

---

# “I Tend to Run to Problems That People Run Away From”: Emotion as an Essential Asset in Trust and Safety Work

Toby Shulruff and Amanda Menking

---

**Abstract.** Platform-side Trust and Safety (T&S) is the crucial paid work of responding to and mitigating harmful content and behavior online and beyond. It is characterized by complexity, ambiguity, urgency, and trade-offs based on competing values across a constantly morphing landscape of technologies, abuses, and actors. Interviews with 47 T&S professionals suggest that their expertise is rooted in affective-relational skills: seeking multiple perspectives, reflexivity, curiosity, and collaboration. Furthermore, our findings suggest T&S professionals are motivated by a desire to protect users and spaces, the intellectual challenges inherent in the work, and the caliber of their colleagues. However, the fundamental challenges of the work are compounded by other conditions: both internal and external misperceptions about T&S, responsibility with limited autonomy, and organizational structures. In contributing a more nuanced and grounded perspective of platform-side T&S, we argue that emotion is not a liability, but rather an essential asset in T&S work. We call for valuing the affective skills and motivations T&S professionals bring to their work with the aim of a shift from coping toward well-being and from individually-borne responsibility toward organizational support.

---

## 1 Introduction

Janitors. Judges. Cleaners. Custodians. Security guards. Governors. Reluctant sheriffs. In recent years, these terms and others have been applied to platform-side Trust and Safety (T&S), that is, the crucial paid work of responding to and mitigating harmful content and behavior online and beyond (Feerst 2019; Riesewieck and Block 2018; Gillespie 2019; Wray 2025; Klonick 2017; Cath-Speth 2021). Prior work on this topic has largely focused on the experiences of front-line content moderators who are exposed to harmful content under conditions of precarity and low pay (Gillespie 2019; Roberts 2021). In contrast,

this study draws on interviews with mid-level and management-level T&S professionals, offering insights into their motivations and the unique combination of expertise and strategies they bring to bear in dealing with some of the most complex and difficult consequences of our digitally-saturated world.

Our findings provide a more nuanced perspective on the stressors experienced by T&S professionals. Rather than resting at points of vulnerability, however, we also relay the motivations, affective (emotional) skills, and strategies they have cultivated both to be effective in their work and to sustain well-being. We argue that the emotional capacities T&S professionals draw on in their work are assets, not liabilities: emotion is foundational to T&S expertise.

This paper proceeds with a review of previous scholarship on stressors and coping strategies in T&S work as well as conceptual frameworks on the role of emotion in expert knowledge work. We then present our findings: the stressors of T&S work, and professionals' motivations, skills, and strategies for managing well-being. In the discussion, we place our findings in conversation with previous scholarship and argue that emotion is an asset rather than a liability in platform-side T&S work.

## 2 Definitional Work

Scholars and practitioners have defined “trust and safety” as both “*the study* [emphasis added] of how people abuse the internet to cause real human harm, often using products the way they are designed to work” (Cryst et al. 2021) and “an umbrella term to describe *the teams* [emphasis added] at internet companies and service providers that work to ensure users are protected from harmful and unwanted experiences” (Eissfeldt, Lazarus, and Shah 2021). Other terms, such as “integrity work,” have been tried, but the phrase “trust and safety” seems to be winning out: it was enshrined by an international standard in 2025 (DTSP 2025). Although not yet a convention, we use Trust & Safety as a proper noun to refer to platform-side T&S (“the teams”) and the term written in lower case as a modifier to refer to the topic, field, etc. (“the study”).<sup>1</sup> We use platform-side T&S to refer to employees of a platform or service provider. We do *not* include volunteers or community moderators who have never been professional T&S practitioners in our sample for analysis.

---

1. Additional phrases help too. We use “ecosystem” to describe human and non-human agents and interactions between them (e.g., how a T&S team responds to a new law outlining requirements for the deployment of an AI agent and how changes to the agent impact users). We use “field” to describe the broad areas of knowledge, expertise, and experience that make up the domain (sphere of activity) and to include people who are not employed by tech companies (e.g., community moderators, regulators). We use “profession” and “professionals” to refer specifically to the paid occupation and the people who are employed by tech companies to do T&S work—from Operations to Product to Legal to Public Policy to Engineering to Research. Finally, we use “practice” to describe how paid T&S professionals apply their skills and expertise. Arguably, Trust & Safety does not yet meet the requirements of a profession. It requires no formal education, training, or licenses; however, the boundaries between occupation and profession have blurred, especially with the advent of the Trust and Safety Professional Association (TSPA), dedicated glossaries, and recent investment from universities in related coursework.

We also note the importance of job titles and roles (Shulruff 2024). The T&S Curriculum created by TSPA provides a helpful overview (Bhatlapenumarty, Davis, and Keller 2021); however, T&S professionals often have other different titles, depending on the business aims, audiences, and affordances of a given platform or service. Critically, T&S work is and always has been more than content moderation, and the majority of T&S professionals are *not* content moderators.

### 3 Literature Review and Background

Platform-side T&S work is challenging. Stressors related to subject matter and working conditions can negatively impact the well-being of those who do the work, at every level of seniority and across many types of platforms. Thus, we begin this section with a review of previous scholarship, much of which pertains to challenges, impacts, and coping strategies related to content moderation and moderators. However, as noted above, T&S work is more than content moderation: it includes work across marketplaces; apps for navigation, dating, the sharing economy, and gig work; Internet of Things (IoT) devices including smart home and location trackers; XR; and infrastructure services (Shulruff 2024). We have included three recent sources that take this wider scope. We note that previous scholarship has framed emotion as a vulnerability rather than an asset. Therefore, we introduce concepts from sociological explorations of the role of emotion in expert knowledge work to lay the groundwork for later discussion of how emotion is leveraged by T&S professionals.

#### 3.1 Challenges, Impacts, and Coping Strategies

Previous research, most notably Gillespie (2019) and Roberts (2021), as well as reports and press coverage, has documented the work and working conditions of front-line content moderators inside platforms and in outsourced positions. They found the content itself can be traumatizing, and that this is exacerbated by working conditions including quotas and accuracy requirements, low pay, and a lack of wellness support. Gillespie's and Roberts's findings related to negative content and working conditions have been widely cited in subsequent scholarship (e.g., Steiger et al. (2021), Denyer Willis (2023), Gorwa (2024), and Tyler, Meares, and Katsaros (2025)).

Building on Roberts's work and popular press investigations about trauma impacts experienced by content moderators, Steiger et al. (2021) compared those impacts with similar professions and reviewed a variety of approaches to wellness grouped as programmatic (training sessions, resilience building, therapy) and technological (automation, interface options like blurring or changing size, virtual reality experiences, and matching for peer-to-peer support). They recommended additional best practices including disclosing risks during hiring, limiting exposure, enabling moderators to create space between themselves and content review, and building connections with others.

Similarly, Spence et al. (2023) examined informal and formal support and individual coping strategies of content moderators tasked specifically with countering child sexual abuse material (CSAM) and child sexual exploitation (CSE). They found that talking to others (informal support), creating boundaries between work and home life (individual strategy), and others' recognition of the importance of the work (formal support) were important, consistent with research on similarly stressful occupations. Their findings revealed ambiguity about how moderators perceived the efficacy of individual therapy or technological tools to mitigate traumatic exposure. Interviewees reported that talking to people other than colleagues was challenging due to either lack of understanding or worries about traumatizing content (both for friends and family as informal support and therapists as formal support), and they saw technological mitigations (a widespread approach to reducing harm) as hindering their work.

Gauthier et al. (2025, 12) evaluated international and US-based samples of content moderators and found that organizational context (i.e., working conditions) and individual response styles combined with exposure to indicate risk for negative impacts to mental health. They advocate for a combined intervention strategy including both (1) "exposure management with work design" through better-designed computer interfaces and (2) modifications to organizational culture and the pace of work to promote resilience and mental health.

On the subject of technological mitigations, Mihalache and Szostak (2026) argued for the integration of well-being into user experience (UX) design for content moderators. They noted that content moderators experience psychological and physiological impacts and a lack of autonomy similar to workers in other highly repetitive environments, psychological distress due to content, emotional and cognitive demands due to task complexity, poor working conditions, and limited support. UX design could, they proposed, mitigate the impacts of task repetition, exposure to sensitive content, and the complexity of decision-making by promoting role effectiveness, connection, and resilience.

In contrast to the work cited above, empirical research on T&S work *beyond* content moderation is sparse. Denyer Willis's (2023) mixed methods research included 64 interviews with T&S professionals and observations of a cross-company group, contending that T&S professionals are primarily oriented toward the protection of their companies—a finding which differs from our own (see below). Moran et al. (2025) interviewed 20 T&S professionals from gaming, marketplaces, and AI platforms in the context of the 2022–23 layoffs in T&S. They found a combination of extreme content, professional pressures and precarity, and public misperceptions produced negative impacts to T&S professionals' well-being. They identified specific challenges including misalignments between policy wording and real-life instances, the weight of responsibility, interdependence with other teams, and the sheer speed and scale of the work. Weigl and Bodó (2025) interviewed 19 people in the T&S field who held roles ranging from moderator to executive within platforms beyond only social media (and including some not working at platforms). While

their focus is largely on implications for the public interest, their findings include interviewees' perspectives on organizational structure, under-resourcing, and harms related to content moderation.

### 3.2 Trauma

The extent to which impacts of T&S work are traumatic is a focus of research and a topic of debate in policy and the popular press. However, terminology related to trauma is sometimes used imprecisely and interchangeably. Therefore, to assist the reader, we pause briefly to offer a summary of various types of traumatic response, common symptoms or responses, and coping strategies. Newell and MacNeil (2010) distinguish between *secondary traumatic stress*, which can occur when someone is empathetic with others who have directly experienced trauma and can result in a similar set of symptoms, and *vicarious trauma*, which is a “cognitive change process resulting from chronic direct practice with trauma populations, in which the outcomes are alterations in one’s thoughts and beliefs about the world in key areas such as safety, trust, and control” (60). *Burnout* develops over time and can stem from exposure to trauma, organizational conditions, or other factors, and shows up as some combination of “emotional exhaustion, depersonalization, and reduced sense of personal accomplishment” (59). *Compassion fatigue* is a combination of secondary trauma and burnout (Voth Schrag et al. 2022). Lastly, *post-traumatic stress* often occurs as a temporary, natural reaction to trauma exposure, while the *disorder* (PTSD) describes situations in which the impacts of trauma are longer-lasting.

Lipsky and Burk (2009) lists potential trauma exposure responses (i.e., symptoms) as “feeling helpless and hopeless, a sense that one can never do enough, hypervigilance, diminished creativity, inability to embrace complexity, minimizing, chronic exhaustion, inability to listen/deliberate avoidance, dissociative moments, sense of persecution, guilt, fear, can’t empathize/numbing, anger and cynicism, addictions, and grandiosity, or one’s identity becoming solely about work” (31). Perhaps confusingly, some symptoms can overlap with common coping strategies, for example, when a coping strategy such as emotional detachment becomes numbing, or an ostensibly healthy strategy like exercise is overdone. Additional strategies include behaviors and activities that support long-term well-being such as physical activities, hobbies, supportive social interactions, mindfulness, and maintaining routines or rituals. Organization-level practices can exacerbate or mitigate harm from the work environment and affect individual well-being (Bloom, cited in Voth Schrag et al. (2022)).

### 3.3 The Role of Emotion in Expert Knowledge Work

Scholarship, reports, and press coverage examining the impacts on content moderators and/or T&S roles more broadly tend toward damage narratives (Tuck 2009), that is, positioning those who do T&S work as victims in need of protection. For example, Pinchevski (2023), a philosopher of Communications and Media, explored the implications of the

idea that trauma could result from exposure to audiovisual media.<sup>2</sup> Drawing on press sources to establish the harms of exposure to disturbing content, he asserted that affect and an ability to be traumatized are essential to effective content moderation: “content moderators are paid for their traumatic vulnerability, which is beneficial only as long as it remains subclinical, that is, not interfering with their ability to put it to work” (217). He further argued this traumatic vulnerability is commodified as content moderators engage in paid labor and platforms benefit from that labor.

Without undermining important claims about the toll of past and current T&S work and working conditions, a damage-focused approach is paradoxical, because protecting users and online spaces is both a main goal of T&S work and a key motivation for T&S professionals—as we share in our findings. Framing emotion as a vulnerability or liability rather than as a resource or asset overlooks the complex interplay of emotion and cognition we heard about from T&S professionals in this study. Sociological research into the role of affect in expert knowledge work proved useful in making sense of our findings, and so we summarize two key conceptual frameworks briefly before presenting our findings.

Positioning emotion in opposition to rationality is an unhelpful inheritance from the European enlightenment, a social attitude embedded in the rhetoric of science and technology (Wajcman 1991). A richer understanding of the interrelationship between rationality and emotion has emerged. For example, Craciun (2018) argues that emotion-based, “affective-relational skills” can be put to work to diagnose, think about, and solve problems, and are therefore instrumental in knowledge work. She proposes a framework of three kinds of uses of emotion in expert knowledge work: *supportive* (as an internal motivator), *didactic*<sup>3</sup> (as a tool for collaboration and communication), and *inductive* (facilitating problem-solving).

Wang (2022) extends Craciun’s framework to propose how a balance of emotion and cognition yields a “desired state of mind,” or “controlled empathy.” Building on sociologist Arlie Hochschild’s groundbreaking articulation of emotional labor as the work of performing a certain kind of emotional response for the benefit of others (or to evoke specific emotional responses in others) even though it might not reflect one’s inner emotional state, Wang notes that achieving this balance is a skilled practice. In addition to internal mediation, she also argues that this emotional-cognitive balancing work “translates complex, messy and ambiguous real-life situations into stable, clear-cut and pre-defined categories in the bureaucratic system” (2). Moreover, as noted above, emotional labor is integral to invoking this state of mind in others. Therefore, emotion is intrinsically valuable in expert knowledge work; it maintains the legitimacy of systems and subsequently produces profit. Structurally, though, Wang says that this kind of work is frequently low

---

2. That trauma can result from electronic media is more than just an “idea.” It is recognized by the Diagnostic and Statistical Manual, 5th edition, as noted by Gauthier et al. (2025).

3. Note that Craciun’s use of “didactic” is nuanced, and differs from a stricter definition of the term in common usage.

status, lacks recognition, and provides few outlets for stress, especially along lines of gender, race, and class.

In the Discussion section below, we apply Craciun's and Wang's conceptual frameworks to our findings. We assert that the resulting perspective allows us to explore the role of emotion in T&S work not as a liability but as an asset. This, in turn, could aid in the development of individual and organizational practices which go beyond basic trauma mitigation or coping strategies toward sustaining T&S professionals' well-being in the long term.

## 4 Methods & Data

This paper is based on analysis of transcripts from interviews with 47 current or former paid T&S professionals. Semi-structured interviews were conducted between June 2023 and July 2024 as part of a broader history of Trust and Safety project; see Appendix A for the full interview protocol. Recruitment included a combination of invitations at in-person events and snowball sampling. Because the study is run by a nonprofit organization working without an academic partner, institutional review board (IRB) approval was not sought. However, we followed rigorous guidelines for ethical research, adapted for T&S and outlined in our Code of Conduct; participation was voluntary and procedures for informed consent were followed; transcripts and a draft of this paper were shared with participants; and we took strenuous measures to ensure data protection. All names used in this paper are pseudonyms.<sup>4</sup> Importantly, interviewees were asked about their professional experiences as T&S practitioners; they were not asked to speak on behalf of their current employer or about only their current role or experiences. We did not ask participants to compromise existing conditions of their employment, nor did we ask questions that were intended to violate legal contracts (e.g., nondisclosure agreements).

Nearly two-thirds of the interviewees were women. Almost three quarters had ten or more years of experience in the field. About a quarter of interviewees had countries of origin outside the US and Europe. Interviewees came from a wide variety of backgrounds, as is common in the field, and found their way into their roles by accident in the early years of the field and, recently, in more purposeful ways. Our interview protocol did not explicitly ask about educational background, but the 27 T&S professionals who volunteered this information had studied in a wide range of disciplines: Political Science, Social Work or Psychology, Computer Science, Law, the Humanities, or Media or Communications (see Table 1).

While we asked generally about how people came to T&S, we did not explicitly request that participants share previous sectors of work. Nevertheless, five shared they had

---

4. Additional information about procedures and protections is available on request.

Table 1: Interviewee Demographics

Question	Response	N	%
Gender	Women	30	64
	Men	16	34
	Nonbinary	1	2
Years of experience	16+	12	26
	10–15 years	22	47
	6–9 years	10	21
	3–5 years	3	6
Country of residence	North America	35	74
	Europe	5	11
	Outside US/Europe	6	13
	Unknown	1	2
Country of origin	North America	28	59.5
	Europe	5	11
	Outside US/Europe	13	27.5
	Unknown	1	2
Areas of educational background <sup>a</sup>	Political Science	8	17
	Social Work or Psychology	6	13
	Computer Science	5	11
	Law	4	9
	Humanities	2	4
	Media or Communications	2	4
Previous professional experience – Sectors <sup>a</sup>	Civil society	5	11
	Law enforcement or national security	2	4
	Academia	1	2
	Government	1	2

<sup>a</sup>Not all interviewees provided this information; percentages based on full sample.

previously worked in civil society, two in law enforcement or national security, one in academia, and one in government. A few participants shared backgrounds in other tech work or that they “grew up online,” and some served as volunteer managers or moderators of online communities. Interviewees had held a wide variety of T&S roles. More than a third had been at some point in content moderation or customer service roles. Most were in or had previously held director or manager-level positions (see Table 2). Of the 47 participants, all but two were still employed in the tech sector and working in either T&S roles or T&S-adjacent roles at the time of their interviews.

Taking a grounded theory approach, we situated the voices of participants as the starting point and ground truth of our analysis. The first author read through the interview

Table 2: Interviewee Roles

Role	Current Role	Former Role
Director	18	5
Manager	12	5
Policy	13	7
Product	6	3
Legal		2
Investigations	4	2
Operations	3	11
Engineer	1	
Consultant	7	1
BPO/Vendor	3	
Customer Support		7
Content Moderator		13

*Note: Totals greater than 47 as some have multiple role categories*

transcripts and then assigned *in vivo* codes developed in collaboration with the second author. She then grouped coded excerpts for re-analysis, weighted based on frequency, and placed them into relation with each other through visual mapping in Miro. The resulting themes form the basis of the findings: challenges of the problem (i.e., issues T&S addresses), challenges of conditions of doing the work, resources they bring to the work, motivations, and well-being strategies (see Table 3). Following this analysis process, we then compared our findings with other recent scholarship on T&S work and the aforementioned conceptual frameworks on the role of emotion in work. Our claims were developed in conversation with these sources.

#### 4.1 Positionality

Toby Shulruff is a PhD candidate in science and technology studies (STS) and mixed methods researcher with over 25 years of experience working at the intersection of technology and gender-based violence. Her previous research has focused on T&S work, expertise, and well-being strategies. Her methodology is informed by feminist research practices rooted in robust forms of consent, participation in shared meaning making, attention to power, and applicability to emergent issues. Amanda Menking is a qualitative researcher with a PhD in Information Science who, during her time in academia, authored papers about gender and emotional labor in relation to knowledge production and online communities. Her methodology is informed by feminist human-computer interaction (HCI) and critical STS. She now works for a nonprofit organization in the T&S ecosystem.

Table 3: Analysis Themes

Themes	Subthemes
Challenges of the problem (i.e., issues T&S addresses)	Crisis driven, fast pace Complexity, ambiguity Trade-offs, competing values Geographies Experience Dilemmas
Challenging conditions of the work	Available resources Pressure, personal safety Visibility, others' understanding Geographies
Resources they bring	Seeing both sides Reflexivity Collaboration via relationship and "translation" Backgrounds (theirs and colleagues)
Motivations	Purpose: passion, protection, impactful Variety, novelty Colleagues
Well-being strategies	Reminders, routines Boundaries, purpose Work-life balance, self-care Celebrate, humor Team, field Leave, burn out

## 5 Findings

Our presentation of findings follows key themes that emerged from our analysis: inherent challenges, expertise, motivations, working conditions, and strategies. The rationale for this order of presentation is: T&S professionals meet the inherent challenges of their work through a combination of expertise and motivation; however, adverse working conditions often undermine their strengths. Their strategies for well-being develop in relationship with all of these factors.

### 5.1 The Inherent Challenges of T&S Work

To begin to understand T&S work and the people who engage in it, it is essential to make an account of the configuration of problems, risks, and harms they grapple with every day. Interviewees spoke of subject matter, scale, pace, complexity, and ambiguous contexts as challenges inherent in T&S work itself. As noted above, the scope of this study extends

beyond content moderation on social media platforms. Therefore, we use the term *subject matter* rather than *content* to better include the range of behavior (also sometimes referred to as *conduct*), incidents, and threat models that T&S professionals encounter beyond discrete pieces of content. In response to repeated exposure, participants described experiences consistent with the symptoms of vicarious trauma, stress, and burnout, including impatience, hypervigilance, and feelings of despair or isolation. As Enise said, “So, definitely day-to-day exposure to explicit images and things like that wasn’t fun anymore, and I needed another outlet or other opportunity to move on.”

Importantly, however, one cannot assume that impacts from exposure to subject matter are universally negative or severe. Participants believed different kinds of subject matter affected people differently, saying self-knowledge or good management practices could match people to work that didn’t cause negative reactions or which “didn’t bother” them, as Ella related:

*One of the areas that is I think very understandably the hardest for most people to deal with is child sexual abuse. It is among the most egregious content that anyone can produce and that anyone can be exposed to. For better or for worse, I’m not as deeply affected by that content as a lot of my peers. You find there are loads of other folks who have that as an experience in the field.*

Rather than making her more vulnerable to negative impacts, Corinne spoke about how her lived experience helped her in her work:

*When I was starting out, there was concern from Trust & Safety management or company management being like, “I think the content moderation might be too triggering, for people with mental illness or these experiences, this job might be too triggering for them.” And I really challenged them to think otherwise. I’m like, no, for me, my lived experiences help me be able to take in the graphic content because I can deeply relate.*

Over the decades of the field’s evolution the magnitude of challenges has increased, gaining momentum and growing in scale and speed across geographies. Jatin summed up this challenge:

*So, the speed at which you are detecting the content should match the speed at which the content is getting uploaded. So, I see that intersection as very, very important because billions of data are getting uploaded, but are we figuring that content at the same speed?*

Compounding the challenge of scale and pace, T&S professionals handle extremely complex issues, resulting from continually morphing behaviors at the interaction of a vast number of people and a wide variety of technologies. Complexity occurs on multiple axes: in individual cases or decision points; the ways those individual cases are part of larger patterns or trends; the convergence of technological and social factors; the

influence of online and offline activities on each other; and varying impacts in different contexts.

To work at scale on issues of such complexity, T&S teams are constantly seeking more information and nuanced perspectives and adapting solutions. In practice, this means that response to the volume of incidents and cases evolved to include larger teams, outsourcing, and automation. Effective processes at scale required honing rules or policies that were broad enough to work in many cases, but also specific enough to be applied by a larger workforce or automated routines. Additionally, Diaa spoke to the need for localization and cultural context, which led companies to bring in “people with regional expertise that can understand or flag something that will be impacting a specific region or a specific market.” As companies based in the United States opened regional offices, regional language expertise was often housed there. Amanda described the feeling in one regional office: “It felt like you were in the UN because there were more nationalities in that building than you would find in an international congress somewhere.” The challenges that came with working in those regional offices will be revisited later.

The problems are always evolving too. Across contexts and geographies, what’s acceptable changes, meaning that there is not one solution. As Sam warned:

*Those things [problems and their solutions] are never done because the world changes, and the company changes, and the way people are misusing the product changes, and the technology changes. So, they’re never done. Also, you’re not that smart, so you certainly could not foresee everything, which meant new stuff was coming up. You can write the set of rules, but the project can never be finished.*

At scale, rules and models for automation also need to be updated constantly. For example, Kirk said, “As vocabulary shifts, as new slurs are introduced, as new kinds of issues come up, we’re paying attention to that and making sure that our models are handling those well.”

Any solutions or approaches taken by T&S teams can have unforeseen consequences and ripple effects regardless of intention, especially when taken under the urgency of crisis. Kirk summarized the thinking process:

*Here are some bad things that we don’t want. Here are some good things that we do want. What’s the trade-off between them? And in some cases even, is this thing going to result in these good things?*

An additional complicating factor unfolding across the decades, and of particular interest now, is the patchwork of regulations across jurisdictions globally that reveals some of the tensions between US-derived freedom of expression and local contexts and laws. Jenn described these challenges this way:

*And so really thinking about how do we bring in a non-Western view and make sure that we're appropriately representing those communities and not over or under enforcing ... particularly when it comes to people who are trying to reclaim slurs that had been previously called out for hate speech, people who were sharing screenshots of the hate speech that they were receiving.*

This tension is not theoretical for T&S professionals, but rather case-by-case and with real-life impacts. Chloe related:

*It's like that realization of, "I'm not going to change the world with policy." You're not going to allow women to drive in Saudi Arabia by keeping this up. You might get that woman killed. So, when she writes in to ask it to be removed, you figure out how to do that so she's protected.*

Not only are these problems high-stakes, complex, and occurring on increasingly vast scales, but the conditions of the work itself compound stressors, which we will elaborate on below. First, we share findings about how T&S professionals meet the inherent challenges of the work with a combination of expertise and motivations.

## 5.2 Expertise

T&S professionals bring a diverse set of skills, knowledge, and backgrounds to their work. Rose ascribed her "intrinsic nature" to an alignment with T&S work: "I'm just a natural protector. I've always been like the person on the schoolyard defending my friends." Speaking about their own and others' various backgrounds, participants noted that variety yields a range of perspectives on the problems T&S addresses, including both specific subject matter expertise as well as the ability to ideate potential implications for a wide set of scenarios.

A strong subtheme that emerged was T&S professionals' ability to "see both sides." For example, Heather said:

*I think some of the skills that you need to be a good journalist, where you have to be able to hold a lot of conflicting ideas in your head and different points of view in your head and say, "Alright, we're just going to have to go from there and acknowledge these different points of view." That's a skill I have to use every day.*

T&S work requires actively listening to people who disagree passionately. To make determinations, T&S professionals seek out more perspectives. Dhvani described the approach as seeking "as much context as possible":

*Where it is possible to speak to someone to get that context, that's something we would try to do. ... And if we do do that and there are two sides to whatever the issue at question might be [we] try and speak to people on both sides of it*

*just to make sure that we're getting a balanced perspective there.*

“Seeing both sides” is more than gathering more information; it also entails affective skills. Participants described the profound importance of reflexivity and awareness of their own biases, while also exercising empathy balanced with skepticism. Harper emphasized that you don’t have to agree with someone’s point of view to listen to them, saying “it truly has changed my perspective on how I meet people on the other side and have discussions with them.” Maisie spoke of an early manager who “gave [her] permission to also humanize people, even the people who were predators,” saying: “Like, I generally separate people’s behavior from who they are.”

This places T&S professionals in the middle, performing a balancing act between both (or many) sides, and taking a case-by-case approach. Across a wide range of geographies, they often have to balance competing needs and claims in which there is room for interpretation and ambiguity, all while under the pressures of scale and pace. Eleanor spoke of the importance of being:

*... a bit more thoughtful and a bit more measured versus panicking ... I've seen people panic and rush to a response and also rush to an answer to just handle it quickly. And then it's like, "Oh, now we have to eat our words and change the decision."*

These skills extend beyond content moderation and investigation to other areas of T&S work, including writing policy, analyzing root causes and risks, testing new products for flaws, creating user journeys, supporting victims, and collaborating with internal and external stakeholders. T&S professionals often collaborate with other teams, including Product, Engineering, Legal, Sales, Data, Enforcement, and with directors; between companies when addressing shared threats like CSAM and disinformation; and with partners in other sectors. T&S professionals’ collaboration skills build on their own ability to see both sides of an issue to provide necessary nuance and complexity by “translating” these multiple competing perspectives and concerns to other professionals inside and outside of companies. Noah shared that he seeks diplomatic skills during hiring to help navigate these functions. In other words, the very skills they need to do day-to-day work makes them well-suited to collaborative efforts.

### **5.3 Motivations**

Interviewees gave multiple reasons for doing the work and staying in the profession: protecting others, intellectual challenge, colleagues, impact, passion, recognition, and the excitement of the work.

Far and away, the most common motivation was a desire to make online and offline worlds safer and to protect users. Isaac summed it up when he said:

*I think that's something that's very central to the idea of trust and safety is it's*

*less about, “Hey, let’s make a specific product look better or shinier or nicer in the space.” It’s about, “How do you make the world a better place for the communities that you are working in or a better place, how do you make people safer?”*

Claire felt a personal connection, saying “I’m really drawn to the mission of protecting people, particularly vulnerable populations. ... I have a very strong sense of justice and fairness, and this very much ties into that.” Heather extrapolated her own motivation to one that holds generally for the field:

*What everyone has in common is there’s an intention to protect the end user, right? And it might be [to] protect the end user from other people, protect them from themselves, protect them from the company, protect them from the government. But, ultimately that’s why we’re trying to govern these behaviors, is because we’re trying to build our own tiny utopia.*

Another common motivation was intellectual challenge; T&S work involves dealing with complex problems with no easy solutions. Laura was not alone in finding this satisfying and enjoyable: “There’s an intellectual, fun problem-solving. ‘Wow, this is hard!’ How do we think about this? This is novel, or this is new context that is really exciting.” Rose told us, “I just really love these problems, they’re basically unsolvable.”

Many participants said that they stayed in T&S because of their colleagues. Josh shared:

*It’s probably some of the smartest, most talented people I’ve ever met have been in this space. Because the ability to work in all of that ambiguity and bring together a product and policy and operations and tooling and regulation and bring that all together in a concise way is, it’s no small feat, right? ... And most people are very curious about the world, about themselves and about people’s behavior and the psychology of why people do what they do. I think that’s something that I encountered quite universally as well.*

Similarly frequent was the sense of being able to make an impact directly and at scale, and the power of seeing results; for example, Jagan said, “You get to work on some stuff that touches hundreds of millions of people or billion people or whatever it is, and that’s amazing too.” Similarly, Tanvi reflected:

*It’s very rewarding when you put in the hard work and you see it materialize into something good, you see harm being reduced, you see protecting people from abuse and you see making a change and it happens not very fast all the time, but many times you get to see that. Even a fraction of it when it happens, it’s quite rewarding.*

Participants also spoke about passion and selflessness, as Layla said:

*The passion that I have seen ... I've never met anybody in any of the Trust & Safety teams I've been a part of, who was there to censor for the wrong motives, to oppress community. It's the opposite. It is like you want to help and to help, you're exposed to horrible pieces of content so that your family, your loved ones won't see it. So there's this selfless mentality that people bring to their role.*

Though less often discussed, interviewees also mentioned compensation and recognition. Even so, the smaller number of participants who did allude to these factors as motivations had conflicting opinions. While some felt that the pay, or at least the excitement, of working in tech was alluring, others were clear that they were not motivated by these factors. Jenn said:

*I think people naturally are like, "Oh, Trust & Safety. You guys went to a social media company because they gave you hoodies and Allbirds and free good snacks" and whatever else. And I work out of my home, my snacks or whatever I buy myself. I buy my own shoes. Thank you very much. But it is not the perks that draw us. The perks are just perks.*

Finally, some participants spoke about the rush or exhilaration of T&S work, as Amanda reflected:

*I have to say, any other role that I've had after that one always felt a little bit boring in comparison. It was fulfilling in different ways, but the pace at which you learned was not the same, the pace at which you made decisions was also not the same. ... And when you get used to that, it's a bit of being like an adrenaline junkie.*

As the phrasing of "adrenaline junkie" points to, exhilaration as motivation can have downsides as well. That ambivalence, or more specifically the intertwined nature of motivations and challenges, is something we explore more in the Discussion section. For now, it is enough to observe that what can be perceived as downsides from the outside are often, in fact, motivations for many T&S professionals.

#### **5.4 Challenging Conditions**

We shared findings above about the inherent challenges of T&S work. Now we focus on working conditions that compound stress, including visibility, responsibility, and organizational structures.

##### **Visibility**

Platform-side T&S work is not well understood outside of the field—either by other teams within companies or by the public. Interviewees shared how their own family and friends, the press, and policymakers either were unaware of their work or profoundly

misunderstood it. However, participants reported this circumstance is shifting over time, largely due to press coverage. They said that with greater awareness of T&S has come a better understanding of the impacts of the work. Specifically, friends' and family's understanding made it easier to talk about what they do. For example, Laura said:

*I don't think my parents knew what I did for the first few years. I would explain it to them or to friends or to former colleagues, but it didn't really make loads of sense to people or they'd never thought about it before. ... And I remember going home for something and ... my mom holding up the paper and going, "Oh, this is your team, right? This is what you guys do." And I think it was the first time it really clicked for her.*

The downside of greater visibility, though, is that it often leads to criticism. Eleanor said, "With more awareness can come unhelpful reactions, like 'What's the worst you've ever seen?'" Carolyn said she found people wanted to talk about trust and safety issues, and would ask for help with specific problems. Consequently, some chose to let their work remain "behind-the-scenes" because the subject matter is "taboo" or "shameful," they may not be at liberty to share details of their work, or because it elicits prurient interest. At an extreme, interviewees also talked about the risk of receiving threats, doxxing, and harassment. In response, Harper told us, "A majority of my team, including myself, use aliases in order to do our work."

Importantly, T&S is not necessarily better understood or appreciated within companies either. Participants frequently talked about being seen as a cost center, or worse, naysayers or blockers for new products or initiatives. Tanvi shared:

*In the beginning, only certain functions in organizations would understand. Your key stakeholders would understand what you're doing. If you had to get a buy-in from a team that you do not work with, you really need to explain why exactly.*

This creates a dilemma for professionals who are often driven by a desire for productive conversations, collaboration, more perspectives, and greater context. The lack of understanding has a cost for those who do T&S work. Amanda pointed to the quick removal of a livestreamed mass shooting, arguing that the team "should be applauded, not criticized." Further, interviewees reported that not only is the work of T&S less legible, but especially externally, the limits of their agency and the location of those teams within organizational structures is poorly understood.

### ***The Weight of Responsibility***

Multiple participants spoke about how the work never ended and about the difficulty or impossibility of stepping away from work. Tanvi summed it up: "You don't unplug from Trust & Safety." Over the decades of the field's evolution, the magnitude of challenges has increased, gaining momentum and growing in scale and speed across geographies.

Participants articulated how their desire to “do right” came with a weight of responsibility and frustrations at the limits to their autonomy when attempting to address complex and multivalent problems.

Perhaps more than most, T&S professionals witness the tangible ways technological affordances and platform policies amplify ideas and behaviors. While the enthusiasm of product designers and engineers may overlook the negative knock-on effects—as Abby said, “when a system is getting started, people are thinking about how great their product is, not how someone is going to destroy it”—interviewees were keenly aware of the real-world impacts of online risks and harms, referencing mass shootings, conflict zones, suicide and self-harm, and stalking and harassment. As Nick said, “Policy decisions can destroy a community.”

Interviewees also reported that there are no “right answers” on a daily basis because T&S professionals are constantly encountering ethical conundrums and conflicting priorities. For example, Corinne described her experience:

*As an investigator, I really have to make judgment calls on so many areas. First being the priority level of what is the biggest danger to the public or to the customer reporting right now, how fast do I need to respond or act on this situation? ... If I get six a day and based on this priority level, who can I help right now? Which tools and resources do I have or do I need to take more time to look [into] something further to identify the abuse or problem I'm trying to resolve? So I'm mitigating as much harm both to the person experiencing such alleged abuse, but also to the people potentially propagating the abuse, making sure that whatever policies they may be violating, they're also being applied equally and fairly, not just to protect the liability of my employer, but also to promote justice and fairness to the individuals even engaging in set of alleged abuse. So that's just two of five, six different judgment calls I have to make.*

Many of these “judgment calls” are not clear or simple, as Lucas said: “All the decisions, ... especially at a high level, are like catch-22 decisions.” For example, the same decision may impact various individuals or communities differently, resulting in negative outcomes for one group, or in unforeseen ways. However, even under these conditions, Sam argued there was a “right way to be wrong,” elaborating: “In this circumstance, we should be doing the thing where we least regret errors because we’re going to make a shit ton of errors.” He went on to give an example in which preventing physical harm to one party would be prioritized over inconvenience to another. The personal emotional toll of making these calls can be high. As Laura said, “The job can be really stressful and difficult and often involves telling people ‘No’ a lot, or having everyone mad at you.”

### **Organizational Structures**

Beyond the day-to-day conditions of the work, interviewees described different kinds of

organizational structures and dynamics that are stressors for T&S professionals. Within younger companies with smaller teams, people wear many hats. As products and services scale, and companies mature, T&S roles specialize. Mandates for greater consistency in responses, adjudications, and workflows involve writing and re-writing rules or policies, delegating or automating routine actions, and collaborating with other teams. Josh gave an example from the mid-2010s:

*When the ISIS execution videos became really prominent, this is around 2014, '15, maybe, you'll remember the orange jumpsuits and the high production values and the terrible murders that took place all on film. And they were then posted to all these channels and those things spread like wildfire because what would happen was you take one down and then someone else would download it and re-upload it. And we realized very quickly that this whack-a-mole of taking down video by video by video was just never going to work. So we went to the engineering team, we were like, "You have to build us a tool ... Here's this video or this image, you go find every other image in video that matches this so that I can take them down in bulk." That was my first "Aha!" moment of like, we absolutely need scale. We need tooling. There's no way humans ... I could have had a thousand humans and we would never have been able to catch them all.*

Participants described the growing role of automation (i.e., AI or tooling), vendors, and regulatory compliance within an evolving trust and safety ecosystem. Consequently, interviewees said T&S teams can lose autonomy and find that creative solutions are constrained. Erin shared, "I think that is, in general, going to formalize and professionalize a little bit. I think [it] is going to have downsides because I think it's going to reduce the flexibility of that space." Participants also relayed stories of tensions or simply a lack of attention because safety issues were seen as lower priority. With rises in outsourcing, automation, and the pressure of political and economic factors, they said T&S teams grapple with inadequate resources for staffing and tooling, as well as layoffs and job insecurity. Carolyn described the multiple impacts:

*It's also a really sad state of where the industry is right now because of the layoffs. I know that every department is getting touched by these layoffs, but Trust & Safety is usually really gutted. Not only is that hurting the users, but it's really hurting the morale within the teams.*

Interviewees reported that these organizational struggles were even more pronounced across geographies, as companies opened offices outside of Silicon Valley. These locations increased the varieties of languages, contextual knowledge, and expertise that could be brought to bear on trust and safety issues, and enabled round-the-clock coverage in urgent and emerging situations. Nevertheless, participants who worked in or with these offices noted important challenges. At a basic level, time zones made it difficult to connect with staff in other locations, a problem that was exacerbated because

leadership positions often remained in the US, creating a barrier to access to decision-making conversations and career advancement for those outside the US, as Seema said:

*There's not a lot of senior leadership in India or any of the APAC locations. So that is one that all of the decision-making was happening elsewhere. And because of the time zone, first of all, you're already at a disadvantage because ... even if you work late hours, you just get three or four hours of overlap with the US. And access to leaders is anyway a little difficult and then you just have a few hours. ... And so I think the burden is on individuals in these regions to almost work twice as hard to gain access to the same career opportunities that trust and safety folks in the US and Europe may have.*

Further, she said, "It's always the marginalized groups who are also carrying the burden of advocacy and explaining why you should pay attention."

Across locations, interviewees said they face both internal and external pressures—sometimes politicized—to prioritize among urgent or severe harms, to make decisions fairly and consistently, and to balance company brand and cost efficiencies with safety and risk management. Isaiah noted:

*There was some idealism for me that we were on the vanguard, and we're innovating and really pushing society forward, when, it turns out, we were doing all those things as long as it benefited the business. And if the business didn't feel like it mattered, they would move on or sell it to the highest bidder.*

Interviewees were often enthusiastic and energized by the challenges they faced daily. Even so, they also disclosed feelings of exhaustion, futility, and even despair with phrases such as, "nothing ever changes," "spinning our wheels," and "my heart sinks." In the next section, we share the strategies interviewees described using to navigate challenges and their bleaker moments.

## 5.5 Strategies for Well-Being

*There are days where I'm like, "I just want a normal job. I just want a 9–5, pensionable job that's going to get me through the day. I don't have to think too hard about [it]." Honestly, Wednesday of this week, I was like, "What the fuck am I doing?" Excuse my language. I was just so frustrated. I was so disillusioned. But then Thursday comes and Friday comes and you get over it. I've found I've had to build personal resilience, and not to do with the content. That was originally. I don't, thankfully, in my job today, I don't get to see much bad content. — Josh*

T&S work is incredibly challenging, and at times interviewees reported finding it disheartening. Consequently, people who choose this work draw on a wide range of strategies to

cope—and ideally thrive—in the long-term. Below we share findings about how interviewees keep going.

### ***Drawing on Motivations***

Above, we presented interviewee’s motivations for doing T&S work, which include purpose (making the world safer and protecting people), intellectual challenge, the high caliber of colleagues, the scale of the impact they can have, as well as passion, selflessness, and exhilaration.<sup>5</sup> These motivations not only bring people to platform-side T&S, they also sustain people day-in and day-out and allow them to persist in the work. Speaking about moving from a more frontline role to a higher-level policy role, Alex reflected that her impact increased:

*It’s been an incredible experience, and it’s made me love Trust & Safety even more, because I’m no longer looking at one particular piece of violent content or one particular account. I am looking at hundreds and I’m saying, this is what we need to do. ... And so, who is supposed to do it, if not the social media companies? And I want to be there. I want to be part of doing that sort of education, that sort of protection.*

Like several others, Zara said that remembering purpose helped to sustain her:

*It’s just passion for these issues and that’s what really keeps you going in this space, because it’s so easy to burn out when you’re looking at terrible content all day every day with minimal support from whatever company you’re working at.*

Maisie also found purpose to be important: “If what it takes to remove these children from danger is me looking at these images and stopping this abuse from happening, I can do it.” Still, she clarified, “It’s not heroic to be able to look at child abuse.”

While motivations drive T&S professionals to continue working in the space, participants also spoke about the practical, day-to-day strategies they implement.

### ***Routines and Practical Strategies***

Participants spoke about the importance of work-life balance, self-care, resilience, authenticity, and having a positive outlook or focusing on the big picture. Added to these broader and more aspirational states of being were more tangible, specific practices. Routines at the beginning and end of the day offer “a moment of affirmation and happiness,” as Alex put it. They said they “use these little things to set myself up for success.” For example, commuting to the office offered opportunities to appreciate or find joy in the world around them.

Many of the routines and rituals described by T&S professionals in the interviews

---

5. As a reminder, there was dissensus about whether recognition and compensation were motivations.

negotiated boundaries in space and time between work and personal life. In addition to Alex's commute, participants also spoke about the importance of physical space in routines. Lara described taking time in between meetings to walk out to the backyard and visit her garden. Many participants also described a practice of closing out their day by preparing for the next day: reviewing emails, catching up on the news, or otherwise getting a sense in advance of what might be coming up at work. While some participants valued physical or temporal boundaries between office and home, others felt that being able to work from home and at times of their choosing was important.

In addition to negotiating spatial and temporal boundaries, participants spoke about how they negotiate boundaries between the thinking and feeling aspects of work and other aspects of their lives: as some called it, compartmentalization. Eleanor said, "I think you have to be able to compartmentalize in order to be able to do this job effectively." Not everyone embraced that term. Dario called it "fortitude" in reference to doing child safety work. Joe shifted emphasis from "compartmentalizing" to giving the following advice:

*I would try to help you think about how to compartmentalize, but I think the most important thing to realize is that, at the end of the day, it's a job. Once you are out the door, do things that you like. Do things that are important to you.*

For some, the ultimate boundary to be negotiated was the work itself and for whom they do it. For example, Jagan was blunt: "I'm here as a mercenary. You're paying me for my time to do a task. The emotional parts are not what I'm here for." However, others found the emotional aspects are part of their motivation. As noted above, Corinne argued that an affective connection via lived experience is not a liability, but rather an asset.

As interviewees shared, colleagues are a key motivator for T&S professionals, and indeed, turning to other team members emerged as an important strategy. For example, Harper described her team's practices during a particularly stressful series of weeks at the beginning of the Israel-Gaza conflict:

*And so one of the agreements has been having a standup meeting every single day for three weeks ... just so we can come together and just have conversations about really difficult customer conversations that we were having and how to navigate them and support each other and ask for help. Because sometimes, people don't want to ask for help, and this was giving everyone permission. It's hard work. ... We want to be realistic here. Let's support each other as much as we can.*

Beyond their immediate co-workers, interviewees noted they now have a larger team available: other professionals in the emerging field of T&S. For example, interviewees talked about the Trust & Safety Professional Association (TSPA) and the collaborative volunteer curriculum project it facilitates, book clubs, and the sharing of "best" practices

as ways that they benefit from the coalescence of the field.

Importantly, humor also showed up throughout the interviews in the way participants talked about their work and their well-being, individually and within teams. Lara had initially described her daily routine in this way: “First, I wake up, and then, I panic. I let existential dread overcome me in the morning.” She paused. “No, that’s not true,” she continued, laughing.

### **Companies’ Wellness Resources**

Interviewees did not speak at length about company wellness resources, though some did mention time off, training, and therapy. Other described hiring practices including content warnings, and options after hiring to shift subject matter types or the focus of work. While the overall trend toward companies acknowledging the challenges of the work and providing resources was noted, interviewees were ambivalent about the efficacy of these measures. For example, not everyone feels that they can make use of resources due to their workloads. Cassie said:

*But I think one thing that’s challenging is that even when you have all these resources, sometimes you’ve got to do what you got to do. So I might want to completely check out. I don’t want to look at this anymore, but if I don’t look at it, a lot of other people will see it. So sometimes we are in that position where we don’t have a choice. I mean, we have a choice. No one is stopping me from taking sick leave based on my mental well-being. But at the back of my mind, I’ll be like, if I don’t do it, it’s going to impact so many more people. So I feel like trust and safety employees continue to make that trade-off of their personal well-being or for the greater good.*

More than company resources, T&S professionals draw on their core motivations for longevity in the work; enact specific practices including routines and negotiating boundaries (physical, temporal, and cognitive); and draw on the strengths of teams and the field at large to maintain well-being. Importantly, as with challenges and motivations, these strategies and practices are not simple solutions with clear benefits.

## **6 Discussion**

*You’re like, “This is a shitty job.” I’m like, “This is the coolest job you could have. What are you talking about?” So, there’s maybe some, just constitutional differences there. — Ella*

Platform-side Trust & Safety work is not, in fact, easy. At any given moment, T&S teams are working to counter harmful content and behavior on a vast scale, involving a high level of complexity and ambiguous context. They come from a wide variety of backgrounds,

and the skills they have honed and their motivations for doing this work are deeply intertwined with the nature of the problems they grapple with. Yet the challenges they face are compounded by the conditions of the work and their positioning at the pinch point of systems and various groups of actors.

As with previous scholarship, our findings relate the negative, stressful, and sometimes traumatic impacts of aspects of this work. Whereas much of the focus of that literature has been on content moderation at the lowest level of pay, the highest level of precarity, and conditions of invisibility, our interviewees held mid-level and management-level T&S positions, and had relatively more choice in their work. Our findings draw on their experiences, reflections, motivations, and strategies to elucidate the complex role of emotion in their work. This perspective can inform a shift from merely identifying coping strategies to promoting well-being.

Below, we discuss our findings in relation to previous literature about stress, trauma, and coping. Then we consider these findings in light of concepts about the essential role of emotions in work. We argue for a greater appreciation of the role of emotion in T&S work as the basis of their affective skills and expertise. Further, strategies to promote well-being should be built with the value of emotion in mind, and the experiences of T&S professionals at the center.

### **6.1 Traumatic Stress, Coping, and Organizational Approaches**

Content moderation—and T&S work more broadly—involves immersing oneself in some of the worst of human expression and behavior every day. Previous scholarship, reports, and investigative journalism have described the challenges content moderators face from both content and working conditions, along with strategies utilized to mitigate impacts. Interviewee experiences align with this literature. While we did not directly ask about trauma responses (i.e., common symptoms of traumatic stress), these were evident in organic, voluntary responses. Interviewees reported hypervigilance, physical stress symptoms, loneliness or isolation, feeling misunderstood, self-sacrifice (e.g., feeling like they can't take a break), shutting down emotionally, feeling powerless, disillusionment (e.g., nothing changes, or work is futile), and diminished empathy (e.g., frustration that people fall for misinformation). They reported being acutely conscious of the weight of responsibility for others' lives, society overall, and how history will judge them. Additionally, some people spoke about how the work negatively shaped their worldview, for example, finding themselves jumping to the worst-case scenario in reaction to events in their personal lives.

They also told us about their strategies for coping with or managing stress, even though we did not directly ask about this. These strategies were often shared in response to an opening question, "How do you spend your days?" They described establishing routines, setting up reminders, taking breaks, doing something else (in the moment or in the longer-term), connecting with colleagues, and making time for moments of appreciation, joy,

celebrating wins, and humor. They used routines, rituals, and other strategies to maintain boundaries between work and personal life. An important part of how they manage the challenges of their work is grounding themselves in their motivations: protection and safeguarding, passion and purpose, and enjoyment of the challenge of complex problems under high-pressure circumstances. These are similar to those described by Spence et al. (2023), which we summarized earlier: talking to others, connecting with colleagues, autonomy, recognition, and boundaries.<sup>6</sup>

There may be “constitutional differences,” as Ella said, in the people who choose this work. Nevertheless, they are exposed to trauma. In the current configuration, T&S professionals individually manage the challenges inherent in T&S work while also grappling with compounding stressors from working conditions. Beyond individuals’ strategies for managing the stress of T&S work, companies are developing wellness resources and other support mechanisms. Spence et al. (2023) and Steiger et al. (2021) included therapy and technological tools as organizational supports; both of these were mentioned by interviewees in this project. Interviewees in this study were not primarily content moderators, and so perhaps for this reason most mentions of technological tools emphasized the capacity to scale T&S work, rather than specific mitigation tools for content review (e.g., greyscale or blurring). Results from a first-ever global compensation survey published by TSPA found that in addition to company-wide benefits, T&S employees reported benefits specific to their roles including support for health and wellness, additional paid time off (PTO) for mental health days, stipends for hobbies or activities, subscriptions for digital protection from doxxing, and flexible benefits for work-life balance. Crucially, among wellness resources they noted that “digital-first programs, in particular, were seen as out of touch with the needs of workers already tethered to screens” (TSPA 2025, 27).

Mitigating the harms of T&S work through a combination of individual strategies and company wellness initiatives is an important minimum. We argue, however, that emotion is an essential component in all levels of T&S work, and should be recognized as such when developing well-being initiatives. Automation is increasingly used to moderate content, respond to user reports, flag bad actors across accounts and platforms, and manage trust and safety workflows. However, many T&S problems—with their high degree of complexity—are not so easily sorted. While much prior research has emphasized the vulnerabilities of front-line content moderators, our findings suggest that T&S professionals’ motivations and affective-relational skills are also assets. Recognizing these skills as assets reframes well-being as a shared, organizational responsibility rather than a burden carried solely by individual workers.

## 6.2 The Essential Role of Emotion in Trust and Safety Work

Now we return to conceptualizations regarding the role of emotion in knowledge work introduced in the Literature Review section. Pinchevski (2023) argued that the ability to be traumatized is essential to effective content moderation, and that it is therefore a commodity in platform-side T&S. Our findings concur with the vital role of affect in T&S

---

6. Absent from this list was physical exercise, though this could be a result of our interview protocol not directly asking about it.

work, though perhaps because those interviewed for this study experience relatively less precarity and have relatively more autonomy than front-line content moderators, other possibilities present themselves. While Pinchevski asserted that emotions are an exploitable resource and therefore a liability to be managed, we argue that a more nuanced perspective can account for emotions as essential to the effectiveness of every aspect of T&S work and point to pathways for more holistic well-being.

As we discussed earlier in this paper, Craciun (2018) proposed a framework of affective-relational skills that can drive expert knowledge work in three roles: as motivation (*supportive*), in relations with others (*didactic*), and in problem-solving (*inductive*). While any framework may be imperfect, we are using Craciun's as a lens to understand and interpret how affective-relational skills become evident in T&S work as described by interviewees. Emotions play a *supportive* role in the motivations participants described, such as a desire to protect others, passion for countering harms, or connections to colleagues. T&S professionals' empathy and collaboration rely on the *didactic* role of emotions as they communicate and relate to others. Finally, emotions are at the core of the *inductive* work of "seeing both sides" and in creative and iterative development of solutions. In short, emotions play key roles in T&S expert work including content moderation, incident response, investigation, policy and tooling development, and product consultation. In Table 4 we apply Craciun's didactic and inductive categories of the use of emotion to a sampling of T&S tasks beyond emotional regulation: for example, classification, considering a case in its context(s), discerning trade-offs, prioritization, and developing innovative solutions. Note: the supportive role simply supports professionals' motivation, and is not included in the table for the sake of brevity.

Wang (2022) articulated a "desired state of mind" which balances emotion and cognition, allowing professionals to fit messy human problems into systemic categories and processes. Playful acknowledgment of the internal manifestation of Wang's desired state of mind can be read into the "This Is Fine" dumpster fire meme popular among T&S professionals: "This is fine. I'm fine. Everything's fine" (Green 2016). Our findings suggest that this desired state of mind is utilized by T&S professionals in three aspects of their work. First, the desired state of mind allows T&S professionals to translate highly contextual, "complex, messy, and ambiguous" human communications and behaviors into data that can be processed through the systems of platform bureaucracy (e.g., policies, classifiers, tooling, user interfaces, workflows; see Shulruff, Lazarus, and Menking (2026)). Second, a balance of affect and rationality supports an empathetic yet consistent response to reports of harm and other urgent situations. Third, T&S professionals invoke the desired state of mind in order to mediate between actors during collaboration (internally and externally).

### 6.3 The Value of Emotion

Platforms benefit from this desired state of mind, and seek it out, as reflected in our findings. An analysis of job postings bears this out, too: platforms let prospective

Table 4: Examples of Inductive and Didactic Affective-Relational Skills in T&amp;S Work

Inductive	Didactic
<ul style="list-style-type: none"> <li>• Classification (i.e., fit a case to policy or make rules from individual cases)</li> <li>• Sense whether something is violative</li> <li>• Consider a case in its context(s)</li> <li>• Discern useful information from an individual's report (story, text, image, etc.)</li> <li>• Assess seriousness</li> <li>• Identify options</li> <li>• Discern trade-offs</li> <li>• Write up incident or programmatic reports</li> <li>• Triage and prioritization</li> <li>• Draw connections between incidents and external contexts</li> <li>• Imagine scenarios</li> <li>• Develop innovative solutions</li> <li>• Argue for the value of T&amp;S</li> <li>• Discern trauma responses vs. other behaviors</li> <li>• Plan for supports in high stress circumstances</li> </ul>	<ul style="list-style-type: none"> <li>• Emotional regulation</li> <li>• "Emotional labor" <ul style="list-style-type: none"> <li>– performing an emotion (e.g. calm, pleasant)</li> <li>– using emotions to move others (calm others down, get them to take action)</li> </ul> </li> <li>• Listen to multiple perspectives</li> <li>• Comfort with "grey areas"</li> <li>• Hold opposing needs and priorities</li> <li>• Awareness of real impacts of policies and systems</li> <li>• Apply affective insights to policy, processes, or UX/UI design</li> <li>• Empathy</li> <li>• Diplomacy, mediation</li> <li>• Relationship building</li> <li>• Separate feelings or values from company's</li> <li>• Modulate interactions with colleagues, managers, friends/family</li> </ul>

employees know they are expected to have "passion" but also "tact," "a positive attitude," and a "calm demeanor," and to be "polite," "fair," "modest," "approachable" (Shulruff 2024). Writing about their approach to building a T&S team at Zoom, Maxim, Parecki, and Cornett (2022) emphasized the importance of hiring people who are able to build relationships across functions, "won't run away" from the subject matter, and are comfortable working in "grey areas." Despite the demand for these characteristics and skill sets, T&S roles and teams are structurally disadvantaged within platforms (TSPA 2025), as Wang (2022, 19) argued is common for those holding affective roles.

Society also relies on this translation work, especially in light of increasing attention to harms facilitated by platform technologies and the related enacting of regimes of regulatory compliance. Yet, unlike first responders who also bridge messy human problems and the systems we've developed to process them, T&S professionals are invisible technicians, hidden inside the tech stack, much like the 19th century laboratory assistants essential to foundational scientific work (Shapin 2011). Steiger et al. (2021) argue that "narratives that disparage moderation as a 'dirty' job do not recognize or honor the essential contribution of this profession in safeguarding the internet for the rest of us" (10). The unglamorous or even ugly side of the internet is often "swept under the rug," something external societal actors may prefer to be ignorant of (Proctor and Schiebinger 2008), becoming visible only when something goes wrong. Invisibility can be problematic

when “visibility may be tied crucially to systems of reward and recognition” (Jackson 2014, 229). Conversely, invisibility may be appealing at times, as T&S professionals in their personal lives experience a range of others’ negative reactions: blame, voyeuristic interest, avoidance, and misunderstanding.

#### 6.4 Implications

T&S professionals bear the human cost of bridging emotion and cognition, harmful content and behavior and the systems they helped build to manage that, and platforms and society. However, we argue that emotions are not a liability but an asset, and so this study has sought not only to faithfully relate some of the costs, but also to elevate the motivations, the skills, and the strategies T&S professionals employ to protect users and keep platforms going. Rather than merely being exploited, these affective resources could be sustained through organizational support for holistic well-being. T&S professionals have thus far developed individual strategies to maintain resilience, work-life balance, self-care, authenticity, and a positive outlook on their own. However, the weight of responsibility for their own well-being should not rest solely with individuals.

Companies, regulators, and societies should acknowledge the essential role of platform-side Trust & Safety in protecting the technology we rely on every day. If the emotion work of T&S professionals drives company profitability and societal benefits, then their emotional capacity, their ability to maintain a “desired state of mind,” and their overall well-being should be cultivated and sustained. Like other organizations, platforms can promote healthy workplaces through measures including decent compensation, benefits, wellness programs, and flexibility (Day and Randell 2014). Technological wellness solutions are also being developed and proposed to remediate the impacts of the work on T&S professionals through skills training, wellness reminders, and interventions. And yet, just as the problems themselves cannot be adequately addressed by technology alone, nor can the well-being of T&S professionals be entirely fixed by technology.

The well-being of those who do this work should be rooted in an acknowledgment of their affective expertise, driven by the needs and goals they articulate, and supported by adequate organizational and societal resources. In other words, T&S professionals should be invited to define for themselves their “desired state of mind” and outline how that is achieved and maintained. Rather than needing to do so individually, initiatives to propose definitions and measures of well-being are already underway within the trust and safety field (e.g., the Trust & Safety Professional Association’s Cost of Caring workshop, the Trust and Safety Foundation’s Psychological Health Research Committee, and the Workplace Wellness Project (TSPA 2024; Trust and Safety Foundation, n.d.; Workplace Wellness Project, n.d.)). Additionally, organizational practices can include disclosing risks during hiring, limiting exposure, enabling moderators and other T&S professionals to create space between themselves and work, and building connections with others (Steiger et al. 2021). Flexibility and autonomy are key; just as different content impacts different people, different strategies work for different people.

## 7 Limitations

This study draws on a dataset of 47 interviews with current and former T&S professionals, and therefore the findings reflect a range of individuals' lived experiences, motivations, and well-being strategies which may not match those of every person who works in platform-side T&S; however, qualitative research is not intended to be generalizable. Less than half of interviewees had held content moderation or customer service roles. Their perspectives broaden our understanding of the nature of trust and safety work beyond content moderation on social media and at the level of lowest-paid work. However, as the trust and safety ecosystem rapidly evolves, continued attention to the changing conditions of the lowest-paid workers is an important area for research and intervention. Our dataset includes limited representation from T&S professionals in the Majority World; further work is needed to examine the commonalities and disparities in their experiences. Similarly, shifts in the T&S ecosystem related to mitigating and utilizing generative AI are not reflected, as our interviews concluded in Summer 2024. The impact of these shifts on the day-to-day work and application of affective-relational skills deserves attention. Finally, a more in-depth study of these affective-relational skills, how they are applied, and how T&S professionals achieve and maintain the "desired state of mind" could be used to develop comprehensive, holistic well-being initiatives for the field.

## 8 Conclusion

Harmful content, abusive behavior, and exploitation happen every day, around the clock and around the world. This is not caused by technology, but it *is* facilitated and exacerbated by technological affordances. This is a "wicked problem" that is complex in dynamics and has no simple solution (Crowley and Head 2017). Platforms have come to rely on the evolving profession of T&S to address these risks and harms beyond social media and content moderation.

There are people—like this study's participants—who are willing to take on these problems, and who have the motivation and expertise to do so. They rely on a mixture of affective and cognitive capacities, a "desired state of mind." As AI responsibility work overlaps increasingly with T&S, this balance is more important than ever (Wright 2025). Platform-side T&S comes at a cost; stressors like subject matter, scale, pace, complexity, and ambiguous contexts are compounded by working conditions, curtailed resources, and organizational structures. Furthermore, a lack of visibility—or when there is visibility, a lack of understanding—from the public, policymakers, and other societal actors adds to the stress. Their strategies for navigating these many challenges often draw on the same expertise and motivations which they use in their work.

Platforms, indeed our societies, depend on platform-side Trust and Safety. Approaches to employee wellness which focus on resilience training and trauma exposure mitigation

help T&S professionals to not be vulnerable, to not become damaged, and to not become a liability. However, by shifting from seeing emotions as a liability to valuing the affective skills T&S professionals bring to their work as an asset, companies can leverage these capacities as essential resources for maintaining functional, safe, and accountable digital environments. Furthermore, the wellness paradigm should shift from coping toward well-being, and from individually-borne responsibility toward organizational support.

## References

- Bhatlapenumarty, Harsha, Eric Davis, and Daphne Keller. 2021. "Key Functions and Roles." Trust & Safety Professional Association. Accessed February 22, 2026. <https://www.tspa.org/curriculum/ts-curriculum/functions-roles/>.
- Cath-Speth, Corinne. 2021. "The Internet's Reluctant Sheriffs: Content Moderation and Political Gatekeeping Through Internet Infrastructure." Video talk, Oxford Internet Institute, University of Oxford, June 16, 2021. <https://www.oii.ox.ac.uk/news-events/videos/the-internets-reluctant-sheriffs-content-moderation-and-political-gatekeeping-through-internet-infrastructure/>.
- Craciun, Mariana. 2018. "Emotions and Knowledge in Expert Work: A Comparison of Two Psychotherapies." *American Journal of Sociology* 123 (4): 959–1003. <https://doi.org/10.1086/695682>.
- Crowley, Kate, and Brian W. Head. 2017. "The Enduring Challenge of 'Wicked Problems': Revisiting Rittel and Webber." *Policy Sciences* 50 (4): 539–47. <https://doi.org/10.1007/s11077-017-9302-4>.
- Cryst, Elena, Shelby Grossman, Jeff Hancock, Alex Stamos, and David Thiel. 2021. "Introducing the Journal of Online Trust and Safety." *Journal of Online Trust and Safety* 1 (1). <https://doi.org/10.54501/jots.v1i1.8>.
- Day, Arla, and Krista D. Randell. 2014. "Building a Foundation for Psychologically Healthy Workplaces and Well-Being." In *Workplace Well-Being*, 1st ed., edited by E. Kevin Kelloway, Arla Day, and Joseph J. Hurrell, 1–26. Wiley. <https://doi.org/10.1002/9781118469392.ch1>.
- Denyer Willis, Graham. 2023. "'Trust and Safety': Exchange, Protection and the Digital Market—Fortress in Platform Capitalism." *Socio-Economic Review* 21 (4): 1877–95. <https://doi.org/10.1093/ser/mwad003>.
- DTSP. 2025. *DTSP Safe Framework Specification*. Digital Trust and Safety Partnership. <https://dtspartnership.org/the-safe-framework-specification/>.
- Eissfeldt, Jan, Jeff Lazarus, and Pia Shah. 2021. "Industry Overview." Trust & Safety Professional Association. Accessed February 22, 2026. <https://www.tspa.org/curriculum/ts-fundamentals/industry-overview/>.
- Feerst, Alex. 2019. "Your Speech, Their Rules: Meet the People Who Guard the Internet." OneZero, February 27, 2019. <https://onezero.medium.com/your-speech-their-rule-s-meet-the-people-who-guard-the-internet-ab58fe6b9231>.
- Gauthier, Gabrielle M., Eesha Ali, Amna Asim, Sarah Cornell-Maier, and Lori A. Zoellner. 2025. "I've Seen Enough: Measuring the Toll of Content Moderation on Mental Health." *arXiv*, <https://doi.org/10.48550/arXiv.2511.09813>.

- Gillespie, Tarleton. 2019. *Custodians of the Internet: Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media*. Yale University Press. <https://doi.org/10.12987/9780300235029>.
- Gorwa, Robert. 2024. *The Politics of Platform Regulation: How Governments Shape Online Content Moderation*. Oxford Studies in Digital Politics. Oxford University Press. <https://doi.org/10.1093/oso/9780197692851.001.0001>.
- Green, K.C. 2016. "This Is Fine." Know Your Meme. Digital image. Accessed March 18, 2026. <https://knowyourmeme.com/memes/this-is-fine>.
- Jackson, Steven J. 2014. "Rethinking Repair." In *Media Technologies: Essays on Communication, Materiality, and Society*, edited by Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot. MIT Press. <https://doi.org/10.7551/mitpress/9042.003.0015>.
- Klonick, Kate. 2017. "The New Governors: The People, Rules, and Processes Governing Online Speech." *Harvard Law Review* 131 (6): 1598–670. <http://www.jstor.org/stable/44865879>.
- Lipsky, Laura van Dernoot, and Connie Burk. 2009. *Trauma Stewardship: An Everyday Guide to Caring for Self While Caring for Others*. Berrett-Koehler Publishers.
- Maxim, Karen, Josh Parecki, and Chanel Cornett. 2022. "How to Build a Trust and Safety Team in a Year: A Practical Guide from Lessons Learned (So Far) at Zoom." *Journal of Online Trust and Safety* 1 (4). <https://doi.org/10.54501/jots.v1i4.81>.
- Mihalache, Diana, and Dalila Szostak. 2026. "Wellbeing-Centered UX: Supporting Content Moderators." In *Trust, Safety, and the Internet We Share: Multistakeholder Insights*. Taylor & Francis.
- Moran, Rachel Elizabeth, Joseph Schafer, Mert Bayar, and Kate Starbird. 2025. "The End of Trust and Safety?: Examining the Future of Content Moderation and Upheavals in Professional Online Safety Efforts." In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, 1–14. Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713662>.
- Newell, Jason M., and Gordon A. MacNeil. 2010. "Professional Burnout, Vicarious Trauma, Secondary Traumatic Stress, and Compassion Fatigue." *Best Practices in Mental Health* 6 (2): 57–68. <https://doi.org/10.70256/607490pbruec>.
- Pinchevski, Amit. 2023. "Social Media's Canaries: Content Moderators Between Digital Labor and Mediated Trauma." *Media, Culture & Society* 45 (1): 212–21. <https://doi.org/10.1177/01634437221122226>.
- Proctor, Robert N., and Londa Schiebinger, eds. 2008. *Agnology: The Making and Unmaking of Ignorance*. Stanford University Press.
- Riesewieck, Moritz, and Hans Block, dirs. 2018. *The Cleaners*. Christian Beetz Filmproduktion.

- Roberts, Sarah T. 2021. *Behind the Screen: Content Moderation in the Shadows of Social Media: With a New Preface*. Yale University Press.
- Shapin, Steven. 2011. *A Social History of Truth: Civility and Science in Seventeenth-Century England*. Science and Its Conceptual Foundations series. University of Chicago Press.
- Shulruff, Toby. 2024. *Trust and Safety in Everyday Technologies: Research Highlights for Stakeholders in Government and Civil Society*. Technical report. Arizona State University. <https://sites.google.com/asu.edu/trustandsafety/2024-report>.
- Shulruff, Toby, Jeff Lazarus, and Amanda Menking. 2026. "Voices of Trust and Safety: Origins and Evolutions." In *Trust, Safety, and the Internet We Share: Multistakeholder Insights*, edited by Maia Levy Daniel, Amanda Menking, Marlyn Savio, and Jean Claffey. Taylor & Francis.
- Spence, Ruth, Amy Harrison, Paula Bradbury, Paul Bleakley, Elena Martellozzo, and Jeffrey DeMarco. 2023. "Content Moderators' Strategies for Coping with the Stress of Moderating Content Online." *Journal of Online Trust and Safety* 1 (5). <https://doi.org/10.54501/jots.v1i5.91>.
- Steiger, Miriah, Timir J. Bharucha, Sukrit Venkatagiri, Martin J. Riedl, and Matthew Lease. 2021. "The Psychological Well-Being of Content Moderators: The Emotional Labor of Commercial Moderation and Avenues for Improving Support." In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14. Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445092>.
- Trust and Safety Foundation. n.d. "Psychological Health Research Committee." Accessed March 18, 2026. <https://www.trustandsafetyfoundation.org/psychological-health-research-committee>.
- TSPA. 2024. "The Cost of Caring: Understanding Trauma Exposure Responses & Resilience in Digital Safety Work." Trust & Safety Professional Association. Accessed March 18, 2026. <https://www.tspa.org/event/the-cost-of-caring-understanding-trauma-exposure-responses-resilience-in-digital-safety-work/>.
- . 2025. *T&S Professionals: Global Compensation Report 2024*. Technical report. Trust & Safety Professional Association. <https://www.tspa.org/ts-professionals-global-compensation-report/>.
- Tuck, Eve. 2009. "Suspending Damage: A Letter to Communities." *Harvard Educational Review* 79 (3): 409–28. <https://doi.org/10.17763/haer.79.3.n0016675661t3n15>.
- Tyler, Tom R., Tracey L. Meares, and Matt Katsaros. 2025. "New Worlds Arise: Online Trust and Safety." *Annual Review of Criminology* 8 (1): 171–92. <https://doi.org/10.1146/annurev-criminol-111523-122337>.

- Voth Schrag, Rachel J., Leila G. Wood, Karin Wachter, and Shanti Kulkarni. 2022. "Compassion Fatigue Among the Intimate Partner Violence and Sexual Assault Workforce: Enhancing Organizational Practice." *Violence Against Women* 28 (1): 277–97. <https://doi.org/10.1177/1077801220988351>.
- Wajcman, Judy. 1991. *Feminism Confronts Technology*. Polity Press.
- Wang, Phoenix Chi. 2022. "The 'Desired State of Mind': Emotional Labor and the Hidden Cost of Symbolic Power in 911 Emergency Response in the U.S." *Ethnography*, <https://doi.org/10.1177/14661381221145353>.
- Weigl, Linda, and Balázs Bodó. 2025. "Trust and Safety: What's in a Name?" *SSRN Electronic Journal*, <https://doi.org/10.2139/ssrn.5404950>.
- Workplace Wellness Project. n.d. "The Workplace Wellness Project." Accessed March 18, 2026. <https://theworkplacewellnessproject.com/>.
- Wray, Ben. 2025. "Content Moderation is What a 21st Century Hazardous Job Looks Like." *Equal Times*, April 27, 2025. <https://www.equaltimes.org/content-moderation-is-what-a-21st?lang=en>.
- Wright, Lucas. 2025. "The Salesforce of Safety: Software Vendors as Infrastructural/Professional Nodes in the Field of Online Trust and Safety." *Platforms & Society* 2. <https://doi.org/10.1177/29768624251390683>.

## Authors

**Toby Shulruff** is a Research Assistant for the History of Trust and Safety project at the Trust and Safety Foundation (TSF), and a PhD candidate at Arizona State University in the Human and Social Dimensions of Science and Technology program. Her research includes the trust and safety field, global futures, and everyday and emerging technologies. Toby works at the nexus of technology, people, and planet to build our collective capacity to understand and shape technologies woven into the fabric of our lives (tshulruf@asu.edu).

**Amanda Menking** is the Research and Program Director at the Trust and Safety Foundation (TSF). A qualitative researcher with an interest in Science and Technology Studies and Human-Computer Interaction, she completed her PhD in Information Science at the University of Washington. Amanda joined TSF after spending almost a decade in academia as a PhD student, postdoc, and Associate Teaching Professor—researching and teaching about bias, knowledge production, and safety in online communities.

## Acknowledgements

We are deeply grateful to study participants for sharing their time, knowledge, expertise, and stories with us. We are also grateful to Lucas Wright, Marlyn Savio, Megan Lindsay Brown, Sarah Knowles, and Maggie Kerr for their early feedback, and to the editors and anonymous peer reviewers.

## Data availability statement

Not applicable.

## Funding statement

The authors declare that they have no relevant or material financial interests that relate to the research described in this paper.

## Ethical standards

## Keywords

Trust and safety; online platforms; emotional labor; affective-relational skills.

# Appendices

## Appendix A: Interview protocol

### A.1: Introduction

Thank you so much for taking the time to speak with me today. My name is [interviewer name], and I'm the [interviewer title].

We've decided to conduct these interviews because we want to learn more about the history of trust and safety and, hopefully, document it for others.

There are no wrong or right answers. We just want to learn from you!

Before we begin, I want to ensure I have informed consent, so I'm going to pause and answer any questions you may have.

[Pause]

I would like to record this interview so that we can transcribe it and then compare data across our interviews. The recording will be transferred from this device to a password protected folder that will be shared only with members of the research team and the transcriptionist.

Do I have your permission to record?

[Pause]

*[If no, proceed to ask for permission to take notes.]*

Do I have your permission to take notes?

[Pause]

*[If no, end the interview.]*

Okay, great! Let's get started. ...

*[Start recording if consent has been given.]*

Do you consent to this interview?

### A.2: Interview Questions

#### 1. Tell me about yourself.

- a. *Note: If interviewee doesn't share the following demographic information, please ask for it specifically: years in T&S ecosystem, country of origin, current*

*country of residence, gender, current role.*

2. **How do you spend your days?**
  - a. Tell me about your current role.
3. **Tell me about how you got started in trust and safety.**
  - a. When did you start working in T&S?
  - b. How did you enter the field?
  - c. What was your first job in T&S?
  - d. Why did you want to work in trust and safety?
4. **How has T&S changed since you began working in the field?**
  - a. What are some key events or incidents that have shaped T&S?
  - b. What are some technological developments that have shaped T&S?
  - c. What are some societal shifts that have shaped T&S?
5. **Who has shaped your T&S practice—for better or worse?**
  - a. Who's inspired you along the way?
  - b. With whom do you most often collaborate?
  - c. With whom would you love to partner?
6. **How would you describe the field of trust and safety today?**
  - a. What's most important/pressing/critical in T&S right now?
  - b. What do people working in T&S today have in common?
7. **What do you think the future of T&S looks like?**
8. **Anything else you'd like to share?**

### **A.3: Closing**

Thank you so much for sharing your time, experiences, and expertise with me.

Once the recording of this interview has been transcribed, I'll send the transcript to you so that you can review it. At the time, you can add anything you wish you had shared with me today and you can also redact anything that you shared with me today that you wish you wouldn't have.

If you have any questions or want to follow up on what you've shared today, please feel free to reach out to me at [interviewer email address].