined as "lazy" and System 1 being the default unless System 2 is poked awake. The System 2-led design of REFLECT being compared to the generally System 1-led baseline already doesn't set REFLECT up for success. It must be as easy to process while still affording all of the right verifications, and engage System 2 with as little effort as possible. This is exactly what surfaced through the participants' responses across the board. They like the idea and recognize the negative effects of the immersive design and algorithmic amplification, but Dual System Theory does not account for preference, it is based on cognitive effort. REFLECT is not competing with the baseline for approval, it is competing for the path of least resistance, and as long as it is System 2-led it never would be. Deliberation will lose unless the design makes it a cheaper move.

This ease of defaulting to System 1 showed up in other ways across the data and these sections. When engagement counts were hidden until context and sources were opened, any metrics on the page after the fact were accepted. If it wasn't the metrics that settled the score, it was the presence of the sources at all. Even when given text-led content units, they would go straight to the page. When given topic controls, they prefer their own natural engagement cues to curate their algorithm, or would only use the high level controls. This was not perfectly present across every single feature, many participants responded well to some features. Credit was given to the presence of sources, and many said they would use the algorithm reset button. This demonstrates attribute substitution (Kahneman, 2011), where users replaced the demanding assessment of whether a claim was well-supported with a more accessible proxy that the design made easily legible.

The project premise was to create an inversion of what is "wrong" with existing platforms. The data shows here that directly inverting every feature doesn't necessarily work as a method, because every inversion comes at cost in every case described. It is ineffective because the affordance's cost and benefit are on the same axis. The interventions taken together succeed in making the environment System 2 permitting. Capacity was protected or restored, evidence was placed within reach, and curation legible and governable. However, permitting System 2 doesn't make a user resistant to System 1. Making room for critical thinking is not the same as causing it. The research questions pass on shifting the environments and putting users in control, but it falls short on some levels of acceptance of the deliberation-supporting affordances, because it will never feel as easy as a platform that leads with System 1. The design question

that follows is therefore not how to add further support for deliberation, but where the cheapest path through the environment already leads.

7.5 Design Implications

The findings are specific to REFLECT, but the implications are not and should be considered in future work which is expanded upon in Section 9. This section states five implications for designing environments intended to support deliberation.

7.5.1 Invert Features Selectively

Removing fluent formats made evaluations possible, but it made the feeds less appealing. As discussed in 7.1.2 and then more generally in 7.4, a common thread was that the same feature that makes something easy to consume also makes it pleasant. A design that inverts every System 1 format pays an appeal cost. The more efficient choice could be to apply text-forward format only where a claim needs to be assessed, and leave other content in the more user-accepted form.

7.5.2 Decide Where the Shortcut Goes

Withdrawing cues does not immediately remove the judgment it supported, it only moves it to whatever cue remains (Section 7.1.2). The design question becomes not whether a shortcut will be used, but which one is the cheapest and what it will indicate. REFLECT may have applied this correctly by attaching the counts only to where items could be checked. It was still used as a shortcut, but arguably a more informative one as it pointed to the sources rather than the post itself.

7.5.3 Spend Effort on Verification, not Ordinary Actions

Effort was most holistically accepted where it produced evidence, seen in fourteen out of fifteen clicked a source, and most spoke highly of them. If this is the intended act of deliberation, this is the part that should be most seamless, and nothing else. Effort was more consistently refused when it was in front of a routine action, even if they understood the purpose of it. In this case, placing sources where the claim is already in sight costs nothing extra to the user. Any extra steps before an ordinary action are not necessary.

7.5.4 Decide on the Tradeoff Between Designing for Occasion Rather Than Replacement

Participants already use different platforms for different purposes, and accepted REFLECT to fill one of those slots (Sections 7.2.1 & 6.1.3). It was not considered to be an alternative for their current feed. A deliberation supporting environment is more likely to be placed alongside scenarios where users are already expected to give intentional thought. Too much of a System 2 environment will put it in this category, so if a design is intended to replace a mainstream (System 1-led) platform, it likely has to carry less System 2 features.

7.5.5 Provide Controls Instead of Explanations

Participants already understood their feed, their algorithm, and how to control it well (Section 7.3.1). It likely did not need extra explanations for why they are seeing certain content, so additional explanation of why something was appearing in their feed was not needed. They liked the transparency in principle but didn't apply the need to themselves. They did not grasp for robust control either, they just needed levers that they could toggle occasionally rather than a continuous tuning task that required more of them to carry out a function they already did naturally.

8. Limitations

The claims in this study are constrained by the artifact and the material it presented, the setting in which it was evaluated, by who took part and how their data were read, and by the comparative analysis that supplied the diagnosis that the design responds to.

8.1 Stimulus and Artifact

The content pool that populated the REFLECT prototype was identical for every participant, not curated for any of them, but their baseline that the REFLECT was compared to was potentially trained on years of their own data and behavior. The comparison is therefore not symmetrical and any preference the participant may have expressed could be attributed to the personalization of their own app rather than the design of the prototype. Participants were told to consider the platform design and features rather than the content, but content intruded regardless and based on their responses, the instructions did not always hold. This most likely affects the UEQ-S ratings as well, as the scores cannot separate the interface's appeal from the content beneath it.

The content pool was also intentionally civil and non-political and this may differ from participants' baseline app content as well. Neutrality was necessary for ethics clearance and to keep reactivity comparable across participants, but it is nonetheless a deviation from the likely norm of what participants are used to seeing in social media. The findings therefore only establish whether deliberative features are acceptable on ordinary content, not whether they actually motivate reasoning since the reasoning would likely not be applied to such neutral content. Several participants do not even use social media for news or information so neutral content may be of further annoyance to them.

REFLECT also lacked structural features participants may have expected, such as a search or navigation bar, settings, or the ability to look at their own profile or that of others. These were omitted because the object of the study is a default feed, and this was stated at the beginning. Participants noted this anyways, and it could affect how they rate their experience if it was understood as incompleteness or a judgment against the design. The same concept applies to the transparency layer in the algorithm settings which is based on static logic rather than a true recommendation algorithm. Since participants were evaluating a simulated algorithm, their reception of it may depend on how believable or how well they accept the mock explanations.

8.2 Evaluation Setting and Measures

Each condition (REFLECT/baseline) was used for about eleven minutes, in a lab, on hand-carded tasks, while thinking aloud in front of the researcher who designed the prototype. The browsing task would be the most affected, as it is not an inherently natural environment in which to scroll your feed "as you

normally would”. In the REFLECT block especially, participants knew it was simulated, so shares were initiated and abandoned not due to confusion, but because they already understood the feature and they knew the content was nominal.

The UEQ-S was the only questionnaire administered. It speaks to acceptance (RQ2) but does not carry any measure of constructs relevant to the design affordances (RQ1) or algorithm control questions (RQ3).

At N=15, the questionnaire results are purely descriptive, and a lack of statistical significance in any UEQ-S results would not be evidence of equivalence. The baseline condition is also technically not one condition but four. Participants used Instagram, TikTok, Reddit, or X, and Chapter 6 explains that these differ greatly in how they afford System 2 processing. A difference between Reddit versus TikTok against the prototype are not equivalent comparisons. Pooling these platforms as the “baseline” assumes an equivalence that the diagnosis does not technically agree with.

Every statement a participant made about future use is an anticipation formed after a single exposure. Nothing measures adoption, sustained use, truth discernment, or susceptibility.

8.3 Sample, Data Quality, Analysis

Participants self-selected into a paid lab session through the HEC-LABEX pool, which skews the sample towards a university affiliated comparatively young and highly educated population in one Swiss city. Screening was on self reported behavior rather than demographics as the intention was diversity of user-type, not age or race, and permits no demographic claim. Three personas roughly guided recruitment via screener responses, but weights were adjusted to create the persona groups, so dispositional coverage is claimed by design rather than demonstrated.

Sessions ran in English, which for most participants is not a first language. Think aloud and interviews reward fluent articulation and coding depending on words or ideas that participants used which may not directly reflect their natural thoughts due to language barrier.

The author designed the artifact, derived its requirements, moderated every session and coded every transcript alone, but precautions were taken to limit any bias and keep the analysis traceable. The codebook was fixed against the six analytical dimensions, so categories were not invented to fit the data. Coding ran in two passes, one for high-level analysis and another for quote extraction and stance. Every segment carried an evidence type: observed, reported, and projected. Every claim made by participants can be traced back with a segment ID.

Automatic speaker diarization was not reliable, so assignments were hand-repaired but not every bit of transcript was compared directly against the audio. The logs were prototype side only, as the baseline was only observed through field notes which were incomplete due to limitations in multitasking, and served as context when coding the interviews rather than considered data. There were occasionally bugs in the task-start action log from the moderator’s side, so any task attribution in the transcripts were from transcript cues.

8.4 The Comparative Analysis and Scope of Claims

The cold-state walkthrough of the comparative analysis phase was produced by cold-state walkthroughs on neutral accounts, which isolates to platform defaults. However, defaults could be the mildest state of the mechanisms being diagnosed. Feeds could be increasingly more or less video-led over time, and

network exposure is minimal as well. The analysis captures the entry stage of environments where System 1 pressure could increase with account maturity and potentially understate the diagnosis.

The Phase 1 walkthrough also represents a single analyst snapshot of the six platforms in May 2026. These platforms are subject to updates and A/B testing. The selection of the platforms may have been non-exhaustive since the sixth dimension was added after the initial comparative analysis during design iteration.

Because the framework, artifact, and evaluation were executed by a single researcher, the prototype was assessed against goals set by its own designer. While bias is mitigated through a traceable design chain, it means the evaluation is a formative exploration rather than an independent test of the initial diagnosis.

Finally, this work is a formative concept evaluation of one artifact. It supports claims about whether a deliberation-supporting feed is legible, tolerable and appealing at first contact, and about which affordances participants took up or ignored. However, it cannot support claims about accuracy judgments, susceptibility to misinformation, sustained behaviour, or how these affordances would fare on politically or emotionally charged material.

9. Future Work

This study was a single exposure evaluation of one artifact, and its claims are bounded accordingly in Chapter 8. The three directions below follow from a specific result that this study could surface but not settle. The first addresses the structural limits of the artifact itself – it may be worth reevaluating the design with personalized content and sample scale. The second isolates the variable that this study could not separate, which is whether a mechanism is fully understood versus encountered. The third takes the result that REFLECT would be put into a separate use slot from mainstream platforms, what it would take to design the app for this slot rather than forcing it to compete for casual engagement.

9.1 Evaluate a Personalized Prototype at Scale

REFLECT was built on a non-personalized content pool with no account, no social network, and no real audience. The construction made the comparison to baseline understandable, but placed limits on what findings could claim. Eight of fifteen participants stated variations of it being under-stimulating, and it's difficult to place this as a fault of the design or the flat and unfamiliar content. The curation transparency was evaluated on a feed that was not personalized to them, so “why” labels are nominal. Sharing and reporting were nominal by instructions as well.

The natural next step would implement the design implications in Section 7.5 and re-evaluate them on a personalized feed with an onboarding sequence that establishes real interests and a content pool drawn from the real material rather than a curated set. A larger sample would also allow for more diversity and therefore more user personas. Many participants seemed to have an idea of algorithmic curation and how to wield it, but it would be useful to understand how it affects people without this fluency, especially as it would probably be the most important target population for such a platform. This study would have to trade some experimental control for ecological validity.

9.2 Isolating Legibility from Prominence

The mechanisms in REFLECT were reliably encountered, but participants did not always explicitly name whether they understood them. If features were prominent but not understood, it could demonstrate that they are not legible. The curation panel for example (which had to be deliberately opened) was opened by all but one participant and discussed at length by many. The feed segment card was reached by thirteen and mentioned by no one. It's true that the latter was not reasonably probed in a way that would suggest to the participant that they should mention it, but this is an example of where it is known that the participant saw the feature (prominence), but it is not known whether they understood its purpose (legibility) unless it was explicitly stated in cases like “Your Algorithm” or “Context and Sources”.

This information cannot be inferred from the data because self-description is linked to the mechanism type, task context and familiarity. For example, the algorithm panel was chosen rather than imposed, gathered several controls behind one recognisable settings metaphor, and was encountered during a task that directed attention toward it. An experiment holding the mechanism constant and varying only whether it explains itself would separate this. The feature, in the same position, with and without a line stating what it is for, while varying how visually obvious the cue is made to the user. Continuation behaviour would establish the immediate effect and unprompted recall after the session would establish whether the mechanism entered the participant's account of their own use at all. The hardest case for such a design is a veracity signal, which must inform a judgment without being read as its conclusion. This is made difficult in cases where the design could be taken as a veracity signal. Section 7.1.2 showed that a legible cue tends to be taken as an answer, so any signal that could serve as a substitution must be made legible enough to inform a judgment without replacing it.

9.3 Designing for the Slot Users Assigned

Participants did not describe regulating their behaviour inside a platform, they described assigning platforms to purposes. Some are for passing time, and some are reserved for learning and application of attention. Eight of fifteen located REFLECT in the “retain” genre, rather than a replacement for the “pass-time” feed they already use (Section 6.1.3). Given the acceptance and under-stimulation they reported in the same breath in the interviews, this suggests that the trade-off observed here is not an issue of the implementation but a consequence of building a deliberation-permitting interface for a purpose that users intentionally use non-deliberatively.

That raises a question this study was not designed to answer: should a critical-thinking platform compete for the pass-time slot at all, or should it be built for the retain slot it was in fact sorted into by this study? The two targets imply different designs, different content strategies and different measures of success, since session length and return frequency are appropriate to the first and misleading for the second. Testing this requires a longitudinal deployment in which slot assignment can be observed as it settles rather than inferred from a single session, and ideally a comparison of two builds of the same platform optimised for each. It would also make it possible to establish whether the relocation observed here is stable, or whether a platform placed in the retain slot is simply one that is eventually not opened.

10. Conclusion

This work set out to understand whether a social media feed could be built to support critical thinking rather than suppress it, and whether people would accept the design. This was approached by using Dual System Theory and affordance theory as a lens through which to build a framework of six interface dimensions, each tied to a mechanism supported by cognitive literature, and applied this framework to six mainstream social media platforms. The diagnosis was consistent in that many present algorithmic feeds with no visible logic, engagement counts before opportunity for evaluation, sharing is cheap, video plays by default. The mechanisms compound each other with depletion of cognitive capacity at the base of any combination that exists on a platform. The platforms sit on a continuum from System 2 permitting with Reddit on one end, and TikTok and Instagram on the other.

This diagnosis produced four design goals and ten requirements, and REFLECT was built against them in an attempt to invert existing features of social media to build a more System 2-led feed. Features included a bounded feed, claims and posts presented with more balanced content format, a topic and reason to every post in the feed, withheld engagement counts until they were attached to something evaluable, and a range of views above a threaded comment section. Fifteen participants used REFLECT and their own most-used platform in the same session. Data was collected through interview recording and transcriptions, action logs, UEQ-S questionnaires, and then synthesized through a strict codebook in order to extract trends in participant responses.

The first research question asked whether interface changes shift a feed environment from System 1 dominant to System 2 permitting, while still appealing to users regarding their use of the app. It was found that the interventions did what they were intended to do. Removing fluent formats slowed judgment down, attaching sources made evaluation of a post easier, algorithm curation became legible and adjustable. This made the environment System 2 permitting, but they did not make participants resistant to System 1. Wherever a cue remained, it would be used as a shortcut, seen as how the vote counts on sources or even the presence of a source at all was used as a substitution.

The second research question asked whether users accept and uptake the deliberation-supporting affordances, and whether they feel like it would change their behavior. Participants accepted REFLECT, but with conditions. The changes that made the feed feel calmer and healthier also made it slower and less stimulating, often with both statements provided in the same breath by the same participant. The UEQ-S rating put REFLECT as indistinguishable from other applications they had used for years, suggesting it was not received entirely negatively. Interface friction was accepted when its purpose was transparent and yielded tangible evidence, like for accessing the sources, but it was rejected when it obstructed a routine action like voting, even when the reason for it was understood.

The third research questioned whether visible, governable, control makes users feel more in control and perceive that it supports their reflection. In this case, the findings prompted a revision of the understanding that was held about the problem. The design assumed that algorithmic opacity was a core issue, but participants already had a functional understanding of how their feed was assembled and already steered it semi-passively through cheaper options. What they were lacking was not an explanation of their feed, and what they grasped at in REFLECT was the higher level algorithmic controls like resetting, though it should be noted that their projected use of it was occasional. Ultimately, users endorsed algorithmic transparency as a general societal benefit, but simultaneously dismissed it as a habitual personal requirement.

The general finding is that the cost and benefit of each intervention lies on the same axis. A feature that removes the ease of consumption, permitting System 2, also removes what made the feed pleasant and hinders adoption. A feature that withholds a cue in order to encourage deliberation only moves the judgment to whatever cue is leftover. Inverting every System 1 feature of a platform therefore cannot be a design method on its own – just because there is room for critical thinking does not mean that it will be engaged, and a feed that supports deliberation will unfortunately lose to one that doesn't unless deliberation is made as cheap as consumption on the other. The design question should be reframed to question not how to add support for reflection, but where the path of least resistance leads and whether evidence can be placed on it.

These claims are bounded by a single-session lab evaluation of one artifact on neutral content with fifteen participants, and they say nothing about accuracy judgments, susceptibility, or sustained use. This work offers a traceable chain from mechanism to property to requirement to feature to observed response, a prototype that can be built on, and a clearer statement of the problem than it started with. The platforms people use were built to make engagement easy. A platform built to make thinking easy will have to be at least as easy to use.

11. Bibliography

- Apuke, O. D., & Omar, B. (2021). Social media affordances and information abundance: Enabling fake news sharing during the COVID-19 health crisis. *Health Informatics Journal*, 27(3), Article 14604582211021470. <https://doi.org/10.1177/14604582211021470>
- Bhowmik, T. (2025). Endless scrolling through social media and work boredom: A dynamic spillover of information overload. *Organization Management Journal*, 22(1), 38–47. <https://doi.org/10.1108/OMJ-06-2024-2225>
- Bhuiyan, M. M. (2023). *Combating problematic information online with dual process cognitive affordances* [Doctoral dissertation, Virginia Polytechnic Institute and State University]. VTechWorks. <http://hdl.handle.net/10919/115989>
- Brady, W. J., McLoughlin, K., Doan, T. N., & Crockett, M. J. (2021). How social learning amplifies moral outrage expression in online social networks. *Science Advances*, 7(33), Article eabe5641. <https://doi.org/10.1126/sciadv.abe5641>
- Cialdini, R. B. (2001). The science of persuasion. *Scientific American*, 284(2), 76–81. <https://doi.org/10.1038/scientificamerican0201-76>
- Darda, P., & Gupta, O. J. (2026). Shadows in the feed: A critical examination of dark patterns and user manipulation on social media. *Asian Journal of Business Ethics*, 15(2), 295–315. <https://doi.org/10.1007/s13520-026-00254-2>
- Garland, J., Bak-Coleman, J., Benesch, S., DeDeo, S., DiResta, R., Eissfeldt, J., Ha, S., Irons, J., Kempes, C., Lovato, J., Roschke, K., Smaldino, P. E., Stephenson, A. B., Wheatley, T., & Semenova, V.

- (2026). The case against efficiency: Friction in social media. *npj Complexity*, 3, Article 5. <https://doi.org/10.1038/s44260-025-00061-z>
- Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.
- Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., & Toombs, A. L. (2018). The dark (patterns) side of UX design. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)* (pp. 1–14). ACM. <https://doi.org/10.1145/3173574.3174108>
- Halpern, D., & Gibbs, J. (2013). Social media as a catalyst for online deliberation? Exploring the affordances of Facebook and YouTube for political expression. *Computers in Human Behavior*, 29(3), 1159–1168. <https://doi.org/10.1016/j.chb.2012.10.008>
- He, Z., & Zheng, Y. (2025). The influence of online news comment valence and like counts on the alienation of media trust. In *Proceedings of the 2025 2nd International Conference on Big Data and Digital Management (ICBDDM '25)* (pp. 691–695). ACM. <https://doi.org/10.1145/3768801.3768912>
- Hochman, G. (2024). Beyond the surface: A new perspective on dual-system theories in decision-making. *Behavioral Sciences*, 14(11), Article 1028. <https://doi.org/10.3390/bs1411028>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Khatri, V., Samuel, B. M., & Dennis, A. R. (2018). System 1 and System 2 cognition in the decision to adopt and use a new technology. *Information & Management*, 55(6), 709–724. <https://doi.org/10.1016/j.im.2018.03.002>
- Lau, J. (2025). If it is hard to understand, is it hard to be hooked? A meta-analysis of processing fluency and transportation. *International Journal of Advertising*, 44(2), 313–333. <https://doi.org/10.1080/02650487.2024.2428551>
- Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, 20(3), 881–900. <https://doi.org/10.1177/1461444816675438>
- Maranges, H. M., Schmeichel, B. J., & Baumeister, R. F. (2017). Comparing cognitive load and self-regulatory depletion: Effects on emotions and cognitions. *Learning and Instruction*, 51, 74–84. <https://doi.org/10.1016/j.learninstruc.2016.10.010>
- Martel, C., Pennycook, G., & Rand, D. G. (2020). Reliance on emotion promotes belief in fake news. *Cognitive Research: Principles and Implications*, 5, Article 47. <https://doi.org/10.1186/s41235-020-00252-3>
- Meinert, J. (2020). *All you need is a (heuristic) cue? An empirical investigation of the use of social media cues and features and underlying mechanisms for credibility judgments of news and political communication* [Doctoral dissertation, University of Duisburg-Essen]. DuEPublico. <https://doi.org/10.17185/duepublico/72856>

- Meinert, J., & Krämer, N. C. (2022). How the expertise heuristic accelerates decision-making and credibility judgments in social media by means of effort reduction. *PLOS ONE*, 17(3), Article e0264428. <https://doi.org/10.1371/journal.pone.0264428>
- Mildner, T., Freye, M., Savino, G.-L., Doyle, P. R., Cowan, B. R., & Malaka, R. (2023). Defending against the dark arts: Recognising dark patterns in social media. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference (DIS '23)* (pp. 2362–2374). ACM. <https://doi.org/10.1145/3563657.3595964>
- Moravec, P. L., Kim, A., & Dennis, A. R. (2020). Appealing to sense and sensibility: System 1 and System 2 interventions for fake news on social media. *Information Systems Research*, 31(3), 987–1006. <https://doi.org/10.1287/isre.2020.0927>
- Morris, M. R., Counts, S., Roseway, A., Hoff, A., & Schwarz, J. (2012). Tweeting is believing? Understanding microblog credibility perceptions. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work (CSCW '12)* (pp. 441–450). ACM. <https://doi.org/10.1145/2145204.2145274>
- Newman, E. J., Garry, M., Bernstein, D. M., Kantner, J., & Lindsay, D. S. (2012). Nonprobative photographs (or words) inflate truthiness. *Psychonomic Bulletin & Review*, 19(5), 969–974. <https://doi.org/10.3758/s13423-012-0292-0>
- Peacock, C., Scacco, J. M., & Stroud, N. J. (2019). The deliberative influence of comment section structure. *Journalism*, 20(6), 752–771. <https://doi.org/10.1177/1464884917711791>
- Pennycook, G., Bear, A., Collins, E. T., & Rand, D. G. (2020). The implied truth effect: Attaching warnings to a subset of fake news headlines increases perceived accuracy of headlines without warnings. *Management Science*, 66(11), 4944–4957. <https://doi.org/10.1287/mnsc.2019.3478>
- Pennycook, G., Cannon, T. D., & Rand, D. G. (2018). Prior exposure increases perceived accuracy of fake news. *Journal of Experimental Psychology: General*, 147(12), 1865–1880. <https://doi.org/10.1037/xge0000465>
- Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. *Nature*, 592(7855), 590–595. <https://doi.org/10.1038/s41586-021-03344-2>
- Pennycook, G., & Rand, D. G. (2021). The psychology of fake news. *Trends in Cognitive Sciences*, 25(5), 388–402. <https://doi.org/10.1016/j.tics.2021.02.007>
- Qi, M., & Tan, X. (2024). Willingness to share debunking information on social media—An exploration of influencing factors from the perspective of media affordance theory. *Science, Technology and Social Development Proceedings Series*, 1, 31–39. <https://doi.org/10.70088/cs978k90>

- Rhodes, S. C. (2022). Filter bubbles, echo chambers, and fake news: How social media conditions individuals to be less critical of political misinformation. *Political Communication*, 39(1), 1–22. <https://doi.org/10.1080/10584609.2021.1910887>
- Schrepp, M., Hinderks, A., & Thomaschewski, J. (2017). Design and evaluation of a short version of the User Experience Questionnaire (UEQ-S). *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6), 103–108. <https://doi.org/10.9781/ijimai.2017.09.001>
- Shugars, S., & Ha, E. (2025). The character of connection: Platform affordances and connective democracy. *Social Media + Society*, 11(2). <https://doi.org/10.1177/20563051251332427>
- Slovic, P., Finucane, M. L., Peters, E., & MacGregor, D. G. (2007). The affect heuristic. *European Journal of Operational Research*, 177(3), 1333–1352. <https://doi.org/10.1016/j.ejor.2005.04.006>
- Small, C., Bjorkegren, M., Erkkilä, T., Shaw, L., & Megill, C. (2021). Polis: Scaling deliberation by mapping high dimensional opinion spaces. *RECERCA. Revista de Pensament i Anàlisi*, 26(2). <https://doi.org/10.6035/recerca.5516>
- Steenbergen, M. R., Bächtiger, A., Spörndli, M., & Steiner, J. (2003). Measuring political deliberation: A discourse quality index. *Comparative European Politics*, 1(1), 21–48. <https://doi.org/10.1057/palgrave.cep.6110002>
- Sundar, S. S., Molina, M. D., & Cho, E. (2021). Seeing is believing: Is video modality more powerful in spreading fake news via online messaging apps? *Journal of Computer-Mediated Communication*, 26(6), 301–319. <https://doi.org/10.1093/jcmc/zmab010>
- Sweller, J., van Merriënboer, J. J. G., & Paas, F. G. W. C. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296. <https://doi.org/10.1023/A:1022193728205>
- Tomalin, M. (2023). Rethinking online friction in the information society. *Journal of Information Technology*, 38(1), 2–15. <https://doi.org/10.1177/02683962211067812>
- Wu, M., Wu, T., & Xiao, Y. (2025). Why people share misinformation on social media? An integration of affordance and flow theories. *Humanities and Social Sciences Communications*, 12, Article 1129. <https://doi.org/10.1057/s41599-025-05511-6>
- Zimmerman, J., Forlizzi, J., & Evenson, S. (2007). Research through design as a method for interaction design research in HCI. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '07)* (pp. 493–502). ACM. <https://doi.org/10.1145/1240624.1240704>

12. Appendices

Appendix A – Original Comparative Analysis Tables

Appendix A presents the full walkthrough observations from the Phase 1 comparative analysis. These were condensed, simplified, and sometimes rows were split for specificity in the coded matrices in Chapter 4. The original descriptive detail is preserved here.

Table A.1 – Algorithmic amplification

4.2 *algorithm logic visible* was added at coding as a subset of *user control over algorithm*. The controls recorded under *user control over algorithm* are input-level only. The row makes that absence explicit rather than leaving it to be inferred from the other rows.

Variable / Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Default feed type	Algorithmic “For You”	Algorithmic	Algorithmic	Algorithmic	Algorithmic “For You”	Algorithmic
User control over algorithm	Moderate: add topics, interests, explore settings (very limited), mute and block certain words	Low: toggle political and sensitive content	Moderate: add topics, interests, mute words, tags. Adult content is off by default and must be changed via web, not app.	Low, toggle off certain ad content, Adult content off by default and must be changed via web, not app.	High : Add topics, interests, mute and block, STEM feed optional but off by default	High: Sensitive and adult content can be moderated (high, standard, low) but not off. Add or block certain interests, words, or expressions.
Steps to reach algorithm control	3, nested under privacy and safety, less intuitive	4, must tap between different controls	3, many subsections, must tap between controls	4	4, many subsections	4, many subsections
Chronological feed availability	No	No	Replies can be chronologically sorted in settings	Replies can be chronologically sorted in settings	No	No

Table A.2 – Immersive design

A.2 auto-advance between units and *immersion entry* were merged into *escalation to stream*, which records whether the card format converts to a continuous stream and what triggers it.

A.2 content unit size was absorbed into *default unit format*, coded as card or stream.

A.2 immersion exit became *exit cost*.

A.2 transition types showed no variation across the six platforms and is reported in prose rather than as a table row.

Variable / → Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Auto-advance between units	Card feed, no continuous stream	Card feed, no continuous stream	No infinite scroll in main feed; continuous only in video section, comments, reposts	Opening a video does not advance to others	Opens into a continuous video stream	Continuous in Reels; main feed card-based
Transition types	Swipe up	Swipe up	Swipe up	Swipe up, must tap into post, swipe back to return to feed when finished	Swipe up	Swipe up
Feed termination / stopping cue	None: continuous loading	None: continuous loading	Main feed does not scroll endlessly; continuous in the video section	None, but tapping in and back inserts a boundary between units	None: no unit edge or point of return	None in Reels; main feed is a continuous card list
Content unit size	Small, enlarges to full screen	Large, larger when video opened	Full screen for video, card otherwise	Mostly text cards	Full screen	Large, full screen when tapped

Variable / → Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
	when video opened					
Autoplay (retention framing)	Yes, in-feed	Yes, default	Yes, default	Yes, default, but toggleable	Yes, core mode	Yes, default
Immersion entry	Low: autoplay, into video scroll if tapped	Low: autoplay, into video scroll if tapped	Moderate: autoplay, but cannot scroll into other videos	N/A: no dedicated immersive video feed to enter	None: autoplay and immediate immersive feed	Low: taps into Reels, often accidental
Immersion exit	Low: swipe to exit	Low: swipe to exit	Low to moderate: two swipes back	Low: symmetric in and out / no	High: must close app	Moderate: swipe out, but a lot of videos also built into feed

Table A.3 – Engagement and sharing metrics

A.3 *visible metrics* was split into one row per metric — like, comment, view, save and circulation count — so that presence could be compared metric by metric rather than read out of a list.

A.3 *counts exact or rounded* became *count precision*.

Variable / → Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Visible Metrics	Likes, comments, reposts, views	Reactions, comments, shares, saves, views	Likes, comments, reposts	Net upvote score, comments, awards; downvotes hidden	Likes, comments, saves, shares, views	Likes, comments, reposts, shares, view counts on profile

Variable / → Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Counts exact or rounded	Exact under 1k, rounded above	Likes rounded to nearest thousand; rest exact	Exact under 1k, rounded above	Exact comments and score; downvotes not shown	Exact under 1k, rounded above	Exact into the tens of thousands

Table A.4 – Friction

A.4 rows are carried into 4.5 unchanged.

Variable / → Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Taps to share	2 to repost, more with a comment , 2 to share privately	2 to share to a story; private sharing depends on the poster's settings; share flow includes audience controls	2 to repost; 3 to share via direct message; 4 via the share-via menu	2+ to crosspost (a community must be chosen); 2 to share privately	1 to repost; 3 to send privately	1 to repost; 1 to send to recents via slide, otherwise 3
Visibility of source when sharing	No preview or context when initiating a share	No context reminder about the post when sharing	Article link shown in-post, but inconsistently	Subreddit, title, and counts shown; poster not surfaced	Sender and poster visible; original source rarely noted	Poster visible when sharing, but original source of the claim not shown
Taps to report or flag	3+ (menu, reason, sometimes more)	3+ (menu, reason, variable steps, confirm)	About 4 steps to submit	5 to 6 steps	2 to reach flag, more to complete	2 to reach flag, more depending on category

Table A.5 – Content format

A.5 format hierarchy in navigation was split: the feed-format half was absorbed into *dominant unit format*, and the navigation half became *video position in navigation*.

A.5 required vs. optional reading was split into *required reading to consume a unit* and *text truncated by default*.

A.5 dominant format in feed became *dominant unit format* and is measured in 6.6 only; 6.3 refers to it in prose rather than reproducing the row.

Variable / → Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Dominant format in feed	Mixed: photo and video, text posts allowed	Video and photo	Mixed: text, photo, video	Text-led; if there is a photo or video it is tied to the post	Video, full screen	Video and photo
Required vs optional reading	Captions short; long posts truncated	Truncated, "see more"	Text posts shown; longer truncated	Reading required to consume a post	Caption optional, easily skipped	Caption truncated, easily skipped
Format hierarchy in navigation	Mixed home; video tab secondary	Mixed home; Reels tab prominent	Text-centric home; video secondary	Text-centric home; no format-privileged tab	Primary tab is the video feed f	Photo and video home; Reels prominent

Table A.6 – Deliberation affordances

A.6 rows are carried into 6.7 unchanged and only simplified for legibility.

Variable / Platform →	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Clickable links	Restricted: links post but are	Clickable in posts and comments	Clickable in posts and	Clickabl e in posts and	None: links not rendered	None: links not rendered clickable in captions or

	de-emphasized in ranking		comments	comments	clickable	comments
Discussion structure	Threaded	Threaded to 2 levels	Threaded	Deep threaded	Minimal 1 level	Minimal 1 level
Default comment ordering	Engagement	Engagement	Engagement	Engagement	Engagement	Engagement
Argument-traceability controls	Reply, @ mention, quote post, permalink to any post in a thread	Reply, @ mention, tag, share of an individual comment	Reply, @ mention, quote post, permalink	Reply, u/ mention, permalink to a single comment, crosspost	Reply, @ mention, video reply to a comment, stitch and duet	Reply, @ mention, tag; quotation only by reposting to a new video or story; no comment permalink
Attached corrective context	Community Notes written and rated by contributors, shown on the post above the replies	Platform: third-party fact-check labels applied to some posts	Platform-adjacent: opt-in labeler services applied at post or account level	None platform-side; context lives in the thread and in moderator posts	Platform: topic warning labels on some content	Platform: fact-check labels applied to some posts

Table A.7 – Synthesis

Ratings indicate the degree to which each dimension's design affords System 1 processing (High = strongly affords heuristic processing). E.x., for the friction row, "High" denotes the absence of friction affording System 1, not to be confused with having "High" friction.

Dimension	X	Facebook	Bluesky	Reddit	TikTok	Instagram
Algorithmic Amplification	<i>Moderate:</i> algorithmic default,	<i>High:</i> algorithmic default,	<i>Moderate:</i> algorithmic default,	<i>High:</i> algorithmic default,	<i>Moderate:</i> algorithmic default,	<i>Moderate:</i> algorithmic default,

	logic invisible, moderate input controls, no chronological feed	logic invisible, narrowest controls, no chronological feed	logic invisible, moderate controls, chronological sorting for replies only	logic invisible, low control, chronological sorting for replies only	logic invisible, but broadest input controls	logic invisible, but high input control
Immersive design	<i>Moderate:</i> card feed, in-feed autoplay, low entry and exit cost	<i>Moderate:</i> card feed with autoplay; full-screen stream secondary in Reels	<i>Moderate:</i> card feed, main feed not continuous, stream confined to video tab, low-to-moderate exit	<i>Low:</i> text cards, advance requires a decision (tap in and out), autoplay toggleable	<i>High:</i> continuous full-screen stream, autoplay as core mode, high exit cost	<i>High:</i> card main feed but stream in Reels, autoplay default, moderate exit cost
Engagement and sharing mechanics	<i>Moderate-to-high:</i> likes, comments, reposts and in-feed view counts (only platform), rounded above 1k	<i>Moderate:</i> reactions, comments, shares, saves; likes rounded	<i>Moderate:</i> likes, comments, reposts; rounded above 1k	<i>Low-to-moderate:</i> exact net score encodes a verdict, but downvotes hidden and no circulation count	<i>High:</i> full metric set including saves and shares, rounding above 1k	<i>High:</i> full metric set, counts exact into the tens of thousands
Friction	<i>Moderate-to-high:</i> two-tap repost, no source shown when sharing, 3+ taps to report	<i>Moderate:</i> two-tap repost, private sharing gated by poster settings, no source shown, 3+ taps to report	<i>Moderate:</i> two-tap repost, three-tap private share, partial source visibility, ~4 taps to report	<i>Low:</i> repost requires choosing a community, partial source visibility, 5–6 taps to report	<i>High:</i> one-tap repost, no source shown, two taps to reach flag	<i>High:</i> one-tap repost and slide-to-send, no source shown, two taps to reach flag

Content format	Moderate: mixed units, in-feed autoplay, text truncated, video tab secondary	Moderate-to-high: video and image dominant, autoplay default, text truncated, Reels prominent	Moderate: mixed but text-centric, autoplay confined to video section, video secondary	Low: text-led, reading required, video must be tapped into, no video tab	High: full-screen video, autoplay as core mode, caption skippable, video tab primary	High: video and image dominant, autoplay default, caption truncated, Reels prominent
Deliberation affordances	Moderate: threaded discussion and community corrective context, but links de-emphasized and comments ordered by engagement	Moderate: clickable links and platform fact-check labels, but shallow threading and engagement ordering	Moderate: clickable links and threading, but corrective context only via opt-in labelers	Low: deep threading, quotes stay in-thread, links clickable — though corrective context is community-side only	High: no clickable links, shallow threading, quotes become new videos, engagement ordering	High: no clickable links, shallow threading, no quote function, engagement ordering

Appendix B – AI Prompt Examples

Representative prompts illustrating the AI use disclosed in Section 3.2. Prompts are reproduced in abbreviated form; project file paths and identifiers are omitted.

- "I'm comparing how social media interfaces afford automatic versus reflective processing, using dual process theory. Suggest search keywords and phrasings for finding HCI and misinformation literature on this — e.g. around friction, cognitive load, and interface nudges."
- "Here is the patch sheet for the next prototype iteration: make the algorithm panel more compact, drop the spectrum element, reorder the ad placement, and add a stopping modal at the end of the feed. Implement exactly these changes and give me a changelog entry for each; change nothing else."
- "In the wireframes, tapping the button on a card opens the context panel. Change this so the whole card is tappable and add small icons for the post metadata. Update the build and log the change."
- "Based on my four design goals, ideate candidate post-session interview questions that compare the participant's own app with the prototype. I'll select and rephrase the ones I use."

- "Build the coding workbook: one tab per participant, columns for segment ID, speaker, timestamp, dimension code, stance, evidence type, flag, and memo, with the summary sheet reading from the tabs for easy organization by dimension when the participant sheets are populated."
- "Render the mapping between the log variables and the coding dimensions as a table I can use as a starting point for the methodology chapter."
- "Suggest alternative phrasing for ____"
- "These are my plans for the study protocol: _____. Please suggest any improvements or shortcomings for me to change."

Appendix C – REFLECT Prototype Changelog

Development history of the prototype from first wireframe to the build frozen for the study. The complete source code is publicly available at <https://github.com/oliviakusch/REFLECT>.

Table C.1 – Prototype changelog and design reason

Date (2026)	Change	Design reason
13 to 17 Jul	Wireframes v1 to v5: v1 and v2 set the bounded card feed and content unit; v3 added the community context layer; v4 added the transparency layer (topic reasoning, algorithm panel); v5 consolidated for development	Concept fixed on paper before any code
by 20 Jul	First working build (index.html + content.js): 29-post fixed pool, all screens, full event logging	Wireframe v5 translated into a working, instrumented environment
21 Jul	Gap analysis: the build audited feature by feature against the comparative-analysis findings	Self-audit found controls not reaching the mechanism and reporting missing; produced the incremental change plan
21 Jul	Content pool v2 (29 to 35 posts, topics densified); topic filters made genuinely functional (more, less, mute, diversify) under the unseen-only guardrail; comments rewritten casual but civil	Filters needed content density to be noticeable; guardrail keeps the early feed identical for everyone; realistic register while tone stays a controlled property
23 Jul	Share sheet given nominal targets (Copy link, WhatsApp, Messages, Other) with honest study note; claim made visually prominent on the context surface; source pages reframed as samples with in-app back navigation	Completes the share funnel for measurement; verification never ejects the user from the app

23 Jul	Algorithm screen matured: status banners (paused, cleared), add-a-topic (nominal, disclosed as such), stable topic colors, reset and pause, diversify as equal-distribution view	One legible control surface; honesty wherever the fixed pool limits real effects
27 Jul	User reporting flow: Report button in the context action bar, two-tap reason sheet, confirmation with undo; disclosure modal gains a report link	Friction symmetry: reporting costs the same two taps as sharing, closing the diagnosed asymmetry
28 Jul	Old-browser layout fix (CSS fallbacks for inset and env paddings)	Lab phone compatibility
28 Jul	Batch A: session survives a reload (same session id, continuous timing); algorithm button and topic chips restyled to read as buttons; photos downscaled 44.6 MB to 2.5 MB	Data integrity for study day; affordance legibility
28 Jul	Batch B: “Community context” renamed “Context & sources” with sources leading the meta line; composer source field promoted (“Back it up with a source”); algorithm topic list collapsed; “changes apply from here on” made visible, including the in-feed “Updated from here” divider; section labels emphasized	Language and emphasis iteration ahead of the first sessions
28 Jul	Own-phone hardening built and tested but deliberately not shipped (parked)	Scope discipline the night before sessions
29 Jul	Study sessions begin on the frozen build	

Appendix D – Prototype Action Logs

Each event was logged with a timestamp, the participant identifier, and the post concerned where applicable. Logs were exported as one CSV file per participant.

Table D.1 – Events tracked via REFLECT’s action log functionality

Event	Description
task_completed	A task was marked as completed
assist	Assistance was given during a task
post_view	A post was scrolled into view
video_play	A video post started playing
context_opened	The context panel of a post was opened
source_click	A source link was opened

source_nav	The participant moved between sources
disclosure_opened	A disclosure notice was expanded
vote	An up- or downvote was cast on a post
source_upvote	A source was upvoted
comment_opened	The comment section of a post was opened
comment_expanded_replies	Replies under a comment were expanded
share_initiated	A share was started
share_prompt_shown	The share-with-source prompt was shown
share_target_selected	A share destination was chosen
share_completed	A share was completed
share_abandoned	A share was exited before completion
flag_opened	The reporting flow was opened
flag_abandoned	The reporting flow was closed without a report
controls_opened	The curation controls panel was opened
controls_adjusted	A curation control was changed
topic_added	A topic was added to the topic filters
personalization_paused	Personalization was paused
personalization_resumed	Personalization was resumed
algorithm_reset	Personalization was reset
stopping_cue_reached	The end-of-feed stopping cue was shown
continue_attempt	The participant continued past the stopping cue
segment_cue_reached	A segment card was shown mid-feed
segment_scroll_attempt	A scroll was attempted at a segment card
segment_continue	The participant continued past a segment card

Appendix E – Screening Questionnaire

This is the screening questionnaire, directly as distributed to participants in Qualtrics via LABEX.

Thank you for your interest in this study. Please read the information below before starting.

What is this questionnaire?

This short questionnaire (about **3 minutes**) helps us select a mix of participants for a study on the design of a social media feed, in which you would try out an early-stage prototype and share your impressions while using it. The study is conducted and financed by Olivia Kusch and Professor Mauro Cherubini of HEC Lausanne, University of Lausanne (UNIL).

This questionnaire asks a few questions about how you use social media, your availability, and your contact details.

Compensation.

Completing this questionnaire is not compensated. **If you are selected and take part in the lab session (about 45 minutes), you will receive 25 CHF in cash at the session.**

Selection

Places are limited and we are selecting a mix of participants, so not everyone who completes this questionnaire can be invited. If you are selected, you will receive an invitation email with a registration link. Completing this questionnaire does not commit you to taking part.

Your data

Your contact details (first name, last name, email address) are used only to organize participation and are shared for that purpose only with the HEC-LABEX recruitment team. They are stored separately from your questionnaire answers on restricted-access storage, treated confidentially, and deleted after recruitment ends for everyone who is not selected.

Your answers are used only for participant selection in this academic research project. You can request access to, correction of, or deletion of your personal data at any time by contacting the researcher below.

Voluntary participation

Taking part is voluntary. You can stop at any time by closing this page; incomplete answers are not used.

Contact

Questions about this questionnaire or the study: Olivia Kusch, University of Lausanne — Olivia.Kusch@unil.ch.

Questions about the recruitment process: the HEC-LABEX team.

- ☐ I agree — I have read and understood the information above and I agree to take part in this questionnaire
- ☐ I do not agree

Q2

Which social media app do you use the most?

- ☐ Instagram
- ☐ TikTok
- ☐ X (formerly Twitter)
- ☐ Reddit
- ☐ Facebook
- ☐ Bluesky
- ☐ Other
- ☐ I generally do not use social media

Do any of the following currently apply to you, or have they applied in the past two years? Select all that apply.

- ☐ I work in or study healthcare or medicine
- ☐ I work in or study management, finance, communications
- ☐ I work in or study finance
- ☐ I work in or study IT, software development, data
- ☐ I work in or study teaching or education
- ☐ I work in or study public sector or politics
- ☐ I work, or have worked in social media, or have participated in a study relevant to social media
- ☐ I work in math, statistics, physics
- ☐ I work in biological science, medicine, or environmental studies
- ☐ I work in retail or trades (manual labor)
- ☐ Please specify your area of work/study if it is not listed here
- ☐ I am not currently working or studying

How much do you agree with each of the following statements about how you use social media?

	Strongly Disagree	Somewhat disagree	Neither agree nor disagree	Somewhat agree	Strongly agree
I often scroll mindlessly through my feed.	☐	☐	☐	☐	☐
I often continue scrolling even when I want to stop	☐	☐	☐	☐	☐
I like to look up information before deciding my thoughts about something	☐	☐	☐	☐	☐
I often mute, unfollow, or change settings in order to shape my feed/algorithm	☐	☐	☐	☐	☐
Trying out a new app comes easily to me	☐	☐	☐	☐	☐
For this row, please select "Somewhat disagree"	☐	☐	☐	☐	☐
I generally believe that social media has done harm to society	☐	☐	☐	☐	☐

Q4

★

If you are selected, the session (about 45 minutes) will take place at the University of Lausanne (Internef building) between Wednesday 29 July and Friday 7 August 2026. Exact time of day can be flexible.

On which days would you be able to attend? Select all that apply.

- ☐ Wednesday, July 29th
- ☐ Thursday, July 30th
- ☐ Friday, July 31st
- ☐ Monday, August 3rd
- ☐ Tuesday, August 4th
- ☐ Wednesday, August 5th
- ☐ Thursday, August 6th
- ☐ Friday, August 7th
- ☐ Monday, August 10th
- ☐ Tuesday, August 11th
- ☐ I am not available on any of these days

Q5

Please enter your contact details so that we can contact you if you are selected:

First Name

Last Name

Email address

Appendix F – Moderator Guide

REFLECT

Moderator Guide

Before the session

- **Device & prototype** — loaded and reset; event logging on and confirmed.
- **Check Participant Number**
- **Materials**
 - Consent form
 - Task cards
 - UEQ-S + concept items (×2)
 - Interview guide
 - Timer/Audio Recording
- **Sign into Moderator View**
- **Validate link**
 - **connect to hotspot**
 - **node server.js in vs code**
 - **ipconfig in terminal – validate what is under adresse IPv4**

1 · Welcome & consent (4 min)

Say:

“

Thanks for coming in. Today you'll use two apps — one you already use, and a prototype that has been designed — and do a few short everyday tasks on each. There are no right or wrong answers, this is not my own application and I will not be offended if there is something you do not like – we are testing the design, not you. I'll ask you to think aloud: just say whatever you're noticing, wondering, or reacting to. It might feel odd at first, but gathering your thoughts about your experiences on both of these applications helps me understand what works and what doesn't. You can stop at any time.

”

SIGN CONSENT FORM

START AUDIO RECORDING

Icebreaker Questions

2 · Block 1 (11 min)

Say: *“I'll give you a task and you can go ahead and perform it as you would normally. Please think through what you are seeing— thinking aloud. I'll mostly stay quiet – don't worry about what I am thinking or doing.”*

If for prototype

Say: *“We'll start with **the prototype**. As you complete these tasks, feel free to click around and see what you notice, what sticks out to you. Please think aloud as you go, tell me what you are doing, noticing, and thinking about what you are seeing as you are comfortable. Feel free to tap around (logos, topics, etc) if you're intrigued by anything, even if it's not explicitly part of the task. Every part of the app is functional except for a profile and settings view as this is outside the scope of the study.”*

If for own app

Say: *“We'll start with **whichever app you normally use**. If you are comfortable, please use your own device, and feel free to mute your volume or use headphones if you prefer to see your feed in private.*

If you prefer to not use your own device, you may use this one, but please consider what your thoughts may be if it were your own feed or the comfort of using a device you are already familiar with.

As you complete these tasks, feel free to click around and see what you notice, what sticks out to you. Please think aloud as you go, tell me what you are doing, noticing, and thinking about what you are seeing as you are comfortable.”

Task 1 — Browse

Hand card. *“Browse the feed the way you normally would. Stop whenever you'd naturally stop.”*

Do: Do not prompt them to continue or to stop — the stopping point is the measure.

Do: Note posts-to-stop and, in the prototype, whether they reach or remark on the end-of-feed state.

Task 2 — Judge

Hand card. *“Find a post that makes a claim. Decide whether you'd trust it, and walk me through how you'd check it.”*

Do: Observe context opens and source-clicks; let them lead.

Do: Report-path noticing (UC6): if they say a post seems misleading, note whether they look for a way to flag it unprompted. If they do not raise it by the end of the task, ask the light probe below — once, neutrally.

Do: Probe (only if not raised): “If you thought a post here was misleading, is there anything you'd do about it?”
Do not tell them to report, and do not point to the control.

Task 3 — Interact

Hand card. “Show me how you would interact with this application”

Do: Observe hesitation or backing out at the accuracy prompt (prototype); do not lead or reassure.

Do: In the prototype, sharing completes locally — let it happen; no need to explain unless asked afterward.

Task 4 — Adjust

Hand card. “Imagine you've been seeing too much of one topic lately. Show me how you would change what your feed shows you.”

Do: Do not point to the controls — whether they find them is the measure.

Do: Note any comments about feeling in / out of control.

3 · Questionnaire (3 min)

Do: Administer UEQ-S

4 · Block 2 (11 min)

Say: “Now we'll do the same kinds of tasks on the other app.”

Do: Repeat the relevant task set (four for the prototype, three for the own app), then administer the identical questionnaire.

Task 1 — Browse

Hand card. “Browse the feed the way you normally would. Stop whenever you'd naturally stop.”

Do: Do not prompt them to continue or to stop — the stopping point is the measure.

Do: Note posts-to-stop and, in the prototype, whether they reach or remark on the end-of-feed state.

Task 2 — Judge

Hand card. “Find a post that makes a claim. Decide whether you'd trust it, and walk me through how you'd check it.”

Do: Observe context opens and source-clicks; let them lead.

Do: Report-path noticing (UC6): if they say a post seems misleading, note whether they look for a way to flag it unprompted. If they do not raise it by the end of the task, ask the light probe below — once, neutrally.

Do: Probe (only if not raised): “If you thought a post here was misleading, is there anything you'd do about it?”
Do not tell them to report, and do not point to the control.

Task 3 — Interact

Hand card. “Show me how you would interact with this application”

Do: If no share is attempted spontaneously, prompt once.

Do: In the prototype, sharing completes locally — let it happen; no need to explain unless asked afterward.

Task 4 — Adjust

Hand card. “Imagine you're seeing too much of one topic lately or want to change some aspect of your algorithm. Show me how you would change what your feed shows you.”

Do: Do not point to the controls — whether they find them is the measure.

Do: Note any comments about feeling in / out of control.

5 · Questionnaire (3 min)

Do: Administer UEQ-S

6 · Comparative interview (10 min)

Overall Comparison

- **"You've now used both [own app] and REFLECT. Which did you prefer overall — and what drove that?"**
 - *Probe:* "Was there a moment in either one that sums it up for you?" [RQ2]
- **"Of everything you saw in REFLECT, what would you actually use if it existed for real? And what would you avoid, or what annoyed you?"**
 - *Probe:* "What makes x feature [useful / annoying] for you?" [RQ2]
- **"Was there anything in REFLECT that slowed you down compared to [own app]? How did that feel — worth it, or irritating?" [RQ2, DG6]**

Behavior Probes

- **"When you were deciding whether to trust [post X], you [opened the context panel / clicked through to the source]. Walk me through what you were looking for."**
 - *Probe:* "How did that compare with how you'd decide on [own app]?" *If they never opened context:* "What would have made you want to open it?" [DG3, RQ2]
- **"In the first task, the feed came to an end instead of continuing. What went through your mind at that moment?"**
 - *Probe:* "How do browsing sessions usually end for you on [own app]?" [DG2]
- **"While you were using [app], I noticed you [hesitated at the accuracy prompt / went straight to X] — can you walk me through what was going on for you there?" (Use a real observed moment.)**
- **"When you adjusted what your feed shows you, did what happened match what you expected?"**
 - *Probe:* "In real life, would you use controls like that — regularly, once and forget, or never?" [DG1, RQ4]
- **"Did you notice the feed had no like or share counts until after you'd voted? Did you miss them?"**
 - *Probe:* "When the numbers appeared, did they change what you thought of the post?" [DG4]

Stepping back

- **"I saw you doing X, explain what led you to this?"**
- **"Imagine using something like REFLECT for a month instead of [own app]. What would be different about how you scroll, what you share, or how you judge what you see?"**
 - *Probe:* "Is that a change you'd want?" [RQ3]
- **"Did anything you did today make you think differently about your own habits — how you scroll or what you share?" [RQ3]**

Would they actually use it?

- **"Realistically — what would stop you from using something like this?"**
 - *Prompts only if silent:* friends not on it · effort · trusting who writes the context · privacy. [RQ2]
- **"If you could change one thing about REFLECT before it launched, what would it be?"**

Design exercise (if time, ~3 min)

Say: "Last thing — no right answers. Sketch or describe one change you'd make to your own social media app to make it better for you. Anything goes — a button, a screen, a rule."

Do: ~2 minutes; have them explain the sketch in one sentence. Photograph the sheet with the participant number. Skip entirely if you're past 42 min.

Anything else you expected me to ask about that I didn't, or anything else you want to add?

6 · Debrief (2 min)

Say: *"That's everything — thank you. This prototype is a research concept exploring how interface design can support more reflective use of social media. The content was curated for the study, and nothing you did was sent anywhere. Do you have any questions for me?"*

SIGN CASH FORM

End of Session Checklist

- Cash form
- Completion in ORSEE
- Close and download moderator session on app

Appendix G – Interview Coding & Summary Sheet

Table G.1– Field definitions / schema for interview coding

One row per coded segment. Column letters match the workbook exactly.

Col	Field	Definition and allowed values
A	seg_id	Auto ID (P01-001 ...), formula-driven off the tab name. Do not type into this column.
B	line_ref	Paragraph number in the source transcript .docx.
C	timestamp	mm:ss in the recording. Included but blank
D	speaker	participant / moderator.
E	block	baseline / reflect / interview / setup. Attribute from transcript cue phrases, never from the action logs (Appendix C).
F	task	T1 discover / T2 judge / T3 share / T4 control / n/a. The study runs four tasks; the comment-reply task was cut 2026-07-22 — there is no fifth.
G	quote	Verbatim quote (or tight paraphrase marked as such in the memo).

H	flag	blank / check audio / unclear / quote for writeup.
I	memo	Mandatory whenever code_primary is not none. Carries the granularity the deleted sub-codes held: what exactly, which feature, what mechanism.
J - K	code_primary	D1–D6, or none for rows kept for context only.
L–M	code_secondary	D1–D6 or blank. Optional — use sparingly, only when a segment genuinely does double duty.
N	stance	positive / negative / mixed / neutral — toward the mechanism the code names, not toward the app overall.
O	evidence_type	observed (it happened in session) / reported (they say they do it) / projected (they say they would). See §5.
P–S	tag_usability ... tag_change	x or blank, one column per tag (usability, barrier, suggestion, novelty, moderator, change).

Table G.2– Example of Interview Coding

	A	B	C	D	E	G	H
21	seg_id	line_ref	timestamp	speaker	block	quote	flag
101	P05-080	80		participant	interview	Yes, I think. The way I use social media is definitely more like, I'm part of society. And I don't really go to it for information. Even though that is probably where I get the majority of my information, I'm not necessarily looking to go there for information. If I have to like look at studies and stuff, then it kind of feels like a chore.	quote for writeup
102	P05-081	81		moderator	interview	Is there anything else about the app in particular that you would find was particularly limiting or frustrating?	
103	P05-082	82		participant	interview	I think your app would attract a different audience to what I would use, because I want to be able to interact with the posts much easier.	quote for writeup
104	P05-083	83		moderator	interview	That makes sense.	
105	P05-084	84		participant	interview	I'm sorry, it's like a different audience. He's always getting... He's always on the news and like doing that stuff sort of thing. I feel like your app is sort of a more condensed... like, condensed news, which is, like, easier to access than, like, having to read a whole...	quote for writeup

	H	I	J	K	L	M	N	O
	flag	memo	code_primary	code	code_secondary	code	stance	evidence_type
quote for writeup		don't use social media for information, but social media gives a sense of community/society. Intentionally looking at stuff for information feels like a chore	D1 — algorithmic amplification	D1			mixed	observed
quote for writeup		reflect would attract a different audience because of the information nature	D2 — immersive design: bound	D2	D6 — source traceability / deliberati	D6	mixed	projected
quote for writeup		reflect would attract a different audience because of the information nature	D2 — immersive design: bound	D2	D6 — source traceability / deliberati	D6	positive	projected

Table G.3 – Action Log Summary

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	
1	Event Summary — REFLECT prototype action logs																				
3																					
4	Participant overview																				
5	Participant	Total events	Last event (min)	ks	stat	comp	Assists needed	end													
30		task_completed	3	4	3	3	2	2	1	3	2	1	2	2	—	—	—	28	12	3	
31		assist	—	—	2	1	1	1	—	—	1	—	—	—	—	—	—	6	5	10	
32	Feed browsing	post_view	35	18	11	12	21	9	11	16	11	35	14	15	11	11	30	260	15	0	
33		video_play	2	2	—	1	—	—	—	1	—	—	1	—	—	—	—	7	5	10	
34	Context & sources	context_opened	7	5	6	8	8	4	1	2	3	5	9	4	7	10	11	90	15	0	
35		source_click	1	5	1	6	3	1	2	1	3	—	5	2	2	3	2	37	14	1	
36		source_nav	1	4	1	6	3	1	2	1	3	—	5	2	2	3	2	36	14	1	
37		disclosure_opened	1	—	—	—	—	—	—	—	—	3	—	—	—	—	4	8	3	12	
38	Voting & comments	vote	1	1	1	2	4	7	—	—	5	—	—	8	8	—	—	37	9	6	
39		source_upvote	—	—	—	—	1	—	—	—	—	—	—	—	—	—	—	1	1	14	
40		comment_opened	5	5	3	6	5	3	1	1	3	3	6	2	3	3	4	53	15	0	
41		comment_expanded_replies	—	1	—	—	1	—	1	—	—	1	2	—	—	—	—	6	5	10	
42	Sharing	share_initiated	2	2	1	3	1	1	1	1	1	1	1	1	1	1	1	19	15	0	
43		share_prompt_shown	2	2	1	3	1	1	1	1	1	1	1	1	1	1	1	19	15	0	
44		share_target_selected	1	1	1	1	—	—	—	—	—	—	—	—	—	—	—	4	4	11	
45		share_completed	1	1	1	1	—	—	—	—	—	—	—	—	—	—	—	4	4	11	
46		share_abandoned	1	1	—	2	1	1	1	—	—	—	1	1	—	—	—	9	8	7	
47	Flagging	flag_opened	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	2	2	13	
48		flag_abandoned	—	—	—	1	—	1	—	—	—	—	—	—	—	—	—	2	2	13	
49	Feed controls & personalization	controls_opened	2	3	6	11	17	7	6	5	5	3	5	7	11	10	7	105	15	0	
50		controls_adjusted	—	6	—	10	21	15	1	—	—	5	6	7	—	2	10	83	10	5	
51		topic_added	—	—	—	1	1	—	—	—	—	—	—	—	—	—	—	2	2	13	
52		personalization_paused	—	—	—	1	1	1	—	—	—	—	—	—	—	—	1	4	4	11	
53		personalization_resumed	—	—	—	—	—	1	—	—	—	—	—	—	—	—	—	1	1	14	
54		algorithm_reset	—	—	—	—	—	—	—	—	—	1	—	—	—	—	2	3	2	13	
55	Stopping & segment cues	stopping_cue_reached	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	14	
56		continue_attempt	1	—	—	—	—	—	—	—	—	—	—	—	—	—	—	1	1	14	
57		segment_cue_reached	1	1	1	1	1	—	1	1	1	1	1	—	1	1	1	13	13	2	
58		segment_scroll_attempt	1	3	1	2	2	—	—	2	1	1	1	—	1	—	1	16	11	4	
59		segment_continue	1	1	—	—	1	—	—	1	—	1	1	—	—	—	1	7	7	8	
60		All events	78	74	46	89	102	61	34	43	39	68	69	49	50	54	79	935			
61																					
Summary P01 P02 P03 P04 P05 P06 P07 P08 P09 P10 P11 P12 P13 P14 P15 + : <																					
Reads Accessibility Investigate																					

Appendix H – Supplemental UEQ Tables

Table H.1 – UEQ-S Scale means by condition

Scale	REFLECT - M (SD) [95% CI]	Baseline app - M (SD) [95% CI]	p
UEQ-S Overall	4.93 (0.90) [4.48, 5.39]	4.75 (0.61) [4.44, 5.06]	.57
Pragmatic Quality	5.13 (1.27) [4.49, 5.77]	5.37 (0.83) [4.94, 5.79]	.60
Hedonic Quality	4.73 (1.04) [4.21, 5.26]	4.13 (0.98) [3.64, 4.63]	.14

Table H.2. REFLECT compared to participant baseline app UEQ data

Scale	REFLECT						Baseline app					
	Mean	STD	N	Confidence	Confidence Interval		Mean	STD	N	Confidence	Confidence Interval	
UEQ-S Overall	4,93	0,90	15	0,45	4,48 5,39		4,75	0,61	15	0,31	4,44 5,06	
Pragmatic Quality	5,13	1,27	15	0,64	4,49 5,77		5,37	0,83	15	0,42	4,94 5,79	
Hedonic Quality	4,73	1,04	15	0,53	4,21 5,26		4,13	0,98	15	0,49	3,64 4,63	

Table H.3 – REFLECT against UEQ-S benchmark

Item	Mean	Variance	Std. Dev.	No.	Negative	Positive	Scale		Short UEQ Scales	
1	↑ 1,5	1,6	1,2	15	obstructive	supportive	Pragmatic Quality		Pragmatic Quality	↑ 1,133
2	↑ 0,8	2,5	1,6	15	complicated	easy	Pragmatic Quality		Hedonic Quality	→ 0,733
3	↑ 1,3	1,5	1,2	15	inefficient	efficient	Pragmatic Quality		Overall	↑ 0,933
4	↑ 0,9	3,4	1,8	15	confusing	clear	Pragmatic Quality			
5	→ -0,2	1,2	1,1	15	boring	exciting	Hedonic Quality			
6	↑ 1,4	1,5	1,2	15	not interesting	interesting	Hedonic Quality			
7	↑ 1,1	2,1	1,5	15	conventional	inventive	Hedonic Quality			
8	→ 0,6	3,5	1,9	15	usual	leading edge	Hedonic Quality			

Table H.4 – REFLECT mean value per item

Y-axis references the items and is ordered the same as in Table H.3

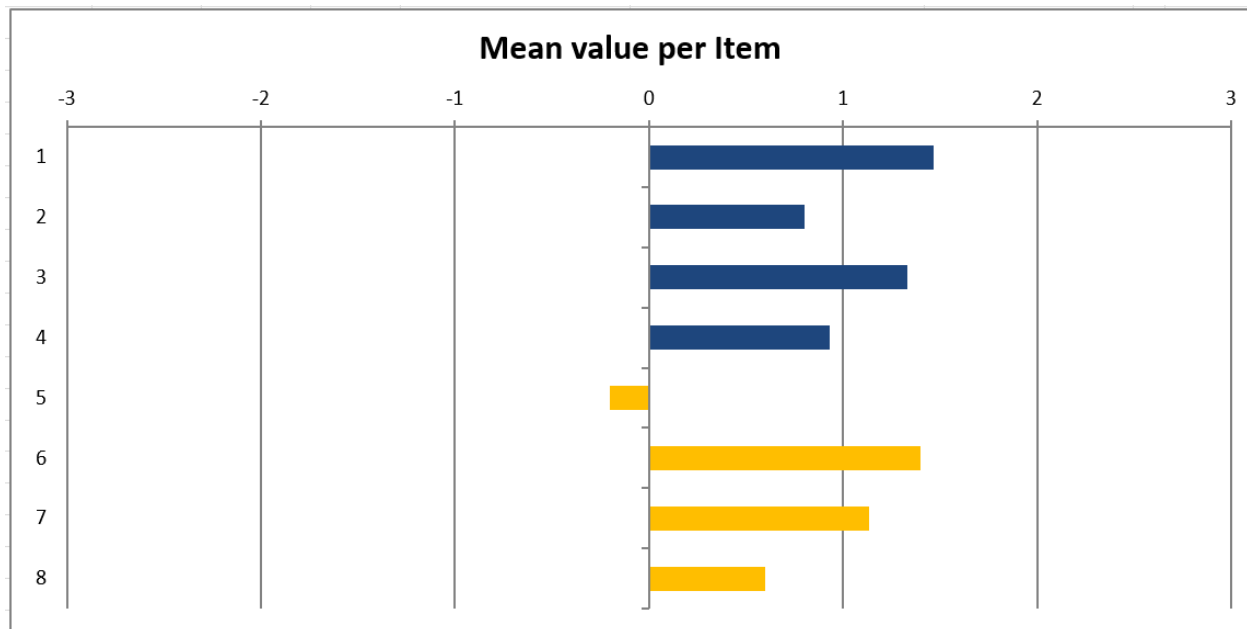


Table 5. REFLECT vs. benchmark

