Emotional privacy: a philosophical investigation of emotion recognition technology

Prégent, A.P.

6. Emotional privacy as a moral right? Normative foundations and technological implications[153]

6.1. Introduction

In the preceding chapter, I re-engineered the concept of emotional privacy and articulated its core functions. Chapter 5 showed that emotional privacy secures the conditions under which emotional expressions fulfil their communicative and regulative roles in social life. This established *why* the concept of emotional privacy matters for assessing new and emerging technologies such as ERTs, yet it made no normative proposition on how those technologies ought to be evaluated nor what type of wrongs, if any, follow from a defeat of expressive ability. This chapter takes up that normative task. I defend emotional privacy as a distinct normative concern that reveals specific structural harms overlooked by existing normative (informational) privacy frameworks.

As argued in the previous chapters, ordinary mindreading and partial emotional legibility are not only inevitable but socially necessary. The aim of this chapter is thus twofold: first, to specify what a moral right to emotional privacy should protect in terms of preserving individuals' expressive ability, so that ordinary, acceptable losses of emotional privacy (actual losses) can be distinguished from wrongful ones (standing losses); and second, to demonstrate its moral relevance by applying it to concrete technomoral scenarios and to established ethical evaluation frameworks.

The chapter proceeds in four steps. **Section 6.2** articulates a normative account of emotional privacy and specifies the conditions under which emotional privacy would be violated, namely the defeat of expressive ability. **Section 6.3** applies this account to three technomoral scenarios to illustrate the types of interests at stake and to clarify the scope of the right across three levels of interaction (individual, interpersonal, societal). **Section 6.4** shows that current dominant normative variants of informational privacy fail to fully capture these interests, thereby demonstrating that emotional privacy fills a genuine normative gap. **Section 6.5** then examines

[153] Parts of this Chapter are published in Prégent (Forthcoming a).

how the reliance on emotional privacy may recalibrate ethical evaluation frameworks, including Contextual Integrity and ALTAI-style proportionality and subsidiarity assessments.[154]

The core claim defended in this chapter is that violations of emotional privacy give rise to a distinctive kind of harms that existing normative accounts of informational privacy do not fully grasp. Fully functional ERTs risk inducing this kind of harms by defeating people's expressive ability thereby collapsing the functions currently served by EEs (especially intentional EEs) in social life. The normative account of emotional privacy proposed in this chapter therefore provides the normative grounding needed to fully grasp this technological reconfiguration of social life.

6.2. Formulating a normative account of emotional privacy

Several privacy scholars, including Warren and Brandeis (1890), Westin (1967), Allen (1988), Taslitz (2002), and more recently, Gremsl and Hödl (2022), and Menges and Weber-Guskar (2025), have considered or defended the idea that our emotions warrant a special kind of protection, sometimes framed as a (moral) right to the privacy of emotions. My argument departs from these approaches by shifting the focus from emotions as inner or mental states to emotional life as a socially embedded phenomenon. I thus treat emotional privacy as a distinct normative dimension of social life.

Foregrounding emotional privacy as a distinct normative dimension of social life immediately raises a familiar normative challenge: what should count as an appropriate expectation of privacy for emotional expressions, given that we both involuntarily leak and deliberately share them across private and public contexts? Not every loss of privacy is normatively problematic, and not every loss should be conceptualised in terms of privacy violations. As argued in the previous chapter, ordinary mindreading and (partial) emotional legibility are not only inevitable but socially necessary because they are part of what makes (intentional) emotional expressions capable of fulfilling their communicative and regulative roles.

Emotional privacy, as I understand it, therefore cannot require individuals to retain full control over the capacity to successfully modulate when, how, and to whom one's EEs will be legible. Nor should it be understood as shielding emotional life from all forms of access. Given ordinary

[154] ALTAI stands for the Assessment List for Trustworthy Artificial Intelligence.

affective leaks and the inherently social nature of emotional life, such requirement would be unrealistic. Instead, the normative account of emotional privacy should only track whether individuals retain expressive discretion over the legibility of their affects within affective communication channels. Hence, in this chapter, I will only focus on the normative stakes associated with standing emotional privacy.

Expressive discretion refers to the ability one has over the legibility of their emotional expressions (notably through the use of privacy mechanisms (see Chapter 5)). Thus, expressive discretion can only exist given the presence of a structural background of opacity against which EEs can be negotiated, contested, or strategically shaped. It is this background of managed opacity that enables individuals to (i) present the self they see fit for a given audience, (ii) sustain role plurality and adhere to role-bound norms of expressiveness, and (iii) contribute to social regulation and cohesion. These interests form a downstream normative trajectory: when expressive discretion is undermined, self-presentation is first compromised, itself destabilising role plurality and adherence to role-bound norms of expressiveness, and the regulative and cohesive functions of emotional expressions begin to thin within and outside affective communication channels.

On this view, emotional privacy *obtains* when individuals retain the expressive ability to control their affective information. Emotional privacy is *lost* when this expressive ability is defeated. The key question, then, is how to distinguish ordinary, acceptable losses of emotional privacy – i.e. leaks, misfires, unusually acute human perception – from wrongful losses, especially in technologically-mediated contexts. To mark this distinction, I rely on the standing definition of emotional privacy. Emotional privacy is *defeated* when a person cannot meaningfully use their EEs and privacy mechanisms to manage the legibility of their affective life, such that emotional expressions cannot fulfil their social and pragmatic functions in that context. Defeat thus concerns the loss of expressive ability in a given configuration of the affective environment, not merely discrete instances of (unwarranted) access or momentary loss of control.

This moral right to emotional privacy can be stated as follows:

> **Moral right to emotional privacy**: A's moral right to emotional privacy is violated iff A's expressive ability over affective information p is defeated.

The basis of a moral right to emotional privacy thus lies in the value of protecting one's expressive ability from defeat. I will come back to what this defeat entails, but first, some clarifications on the kind of moral right in question. Emotional privacy should be understood as a distinct moral right within the normative informational privacy cluster. Consequently, I follow most of the normative accounts in informational privacy and assume here that it is widely accepted that a moral right is justified insofar as it protects certain interests of X, whose realisation contributes to X's well-being. In the case of the moral right to emotional privacy, it protects the interests of (i) self-presentation, (ii) role plurality and adherence to role-bound norms of expressiveness, and (iii) social regulation and cohesion. Within the affective communication channels, these interests cannot be secured without the individual retaining her expressive ability, hence why expressive ability is the object of protection in the case of emotional privacy. The moral right to emotional privacy is a pro tanto, interest-based moral right.[155] Moreover, since the interests follow a downstream trajectory, securing (i) is a precondition for the realisation of (ii)-(iii). The right is thus individually *held* yet it secures interests that are also societal interests as self-presentation, role plurality and norm conformity, and social regulation and cohesion are *constitutive* of both individual and societal well-being.

6.2.1. Human vs technology-based emotion recognition

While Chapter 5 demonstrates that actual emotional privacy is often lost through affective leaks and (successful) ordinary mindreading, it would be counterintuitive to argue that such losses constitute a violation of privacy as they are part of social life. Hence, not all losses of emotional privacy should be morally conceptualised in terms of privacy rights, and a robust account of privacy should meet this difficult challenge.

The distinction between actual and standing forms of emotional privacy helps to distinguish ordinary losses of emotional privacy from wrongful ones by excluding affective leakages, misfires, and (unusually) acute human perception from counting as violations of a person's

[155] Justifying the moral right to privacy following the argument that it protects certain interests of X, whose realisation contributes to X's well-being, subscribes to an interest-based approach to individuals' right, as opposed to a will-based (choice-based) approach to individuals' right. The latter instead defends the idea that a right is essentially an opportunity for an individual (or group) to make choices. Instead of tying rights to various aspects of the right-holder's well-being, the Will Theory of rights holds that each right gives its holder a certain measure of control over their situation. To say that someone has a right is to say that they are empowered to decide whether and how another person's duty towards them is fulfilled (see Kramer, Simmonds, & Steiner (1998) for a debate on the two approaches. See also Raz (1994, chap. 2-3) for a defense of the interest-based view and Hart (1955) for a defense of the will-based view).

moral right to emotional privacy, while including recognition by technology such as fully functional ERTs. This allows me to give a clear answer to the question of what kind of normative privacy expectations should be appropriate regarding the case of emotional privacy.

This same difficulty animates long-standing debates about surveillance. While surveillance technologies are often assessed in terms of privacy, several scholars have expressed dissatisfaction with privacy-based objections to new and allegedly intrusive technologies (Ryberg 2007, 2017; Macnish 2018). Jesper Ryberg (2007, 2017) provides one of the most influential challenges to privacy-based objections in this domain. Discussing CCTV in public spaces, he argues that privacy theorists have failed to show why CCTV is morally problematic as a privacy violation. If there is a morally relevant difference between an ordinary person observing a street from their window and a CCTV camera recording it, he claims, "it must lie inherently in the way each of the two operate" (2007: 131); yet, according to Ryberg, no such difference obtains. He later deploys a similar rationale against advocates of neurorights (2017), arguing that they have not shown why neurotechnological mindreading violates mental privacy if ordinary mindreading does not. Since much of social life depends on mutual access to affective information, and since EEs routinely reveal sensitive information (e.g. through blushing, sighing, or frowning), it might seem that technological emotion recognition is simply an extension of human recognition, and thus normatively continuous with it.

Two notable responses by Annabelle Lever (2008) and Benjamin Goold (2008) already question Ryberg's analogy by pointing to power asymmetries and institutional context. Lever situates privacy within a broader democratic framework, arguing that state-operated CCTV differs categorically from informal observation between citizens mainly because of the coercive power of the state and the requirements of democratic justification (Lever 2008: 38). Goold similarly emphasises the institutional nature of state surveillance and its capacity for monitoring and data retention (Goold 2008: 45).[156, 157] Yet Ryberg (2008, 2017) remains unconvinced, claiming that these responses do not demonstrate how the use of CCTV, or neurotechnological mindreading, wrongs individuals' privacy.[158]

[156] Goold's argument also includes the possibility that the old lady is violating people's privacy.
[157] Interestingly, both critics points point to the risks that follow from arguments that depoliticise new technologies that are intrinsically political, such as state-operated CCTV cameras.
[158] Although it sometimes seems like Ryberg would want privacy scholars to provide a dichotomous privacy framework where technology-based practices are *always* wrong and human-based practices are *always* legitimate, the view I am defending is more nuanced, and, as the reader shall see, I do not exclude the possibility that technology-based recognition practices may, in some contexts, be warranted.

Below I anticipate Ryberg's challenge as a main objection for the case that ERTs may not be a threat to individuals' privacy, in the same way as CCTV cameras and neurotechnologies are described as not being a threat either. I rebut his challenge by taking up his suggestion that the relevant difference lies in the intrinsic differences between human- and technology-based emotion recognition. I show how these differences are inherent to the way the two operate and how in the case of ERTs, they become normatively salient precisely through the lens of the moral right to emotional privacy.

In Chapter 4, I argued that human and technological recognition rely on different *epistemic routes*. Human recognition typically occurs *in situ*, on a pro tanto basis, within shared contexts of meaning and expectation. It is inherently defeasible because it depends on background knowledge, social conventions, and mutual adjustment. Human perceivers can misread, but they can also revise their interpretations. Uptake is therefore both negotiated and constrained (the latter being the case notably through norms of expressiveness and affective leaks). Recognition here thus belongs to a communicative regime, in which EEs typically function as contributions to an ongoing interaction, serving the emoter, the perceiver, and society as a whole. ERT-based recognition operates differently. In Chapter 4, I also argued that, in contrast with human-based recognition, ERT-based recognition is one not of communication but of diagnostic, as it uses data points used to infer, classify, and predict inner affective states.

Innovation theory helps to further capture the fundamental differences between human-based and ERT-based recognition. In innovation theory, a common distinction is drawn between what is called incremental innovation and radical innovation. Incremental innovation involves refinements that improve performance while leaving the core concept underlying the technology intact. For instance, better lenses or higher resolution in smartphone cameras sharpen images without altering their basic purpose. Radical innovation, by contrast, overturns the core concept of an entrenched technology or practice, as with the first photographic camera replacing drawing and painting with chemically captured light, or the airplane introducing a new mode of transportation (Christensen 1997). With radical innovation, the purpose of a given technology is reshaped to satisfy new ends.

What Ryberg argues is that CCTV cameras, or even neurotechnologies (and I presume he would say the same for ERTs) should be seen as incremental improvements vis-à-vis basic human capacities and practices as they merely enhance what humans already do, but they do not

overturn the practice; surveillance is still surveillance, and mindreading is still mindreading (granted, in a much more efficient way), in a similar way as a smartphone camera is still a smartphone camera, regardless of whether it has 8 or 16 megapixels. Since it is agreed upon that human 'surveillance'[159] or mindreading pose no privacy issues, to be considered a new privacy threat, privacy scholars would need to show that these technologies are something at least equivalent to radical innovations, meaning that they fundamentally depart from human practices. The problem is that, as they currently stand, they embody incremental innovations.

However, a long-standing lesson from technological history – whether in industrial automation, biotechnologies, or information systems – is that incremental innovations can also, through small, cumulative modifications, transform the very nature of a system once they pass a critical threshold (Henderson & Clark 1990; Christensen 1997; Garcia & Calantone 2002; Dosi 1982). This is what Henderson and Clark (1990) call "architectural innovation" and what more recent work describes as the emergence of new technological paradigms from the cumulation of incremental changes (Pedota et al. 2021; Berggren 2019; Hacklin, Raurich, & Marxt 2004). Once such a critical threshold is passed, what initially seemed like mere optimisations reorganise the system's functional architecture and its purpose. The shift from mobile phones to smartphones exemplifies this dynamic. When mobile phones replaced landlines, they retained the core practice of voice communication, following a traditional path of incremental changes. When smartphones replaced mobile phones, however, they redefined the device as a general-purpose, networked computing platform in which voice calls are no longer central to the core practice. While it is sometimes difficult to pinpoint the exact moment at which this conceptual discontinuity occurs in the evolution of a given technology, the idea is that the accumulative improvements ultimately produce a different technological kind, with new roles, expectations, and practices.[160]

Atoosa Kasirzadeh (2025) identifies a parallel in AI ethics, differentiating between accumulative risks and decisive risks. Accumulative risks are the equivalent of incremental changes reaching a threshold, where the core concepts are overturned, and decisive risks are the equivalent of radical changes (i.e. general AI). Hence, she notes that alongside the risks brought about by decisive breakthroughs in AI are also accumulative changes that slowly and pervasively alter the

[159] As Goold points out (2008), it seems that some cases, such as a peeping Tom or a malicious stalker, should be considered as privacy-threatening. But this lies beyond the argument I am making vis-à-vis technology-based vs human-based recognition.

[160] This threshold is also the one on which Nissenbaum relies for determining prima facie privacy violations. (see Nissenbaum 2010).

world. At first, similarly to incremental changes, such accumulative changes seem continuous with what preceded them. Yet, at a certain point, the accumulation of refinements produces a qualitative shift – a reorganisation of the system's or practice's core logic, reaching a point where the underlying practice is ultimately transformed.

The key element in what innovation theory calls "architectural innovation" or "new technological paradigms", and what Kasirzadeh identifies as "accumulative risks" is that if we accept the central idea, Ryberg's methodology of presenting technologies as simple improvements of human capacities leaves his account blind to that specific kind of pervasive incremental changes.[161] If similar changes in emotion recognition do overturn the core concepts grounding the human practice, then we have grounds for assessing them as new (technological) privacy threats in the technology-based practice.

Concerning privacy and privacy rights, I suggest that what others have attempted to point out with the use of ubiquitous CCTV cameras and anticipate with neurotechnologies is similar to what I am doing with fully functional ERTs. I am arguing that fully functional ERTs constitute a new kind of privacy threat because they have passed a decisive turning point, becoming an inherently different form of emotion recognition and posing a distinct normative kind of privacy risk from human emotion recognition. As both innovation theorists and AI researchers note, however, a central difficulty with such transformations is that, in most cases, it is almost impossible to identify the precise moment at which the technology crosses the relevant threshold and the core practice is recognised as having been overturned. The example of CCTV cameras shows clearly this challenge. It is almost impossible to pinpoint when, exactly, the practice of surveillance through CCTV cameras has stopped being simply what could be considered an extension of human surveillance and reached the point where it created a new normative paradigm of ubiquitous surveillance, but it seems fair to say that we have, somewhere down the line, crossed that threshold. This change of paradigm was pervasive and pernicious, leaving most people, including policymakers, under high uncertainty regarding the many different new ethical issues it created (including privacy issues). In the case of neurotechnologies and fully functional ERTs, what privacy scholars are attempting is to anticipate these ethical issues before they are actualised.

[161] In innovation theory, however, architectural change is not always seen as negative, and is often even seen as a positive kind of breakthrough.

Hence, I assume that fully functional ERTs have, by definition, already reached a tipping point where the technology overturns the core concept of the human practice of emotion recognition. Under fully functional ERTs, emotional legibility is no longer situated, defeasible, and negotiated, but systematic, transparent, and unilateral. At that point, the technology ceases to be a better version of human recognition and instead becomes a different practice altogether. The underlying concept of emotional legibility changes, and with it the normative space in which emotional privacy operates.

From the perspective of emotional privacy then, the difference between human- and technology-based recognition is not merely a difference in degree but in kind. Human recognition, constrained by perceptual limits and social norms, leaves room for opacity, ambiguity, and strategic performance. It presupposes and preserves expressive discretion. Fully functional ERTs, by design, threaten to introduce a condition of systematic exposure, in which inner affective states are made legible regardless of individuals' use of expressive discretion. They enact what I refer to as the defeat of expressive discretion: the disabling of a person's ability to meaningfully manage the legibility of their EEs.[162] Importantly, by 'design', I mean that this is not because an ERT bypasses expressive discretion on an x-number of occasions that at some point the ERT's legibility becomes problematic. Rather, the background conditions of affective opacity begin to erode from the first instance in which technological recognition is made reliably and continuously available. The wrong, in such cases, does not lie in a single instance of exposure, or in the accumulation of such instances of exposure, but *in the structural disruption of the social and pragmatic functions of emotional expressions*.

Once we understand the conceptual and structural differences between human and technological recognition, it becomes clear that fully functional ERTs cannot be treated as a mere extension of ordinary human emotion recognition, as Ryberg's earlier critique of CCTV cameras and neurotechnologies presumably would assume, because unlike human recognition, which is an essential part of social life, ERTs reconfigure the practice of emotional legibility itself. Ordinary human recognition and technologically mediated recognition belong to different normative kinds, and only the latter threatens the affective infrastructure on which self-presentation, role plurality, and social regulation currently rely. This is what the normative account of emotional privacy is conceptually designed to shed light on. Emotional privacy,

[162] As mentioned earlier, this does not entail the successful control of EEs, but only the ability to exercise expressive discretion.

understood as a normative dimension concerned with the preservation of expressive discretion and managed affective opacity, is therefore directly threatened by fully functional ERTs because their use defeat individuals' expressive discretion.

Now that I have explained the inherent difference between human and technology-based emotion recognition, I use three technomoral scenarios to analyse the extent to which the use of fully functional ERTs may threaten the interests at stake in emotional privacy.

6.3. Illustrating emotional privacy losses through technomoral scenarios

EEs are essential elements of most of our interactive practices. Nonetheless, they are informative only conditional to the perceiver's successful interpretation of them, which is itself conditional to the perceiver's access to relevant epistemic inputs (e.g. background knowledge, context, prevailing norms).[163] Hence, ordinarily, social life proceeds under a degree of affective opacity. Emotional states and social motives are not fully legible by default and an individual's willingness to share raises or lowers the probability of their own affective legibility. The privacy mechanisms that one deploys to this end constrain the inferences and predictions others can draw from one's EEs and enable individuals to manage the strategic ambiguity necessary to navigate social life. Humans heavily rely on this expressive ability to maintain self-presentation, manage role plurality, and collectively help to regulate social life, ensuring social cohesion.

This affective infrastructure thus sustains social life across multiple levels of interaction: individual, interpersonal, and societal (see Chapter 5). Fully functional ERTs threaten to dismantle this affective infrastructure by collapsing the fragile equilibrium between social norms of expressiveness, what we choose to display, what we withhold, and how others interpret our EEs, given also possible affective leaks. By automating, accelerating, and amplifying the process of affective inference, they drastically reduce the opacity that normally obtains in affective communication channels, defeating our expressive ability. When privacy mechanisms, and, correlatively, interpretive labour, and perceptual limits no longer operate as constraints, the emotional lives of others become, in principle, continuously and effortlessly legible. What is threatened is not only "emotional data", or the sanctity of an "inner" life, but the *functional architecture of emotional life as a social practice*. Through the anticipation of fully functional ERTs, it is possible to expose the moral cost of this collapse.

[163] See Chapter 4.

In this section, I first lay out three technomoral scenarios in which a (standing) emotional privacy loss has happened due to the defeat of expressive ability by an ERT. I use these three scenarios to track the harms back, exposing the threat to our interests in self-presentation, role plurality and conformity to role-bound norms of expressiveness, and the social regulation and cohesion that stabilise collective life.

To do so and to be in a position to focus on emotional privacy violations, each scenario treats an idealised, fully functional ERT.[164] As a reminder, this means that the ERT is perfectly accurate, universally applicable across cultures and individuals, and capable of real-time interpretation of emotional expressions. It does not suffer from the technical limitations, bias, or contextual insensitivity that preoccupy current critics. Its inferences are flawless, context-aware, and seamlessly integrated into various social domains such as workplaces, education, healthcare, public spaces, interpersonal communication, and law enforcement.

At first glance, such a technology might appear ethically unproblematic. It produces no false positives or negatives, avoids misinterpretation, and could be deployed for seemingly beneficial purposes such as enhancing empathy, improving mental health support, or even flagging violent individuals before they commit any aggression. Yet it is precisely in this perfected form that a fully functional ERT sheds light on the moral significance of emotional privacy. The moral wrong lies not in what ERT misrepresents (or might misrepresent), but in what it renders impossible, namely, the maintenance of the functional architecture of social life sustained by emotional expressiveness.

As is often the case with social matters, the levels of interactions discussed so far, namely individual, interpersonal, and societal, are not completely impermeable from one another, and sometimes overlap. I aim thus not at attempting to show how ERTs may disrupt specific social functions of EEs while leaving the others untouched, but rather how specific functions of EEs, mainly from intentional EEs, at least at these three levels, are severely disrupted by the defeat of people's privacy mechanisms, leading to harms that affect both individuals' and society's interests.

The three scenarios follow the three levels of interaction presented in Chapter 5. Moreover, they all have a *distinct social setting* in which the ERTs are "owned" or at least in the possession

[164] See Chapter 2 for a detailed definition of fully functional ERTs.

of different types of individuals. Scenario I is designed to analyse issues related to ERTs that are "owned" or in the possession of the data subject. Scenario II is designed to grasp issues related to ERTs that are "owned" or in the possession of someone in a position of authority over the data subject. And scenario III is designed to grasp issues related to situations where ERTs are already a widespread and ubiquitous technology, possessed and used by everyone, similar to what phone cameras are nowadays. I first overview each technomoral scenario by explaining what would have happened in that particular situation if there had been no ERT involved. Then, I pinpoint where and in what the loss of emotional privacy consists, the harms that follow from this loss, and whether this loss appears to be wrongful or not, leaving the final verdict for the next section. While reading the scenarios, some might think that they beg a question on the role of consent. The technomoral scenarios are designed to clarify what is structurally at stake; please bear with me as I will keep the discussion around whether and to what extent consent can legitimise the use if ERTs in subsection 6.5.3.

6.3.1. Scenario I: Alfred's self-monitoring

Alfred wears affect-sensing glasses and watch to better understand and manage his stress. At a small dinner with close friends, the system detects a sustained pattern of physiological and behavioural markers for sexual desire when his friend's spouse speaks; on his phone, a push notification reads "arousal spike (87%)". Alfred freezes. He also realises that two other guests have seen the notification.

Ordinarily, Alfred has the ability to, and exercises his ability to, deploy privacy mechanisms, ranging from direct concealment to selective disclosure. Even though his attempts are not always a success, he usually manages to meaningfully regulate the affective opacity in the interactions he sustains and, if at all, turns felt states or social motives into emotional expressions. Even though regular leaks are noticed, they rarely contradict his deliberate expressions in a way that would make the latter useless. The device changes this social structure. The protected space in which conflicted affects can be worked through, without becoming socially actionable, disappears.

In the technomoral scenario, Alfred loses actual emotional privacy over his sexual desire for his friend's spouse.[165] By exposing his desire, the ERT dismantles the affective scaffolding that supported the dinner conversation between friends. The dinner conversation moves from negotiated meaning to device-mediated exposure.[166] In such a situation, the exposure of an

[165] As mentioned above, I will discuss the question of consent in §6.5.3.
[166] Note that, on the perceivers' end, the non-acknowledgement ability is also diminished, as a formal exposure like this one makes it harder to pretend to "not notice".

inappropriate sexual desire would likely prompt others to question Alfred, to seek explanations, or to reassess their view of him, even though he has done nothing wrongful. The privacy-preserving pathways of reflection, retrospection, and growth, that were ordinarily within Alfred's reach, are short-circuited. On the account of emotional privacy developed here, Alfred therefore loses not only actual but also standing emotional privacy. The loss is wrongful not because of the content of the desire itself, but because it results from the defeat of his expressive ability.

6.3.2. Scenario II: Marie's hospital policy

Marie is a senior surgeon. She is also in the middle of a bitter custody dispute and has sought counselling to stay on top of things. After another rough morning, she walks onto the surgery floor, happy to finally be able to clear her head at work by focusing on the procedures she has to perform. She is handed a small electronic pad that links to the hospital's affect monitors to check fitness to operate. Before she reaches the first patient file, the screen turns on and reads "Not emotionally ready to proceed." Her authority to sign pre-op orders is temporarily locked and a prompt invites her to a short well-being debrief. Nurses and doctors look up, surprised. Marie is known across the department for a cool-headed temperament and almost irreproachable composure. The coordinator follows protocol and books the debrief. Marie had planned to keep the morning's dispute out of the hospital as she does not like to mix personal and professional life, but she has no other choice now, especially since the affect monitor was installed following due process.

Ordinarily, in the surgery unit, Marie sustains her professional and personal boundaries through three privacy mechanisms. Through direct concealment, she attempts to keep her inner affective life tied to personal matters away from hospital chatter. Through indirect concealment, she maintains the convention-fit calm expected of a senior surgeon, preserving both professional relationships with her colleagues and respect for her authority.

The hospital's ERT collapses this arrangement by stitching together her current two main social roles– surgeon and soon-to-be divorcee – and converting what she meant to keep opaque into a matter of professional fittingness. Direct concealment is bypassed because her effort to maintain neutral behaviours is publicly contradicted by the ERT's alert. Indirect concealment is dissolved because the system announces the very tension that her convention-fit behaviour was holding in place. Selective disclosure is pre-empted because the protocol forces her into a mandatory debrief with the coordinator. Here, too, standing emotional privacy is lost. Marie's capacity to maintain a composed professional demeanour despite experiencing a complex mix of emotions, is defeated. What is undermined is not merely the concealment of a particular feeling, but her capacity to regulate how her affective state becomes socially legible within the practice.

The moral harm, however, extends beyond the individual level. It compromises her ability to present the self she intends to be in this context and damages her ability to manage her distinct social roles and to conform, as she intended to, to the social norms of conduct of a senior surgeon. As a consequence, colleagues recalibrate their own portrait of her professional self to the device output rather than to her seniority, record, and typical expressions. Her calm is re-read as inauthentic, and her practical authority is reduced. In short, the ERT defeats the affective scaffolding that ordinarily regulates the workplace by lowering informational opacity and flagging her inner affective states as above the acceptable threshold.

6.3.3. Scenario III: The minister's public apology

A teenager is killed in a late-night knife attack. The Interior Minister schedules the typical process in such case: a closed briefing with the victim's family, a short televised address to acknowledge the harm and apologise, then a written plan for policing and prevention released the next morning. The minister begins her public apology, but this time, people use their app-based ERTs on their phone to fact-check her emotional expressions. The minister's emotional score reads "remorse (10%)". Commentators pivot to the scores rather than the reasons she is giving. The apology has not done its regulatory work as the minister's verbal apology is overridden by her emotional score, drastically altering the population's interpretation of the whole scene.

Ordinarily, the minister's address is designed to perform two intertwined social functions. It attempts to regulate society's frustration by acknowledging the breach, explaining the next steps, and anchoring a proportionate path forward. It also attempts to sustain cohesion by offering a shared moment to recognise the loss and offer reassurance. These social functions rely at least partly on selective disclosure and concealment. The televised apology is curated. Reasons are given in an order that maintains role distance and tempers any potential community outrage, before sanctions are considered.

This time, the phone-based ERT prevent the success of the typical process and collapses the minister's attempt at regulating collective outrage and ensuring the continuation of social cohesion. By rendering the minister's affects continuously legible to everyone watching in real-time, the system defeats the minister's ability to rely on privacy mechanisms to perform her emotional expressions for her public. The score appears as part of the message and the typical opacity that lets the apology play its regulative role disappears. Hence, there is loss of emotional privacy. Selective disclosure is pre-empted and the apologetic expression, that ordinarily support the verbal apology, ceases to function as a regulative practice that guides collective response. The main problem is that the new state of legibility defeats the privacy mechanisms that make

public emotional expressions reliable sources of regulation and cohesion. That shift marks a loss of regulatory traction and a thinning of social cohesion, exposing the societal-level harms caused by a defeat of emotional privacy. In this case, there seems to be a violation of the moral right to emotional privacy.

Importantly, these scenarios are not intended to suggest that emotional privacy ought to shield individuals from legibility altogether. The protection of expressive discretion does not (and is not meant to) secure an absolute state of opacity. Emotional privacy is necessarily partial and remains porous to affective leakages, misfires, and unsolicited inferences that shape how one is perceived. Within these limits, however, emotional privacy preserves a zone of strategic ambiguity in which emotional expressions can perform their social functions. It is this zone that warrants protection against technologically mediated disruption. The technomoral scenarios aim to show how fully functional ERTs can undermine this zone in distinct ways, thereby jeopardising the work of intentional EEs within social practices.

6.4. Moral right to emotional privacy protects a distinct kind of interests

The concept of emotional privacy uncovers a set of distinct interests that are not adequately captured by existing (moral) rights of informational privacy. The moral right to emotional privacy thus aims to protect these new interests whose loss risks severely damaging the current structure of social life. These interests form a downstream normative trajectory. At the upstream level is expressive discretion, the ability to manage the legibility of one's emotional expressions within affective communication channels. When expressive discretion is defeated, individuals lose the ability to present the self they judge appropriate. This upstream loss then cascades. First, it undermines self-presentation, then, it collapses role plurality and the capacity to meet the expressive expectations tied to specific social roles, and finally, it erodes the regulative and cohesive functions that emotional expressions ordinarily perform in sustaining social coordination and collective life.

The three technomoral scenarios presented above each focus on a particular stage of this progression. Alfred's scenario shows how technologically induced legibility defeated the capacity for self-presentation. Marie's scenario illustrates how this legibility collapsed role plurality and conformity with role-bound norms of expressiveness by defeating her efforts to keep personal matters away from her work. Finally, the systematic exposure of the Minister's affective states disrupts the regulative and cohesive functions of emotional expression in public

life. In each case, the defeat of expressive discretion triggers individual, interpersonal, and societal effects.

If these interests do indeed form a downstream trajectory, then a moral right designed to protect expressive discretion against defeat must also be understood as protecting the downstream interests embedded in affective communication channels. In this section, I therefore examine how the familiar variants of the moral right to informational privacy (Control-based accounts, Access-based accounts, and Contextual Integrity) fare in protecting affective information carried by emotional expressions. I argue that while affective information naturally belongs within the broader informational privacy cluster, the definition proposed by the mainstream normative accounts of informational privacy fails to secure the specific interests at stake. The conclusion is that none of them offers adequate protection to this particular set of interests, hence emotional privacy must be recognised as a moral right within the informational privacy cluster.

6.4.1. Variants of Control-based accounts

I now turn to the moral right to privacy as understood from the Control-based accounts already introduced in Chapter 5. The point I am trying to make here is *not* that these accounts *cannot* capture any privacy wrongs caused by ERTs, only that they are ill-equipped to grasp the wrongs that follow from the defeat of expressive discretion, therefore overlooking something of moral value. Hence emotional privacy is a complementary moral right.

The normative Control-based approaches (Classic Control, Negative Control, Source Control), ground the *moral right* to informational privacy in a person's authority vis-à-vis informational flows (Westin 1967; Mainz & Uhrenfeldt 2021; Menges 2021). Although the approaches disagree on what the authority is over and how it is exercised, they usually agree on the core idea that privacy wrongs stem from *failures of rightful control.*

On the Classic *normative* Control account, person A's privacy has been violated iff person A has involuntarily lost control over access to personal information p about person A.[167] The main problem with this formulation is that it includes ordinary affective leaks as "involuntary loss of control", which renders any (semi-)involuntary EEs (e.g. natural EEs, semi-natural EEs, and

[167] I took this normative definition of the 'classic' control-based approach from Mainz and Uhrenfeldt (2021: 289), specifically their CA1 definition, but I have removed the "unwanted" clause to escape issues over the subjective evaluation of what is considered "unwanted access".

affective leaks) a wrongful loss of an individual's informational privacy. This is highly problematic for obvious reasons, namely that people would have their informational privacy constantly violated, meaning that there would also be no difference between ordinary losses and wrongful losses.

On the last *normative* form of the negative control account, as proposed by Mainz and Uhrenfeldt (2021), which they characterise as the strongest one, person A's right to privacy is violated iff person A has involuntarily lost negative control over the access to (personal) information p about person A, due to action(s) of person B, of which B is responsible (2021: 298). Recall from Chapter 5 that by "negative control" the authors mean "the ability to prevent".

This formulation captures clear cases in which a person is subjected to access they cannot block, for example coercive interrogation or unauthorised surveillance (Mainz & Uhrenfeldt 2021). However, it does not translate well to cases involving emotional expressions. The central difficulty is that the account assumes a discrete and attributable act of access by a specific B whose responsible action causes A's involuntary loss of control. Emotional expressions do not function in this way in ordinary social life. People routinely perceive and interpret the emotional expressions of others, and while this does result in a loss of negative control, B's behaviour consists merely of ordinary looking, listening, or attending. It is a normal feature of interpersonal engagement rather than an attempt to invade someone's privacy. Without any technological intervention, Alfred's friends may notice his attraction at dinner, Marie's colleagues may see that she is tense, and attendees present at the Minister's apology may form judgements about her sincerity. In each of these situations, someone accesses information about another person and the loss of control on the part of the person observed is involuntary, yet none of these instances are plausibly classified as wrongful privacy violations. The negative-control account therefore risks treating routine and socially appropriate forms of affective uptake as violations simply because they involve an involuntary loss of control caused by another's ordinary actions.

The account also encounters difficulties when applied to technologically mediated emotional legibility. In environments where fully functional ERTs are continuously operating, as in Marie's hospital scenario, emotional legibility results from the design of the environment rather than from any single agent's deliberate action. If we insist on identifying responsible agents in such cases, responsibility becomes so widely dispersed across engineers, managers, system operators, and others that the notion loses any meaningful grip. The alternative is to conclude that no

violation occurs because no single act of access can be attributed to a specific B. Yet the harm in Marie's case clearly does not stem from any specific observer B actively acquiring more information than she intends to reveal. It stems from the technological configuration of the environment itself. Once the system is installed, her expressive discretion is undermined regardless of what any individual colleague or supervisor does. A framework that recognises only discrete and agent-driven acts as privacy-relevant cannot account for this structural defeat of expressive discretion.

The result is that emotional expressions generate two types of losses of negative control. The first is benign and unavoidable in ordinary interaction. The second is wrongful and arises when emotional legibility is engineered into the environment through technological means. On Uhrenfeldt and Mainz's account, a person's privacy is violated when they involuntarily lose negative control over access to their personal information due to actions attributable to another agent (Uhrenfeldt & Mainz 2021). Framed in this way, the account appears both overinclusive and insufficiently precise regarding the scope of wrongful privacy violations. It risks labelling ordinary interpersonal perception as a privacy violation, while also failing to register violations in technologically-mediated contexts where emotional legibility is structurally imposed. For these reasons, the negative-control account cannot capture the practice-disrupting defeat of expressive discretion brought about by fully functional ERTs and cannot distinguish such cases from the ordinary losses that structure everyday life. It therefore cannot absorb the normative work that the moral right to emotional privacy must perform.

Finally, Menges' source-control account is, as I understand it, developed into a normative account called the Deep Self View (DSV). On the descriptive account of source control, Menges holds that privacy obtains when person A is the *right kind* of source for information p's flow (if information P flows at all). Building on this descriptive foundation, Menges formulates DSV as a normative refinement of the control tradition by grounding the moral right to privacy in the protection of a person's deep self as the legitimate moral source of information about them. On the normative account, he thus holds that:

> A has a right to privacy regarding personal information p just in case others have the pro tanto obligation directed at A to not make p flow unless (i) this flow is an expression of and (ii) does not conflict with A's deep self. As long as A has neither waived nor forfeited the right to privacy, B infringes A's right to privacy regarding p just in case B makes p flow even though it is not an expression of or conflicts with A's deep self (Menges 2024: 7).

The "deep self" can be summarised as a person's "set of [relatively] stable intrinsic desires" (Menges 2024: 9).[168] Examples of such intrinsic desires include wanting one's children to flourish or wanting to leave future generations a world worth living. According to the DSV, information p should fall under A's moral right to informational privacy when, in a given context, it is instrumentally, representatively, or symbolically valuable that A's deep self be the source of any associated informational flow. The moral right therefore imposes on others a pro tanto obligation not to initiate informational flows about A unless doing so expresses, and does not conflict with, A's deep self. The decisive normative question is not whether A voluntarily chooses the flow, but whether the flow properly originates from A's genuine and relatively stable intrinsic commitments.

Although the view is internally coherent, certain tensions arise. First, Menges holds that "one does not need to be aware of one's intrinsic desires" (Menges 2024: 9) but also requires that others have a *pro tanto obligation* not to make information about A flow unless it expresses (and does not conflict) with A's deep self. While Menges' account does not demand omniscience but only a reasonable-belief standard on person B, this does not entirely resolve the tension. Given this epistemic opacity, it is unclear how others can reliably fulfil the pro tanto obligation not to make information flow unless it expresses or aligns with A's deep self. Even with a reasonable-belief standard and a default presumption against flow, if we accept that seemingly banal facts, such as whether one has two hands,[169] can possess instrumental, representative, or symbolic significance in particular contexts, the DSV places a considerable practical burden on others. They must determine, often immediately and without proper contextual insights, whether a given informational flow would appropriately express A's deep self or instead conflict with it. This problem becomes even more acute in large-scale or multi-agent technological environments, for example those involving ERTs in classrooms, hospitals, or airports. In such settings, it is effectively impossible to assess, person by person, whether a specific informational flow would or would not originate in the individual's deep self.

But the central difficulty for DSV in relation to emotional privacy is that it evaluates the legitimacy of the flow of information through the lens of authenticity and alignment with stable

[168] I will not go into the details of the notion of deep self here, but note that these stable desires constitute, in a sense, their sense of agency; "the agent's ability to write their own life stories" (Menges 2024: 16). See also Menges (2024: 9-11) for a detailed explanation of the deep self.

[169] Let's say that I know person A has two hands, and a couple days later our country declares mandatory military service for everyone who is apt to fight, which implies that the people need to have two hands, this suddenly becomes highly relevant for A's deep self.

intrinsic desires, whereas emotional privacy evaluates the preservation of expressive discretion within affective communication channels. As a result, the two frameworks do not track the same kinds of harm.

The DSV view thus allows for a flow to align with one's deep self and yet undermine one's ability for expressive discretion. The DSV can judge an informational flow to be legitimate even when that flow undermines the individual's ability to manage the opacity and selectivity of their emotional expressions. Marie's case illustrates this problem. Suppose that part of Marie's deep self, as a committed surgeon, includes the intrinsic desire to safeguard patient well-being and to maintain high professional standards. If so, the informational flow produced by the hospital's ERT system can plausibly be interpreted as expressing or at least not conflicting with Marie's deep self. Under the DSV, this flow would not constitute a privacy infringement. Yet emotional privacy is clearly violated in Marie's case. Her expressive discretion is defeated by the system's continuous extraction of affective information, which prevents her from managing her professional self-presentation in context. The DSV therefore fails to protect her ability for expressive discretion because in this case it does not directly correlate with her deep self but instead concerns the functional and relational significance of emotional expressions in everyday interactions.

In other words, the DSV operates at the level of the individual's deep, relatively stable commitments, while emotional privacy operates at the level of social practices in which expressions serve regulative and coordinative functions. The DSV therefore overlooks something of independent moral importance, namely the preservation of expressive discretion in affective communication channels. Consequently, the harms that emotional privacy seeks to protect against are not adequately captured by the DSV.

6.4.2. Variants of Access-based accounts

The normative Access-based approaches (Classic Access, Restricted Access, Actual Access), ground the *moral right* to informational privacy in a person's information accessibility given some conditions (Gavison 1980; Allen 1988; Macnish 2018). Despite their internal differences, all normative access-based accounts of privacy share a structural commitment. What is protected, in each case, is some informational p about person A, and correlatively, what is condemned is *an act of access to that information under the wrong conditions*. Whether they focus on unjustified

intrusion (Gavison 1980), contextually inappropriate accessibility (Allen 1988), or unauthorised epistemic access (Macnish 2018), these accounts locate privacy's moral content in the act or condition of access across an informational boundary.

On the normative Classic Access-type accounts, following Gavison (1980) and Reiman (1976), person A's informational privacy is violated iff A is subjected to unwanted access to information p by some person B. The central idea is that privacy protects individuals against being subjected to access they do not want. Applied to paradigmatic cases, this account-type can explain the wrong of intrusive surveillance when A never had any meaningful control over the situation in the first place, for instance when B observes A through a hidden camera or wiretaps a telephone line. In these cases, it seems clear that A's privacy is violated because B exposes A to a sustained unwanted access.

However, once we turn to emotional expressions, the strengths of the Classic Access account become a source of difficulty. Applied to the technomoral scenarios, this account faces an immediate problem. Unwanted access to affective information is part of ordinary social life (and even important to it). If the classic access account treats all such unwanted access as a violation, it collapses benign and even valuable forms of emotional perception into the category of privacy harms.

In technology-free versions of the scenarios for instance, Alfred may not want his friends to notice his attraction, yet their uptake of his demeanour is clearly part of the normal dynamics of friendship and social attunement. Likewise, Marie's colleagues may notice that she is unusually tense, and in the Minister's case, members of the public will inevitably form views about her sincerity. In each situation, someone accesses affective information that the person would rather not share, and yet these instances of unwanted access do not seem to be wrongful privacy intrusions. They are ordinary losses of emotional privacy that are woven into the fabric of interpersonal and public life. Hence, the Classic Access account cannot respond to Ryberg's challenge because it does not have the tools to differentiate ordinary mindreading from ERT-based recognition. What emotional privacy highlights is that the difference between ordinary mindreading and ERT-mediated access does not lie only in the presence or absence of an unwanted act of access. It lies in the way ERTs structurally defeat expressive discretion by making affective states continuously and determinately legible, and by transforming the practices

within which emotional expressions operate. For this reason, the Classic Access account cannot absorb the moral role of emotional privacy.

On Restricted Access-type accounts, person A's informational privacy is violated iff A is subjected to unauthorised access to information p by some person B. In this tradition, inspired by Allen's (1988) and by Tavani and Moor's Restricted Access account (2001), the key question is not simply whether access is wanted, but whether it is normatively permitted in light of social and institutional rules (Allen 1988; Tavani and Moor 2001). What matters is not whether A wants access to occur, but whether B is authorised (institutionally, socially, or even morally) to acquire the relevant information. Privacy is therefore understood as a system of restricted access zones: certain agents may access certain information only within the limits defined by social roles, institutional norms, or moral duties. In this respect, the Restricted Access account has an important advantage over the classic access view. Whereas the classic view relies on unwantedness and thereby risks treating any unpreferred perceptual uptake as a privacy violation, the Restricted access view replaces subjective attitudes with objective permissions. This allows it to distinguish between unwanted but permissible access, such as a colleague noticing fatigue, and unwanted but impermissible access, such as an employer secretly reading private emails. It also explains why some forms of surveillance constitute privacy violations even when individuals have little to no control (against weaknesses of the Control-based accounts) and even when they might not object in a subjective sense. By grounding privacy in socially recognised boundaries rather than personal preference, the Restricted Access framework is better equipped to analyse privacy in institutional contexts where entitlement, not desire, determines who may know what about whom.

However, once applied to emotional expressions and fully functional ERTs, this advantage becomes insufficient. Consider Marie's case. In the technology-free version, some colleagues might notice that she seems unusually tense before surgery. Their uptake of this information is not only socially authorised but also expected, since mutual attunement is part of professional collaboration. On the Restricted Access account, no privacy violation occurs, and this seems correct. In the ERT-mediated version, the hospital installs a system that continuously analyses Marie's EEs and outputs a classification that she is not fit to work. However, since the system operates within proper institutional authorisation, the Restricted Access accounts register no wrongful loss of privacy, even though Marie's expressive discretion is defeated in a way that deeply affects her professional and personal roles. Emotional privacy, by contrast, identifies this

as a central harm. The mere fact that the hospital is authorised to monitor her EEs does not mitigate the architectural change in the conditions under which her emotional expressions function nor the defeat of her expressive ability.[170]

A third variant within the access-based tradition strengthens the definition of a privacy violation by requiring not merely unwanted or restricted possibility of access but Actual (epistemic) Access (see Macnish 2018). On this view, person A's informational privacy is violated iff A is subjected to actual access to information p by some person B. This account narrows the scope of privacy violations, since mere opportunity for access or the mere increase in accessibility does not suffice. There must be an identifiable moment in which B comes to possess, apprehend, or register the information in question (Macnish 2018). The strength of this account lies in its ability to avoid treating only hypothetical or merely potential access as wrongful. If surveillance cameras are installed but no one watches the footage, or if a database is created but never queried, the mere possibility of knowing is not automatically treated as a privacy harm (Macnish 2018). This makes the Actual Access account stricter in its violation-threshold than both the classic and restricted variants, especially in situations where informational risk is high due to the automatic collection of information, but no knowledge has yet been formed because no agent has accessed it yet.[171]

Applied to emotional expressions, however, if actual access means *genuinely* acquiring information p about A, it seems that the Actual Access account gives us the opposite result than the one we are looking for intuitively. Because emotional expressions are inherently public-facing and partially involuntary forms of behaviour, and ordinary uptake of them is not only unproblematic but essential to social life, B acquires information p about A in ordinary settings, all the time, subjecting A to actual access over their information p. Let's call this one interpretation 1. If we do treat human-based inference as actual access, then ordinary perceivers in the technomoral scenarios often count as having actual access, and the view would then condemn human-based uptake over the practice-disrupting effects of ERTs. Yet it would be implausible to treat these episodes as privacy violations.

[170] But as I will discuss in the next section, this authorisation may be used to warrant the use of ERTs against the pro tanto reasons in favour of emotional privacy.
[171] It has also been used in neurorights debates to distinguish between technologies that "actually access" inner mental content and ordinary interpersonal inference, suggesting that some forms of technologically mediated knowledge acquisition are categorically different from ordinary observation (Yuste et al. 2017; Lavazza 2018; Ienca 2021).

One might reply that Actual Access is best understood as a strong kind of epistemic penetration that ordinary perceivers lack. Let's call this one interpretation 2. On this interpretation, though, then both ERTs and humans would never "actually" access emotional states, because the Access view already discard technological uptake from "actual", human agent, uptake. This results in no privacy violations ever, which also seems implausible. On another interpretation, we might say that Actual Access is only happening when human agents actually access ERTs' output (interpretation 3). Let's consider this interpretation in the light of the technomoral scenarios. In Marie's hospital, the ERT system continuously analyses behavioural and physiological changes, storing affective classifications even if no supervisor personally looks at the data. Marie's expressive discretion is already defeated at this stage, because her emotional life is rendered structurally legible in a way she cannot meaningfully regulate. Yet no discrete moment of actual access by a human observer need occur for the harms to happen. When the "Not emotionally ready to proceed" notification popped up on the screen, people around Marie did not have access to the 'actual' state of her inner emotions[172] and yet this was sufficient to defeat her expressive ability. This means that if we adopt the last interpretation of the Actual Access view, many privacy violations are overlooked and the harms resulting from those violations are left unflagged.

The Actual Access account therefore ties privacy violations to the instant at which information is obtained (no matter how we conceive the conditions that needs to be met for this obtention to realise), whereas emotional privacy concerns the degradation of the background conditions that make emotional expressions meaningful, interpretable, and socially functional. Ordinary mindreading in the technomoral scenarios involves actual access to some EEs (If we accept the first interpretation of the Actual Access view) but does not undermine these conditions. ERT-mediated environments defeat expressive discretion even if no individual ever accessed the information produced by the system. Hence, regardless of the interpretation we use for understanding the Actual Access view, it cannot systematically identify practice-level transformations in affective communication channels where the wrong (viz. the defeat of expressive ability) arises prior to, and independently of, any discrete moment of actual epistemic access. It either takes every, ordinary, human uptake as a privacy violation (interpretation 1), no human or technology-based uptake as privacy violation (interpretation 2), or only human uptake of a technology-based output (interpretation 3), where the latter results in overseeing privacy

[172] For the sake of this argument, let's hypothesise that such thing could have been possible if they had taken a look at the fine-grained data provided by the system rather than just the final decision based on those results.

violations that arise despite this condition being met. What emotional privacy highlights as harmful is not the moment at which a person actually acquires information about another person. It is the architectural transformation of the communicative environment that makes expressive discretion impossible, independent of any actual (epistemic) uptake of a human on a technology-based output.

6.4.3. Contextual Integrity

Helen Nissenbaum's framework of Contextual Integrity (CI) (2010, 2019) offers one of the most influential normative approaches to privacy. Unlike control- or access-based theories, CI grounds privacy in the *appropriateness* of a given information flow relative to a given context. Privacy is preserved when information flows conform to the prevailing entrenched informational norms of a context, specified by roles (sender/data subject/recipient), attributes (information types), and transmission principles (e.g. consent, confidentiality, reciprocity).

A deviation from those contextual norms is first flagged as a *prima facie* violation of privacy and subsequently confirmed (or refuted) as *genuine* once a moral assessment has deemed that the new flow does *not* uphold the context values, ends, and purposes in an equivalent or better way than the entrenched flow. CI's strength lies in its ability to connect privacy with social life, recognising that informational norms are grounded in the values and functions of distinct social contexts such as medicine, education, or politics.

This makes CI exceptionally well suited for evaluating technological practices that alter social life. The framework is therefore capable, at least in principle, of recognising new informational norms that arise in response to changing conditions in the flow of information.

Yet when the time comes to use the CI heuristic on new and emergent technologies, many challenges arise. First, scholars have pointed out its inherent conservatism (Rule 2019; van de Poel 2022). The inherent conservatism of CI makes it possible to justify problematic technologies when they follow established norms (Austin 2003; van de Poel 2022). The case of Marie is a paradigmatic example of this type of situations, where the ERT has *already* been incorporated into hospital's protocols, ensuring patient-safety and professional excellence. Marie's case assumes that the ERT has already been incorporated into hospital's protocols. This matters for CI because institutional contexts are precisely those where entrenchment is strong.

This means that the norms about roles, attributes, and transmission principles are formally codified and reinforced by hierarchy, compliance systems, and professional standards. Once an informational practice becomes part of the institutional fabric, CI risks classifying it as already appropriate, on the grounds that the relevant flows now match the prevailing norms.[173] By grounding its analysis in entrenched norms, CI struggles to ratify harmful transformations once they have become "normal." In Marie's case, CI could conclude – if we follow the official norms strictly – that ERT-mediated affect assessments are contextually appropriate because they are authorised by the institution, justified by a mission of patient safety, and embedded in professional role expectations. But this verdict would miss the core harm. Marie's loss is not primarily tied to the context's values, goals, or ends, but about the structural defeat of *her* expressive discretion as a human, which undermines her ability to sustain the self that she sees best fit, but also that her role and authority expect.[174] The hospital's safety aim must be weighed against the fact that emotional discretion affords individuals like Marie the capacity to sustain a plurality of roles (mother, divorcee, surgeon) and to follow role-bound norms of expressiveness, such as those tied to her role as a surgeon (see §5.5.2).

Emotional privacy, by contrast, does not treat existing norms as the ultimate criterion. Instead, it evaluates whether the introduction of ERTs in a specific context undermines individuals' ability to control their affective information and whether the defeat of intentional EEs' functions disrupt the social practices and their own societal benefits.

In the case of Marie, by defeating the functions of intentional emotional expressions, the technology risks weakening social practices that rely on them. Emotional privacy is sensitive to disruptions caused within affective communication channels. Hence, it can explain why some use of ERTs are problematic even in contexts that appear formally acceptable. By contrast, CI's reliance on entrenched norms can normalise harmful practices once they become institutionalised.

Even when access, purpose, and authorisation are satisfied, the defeat of expressive discretion damages the practice itself through the undermining of individuals' expressive discretion. Focusing only on the context's internal structure, without an interest to the individuals' well-being, result in missing harms that affect the very practice under evaluation.

[173] See Austin (2003), on the argument of 24/7 CCTV-camera surveillance.
[174] See Gstrein & Beaulieu (2022) for a critique of CI as unable to protect individuals' rights.

Moreover, as pointed out by Rule (2019), CI also makes it hard to reach a normative diagnostic when, contrary to the hospital's case, the context is ill-defined, and/or partly merging with other context(s). This is the case of Alfred's situation, where the ERT seems to be both in the context of friendship, which would include everyone physically present at the table (or at least the ones that saw the push notification), and the context of self-care, which focuses only on Alfred's personal assessment via the use of the ERT. This merging of contexts, which could in some other cases be neatly divided, is in this case produced by the social disruption of the technology that is under investigation. This disruption causes, as I have argued elsewhere (Prégent 2025b), the context to undergo significant reshaping, which may also involve, alongside a severe collapse with other contexts, a re-hierarchisation of the prevailing norms of the new (post-disruption) context. But focusing only on the fact that oftentimes, contexts such as friendship or self-care are often ill-defined, we can see that in Alfred's case, there are no codified informational norms governing *affective legibility* among friends, and the values at stake are at best implicit rather than formally articulated such as in Marie's scenario. CI can flag that a self-care emotional inference ("arousal 87%") becoming visible in a friendship setting is a prima facie violation, since the flow deviates from the informal expectation of defeasible emotional interpretation. But when CI reaches its normative stage, it must determine whether the practice should evolve to accommodate this new flow. Here the framework lacks the resources to explain why this new degree of legibility is objectionable because CI ties its assessment to entrenched norms. In informal settings, such as friendship, where norms are weak, shifting, or tacit, CI has difficulty identifying the decisive contextual value that must be upheld and hence it is difficult to determine whether this new degree of legibility is upholding the context's integrity. As Rule (2019) and van de Poel (2022) warn, CI risks either normative drift, which would allow the practice to evolve with a harmful technology, or normative hesitancy, which would result in a final decision that lacks moral support.

Finally, the last challenge pertains to the fact that some contexts, even though in principle formally regulated, struggle with never-ending disagreements on the norms that should regulate the flow of information. This is what is happening in the case of the Minister, and ultimately, the question of whether and to what degree public authorities should have privacy when in office. Some hold that transparency maximises accountability; others hold that maintaining privacy or even secrecy is essential to public trust, political unity, and institutional dignity (Hood 2006; Rosen 2001). CI performs poorly in such cases because its normative tool depends on the

very contextual values that are under dispute. When the context's purpose is contested, CI can flag that a new flow deviates from entrenched norms, but it cannot decisively say whether those norms ought to change.

In the next section, I will propose a way to supplement the moral part of the CI framework with emotional privacy, when relevant, to help remedy some of its internal tensions regarding the assessment of ERTs.

In sum, while emotional privacy is largely compatible with the different variants of informational privacy, none of these variants can on their own cover what emotional privacy protects. Hence, in the context of ethical analysis concerning ERTs and technologies of this kind, the moral value of emotional privacy can be added to the already existing weight of privacy concerns. This is what I turn to now.

6.5. Assessing emotional privacy's moral weight in future societal trade-offs

There are two reasons to define the moral right to emotional privacy as an interest-based, pro tanto moral right. First, establishing it as a moral right secures its status. It ensures that its ethical and normative weight must be acknowledged in relevant analyses. Second, identifying it as a pro tanto right has practical implications. As with other moral rights in (liberal-democratic) societies, emotional privacy may be subject to limitations when it conflicts with other important rights or interests. This approach is consistent with the standard treatment of rights conflicts in modern liberal-democratic societies and with how privacy is usually conceived in normative analyses,[175] where limitations may be placed upon it, but any compromise must be reasoned, proportional, and accountable.

One of the aims of this chapter is to reconsider the weight of privacy in the inevitable trade-offs that liberal-democratic societies will have to make vis-à-vis the use of new technologies that profoundly reshape our social life. I thus here pre-empt techno-optimists doubts on the actual "gravity" of privacy violations compared to the potential uses and opportunities that these new technologies will bring to the table. In the particular case of ERTs, at first sight, the prospect of exposing people's emotional lives may appear beneficial. One might think, at least on a prima facie basis, that such transparency would foster (or force) honesty, eliminate the space for

[175] See Moore (2010: 99) and Allen (2015: 174).

manipulation, and effectively counteract in real-time the everyday strategies of deceit or concealment that often complicate human interactions. ERTs, in this techno-optimist view, would cut through the noise of social performance and expose the ugly but nonetheless authentic picture of how people truly feel. Such exposure could seem especially welcome in contexts where sincerity is deemed valuable – in politics, in the workplace, or even in intimate relationships. By piercing through emotional expressions' shield and exposing concealed emotions and social motives, ERTs might appear as a new way to eliminate ambiguity and ensure social order grounded in (forced) honesty and trust.

Yet this line of thought now not only has to dialogue with the "mainstream" privacy losses that are the (il)legitimate collection, storage, and commodification of individuals' (emotional) data, it also needs to demonstrate that ERTs' use is not going to disrupt society's current (affective) social equilibrium. Ethical assessments typically weigh competing values involved in using technological tools. Within the cluster of privacy, emotional privacy brings new reasons to favour privacy in those settings. Below I aim to show practically what the proposed concept and the following moral right to emotional privacy may bring to the overall value of privacy in relevant trade-offs. Importantly, in these trade-offs, ERTs' advocates will have to justify the use of the technology by weighing societal interests against another. This is because the interests served by emotional privacy are held by individuals but are also constitutive of society's well-being. Trade-offs involving ERTs therefore no longer oppose individuals' (privacy) interests to society's interests (e.g. security, efficiency, or community), instead, they require an explicit balancing between competing societal interests.

By extending privacy as not only an individual but a societal interest, and by carefully unfolding the various functions of privacy in sustaining essential practices of social life (and the consequences of a disruption of such practices), emotional privacy adds a significant weight to the more general value of privacy in our techno-social world. What is protected by emotional privacy and threatened by ERTs is nothing less than the affective infrastructure that supports the self and social life.

If the core interests at risk also include the ones inherent to social practices constitutive of a liberal-democratic society, then the evaluation of ERTs must ask: Does the deployment preserve enough selective opacity for expressions to fulfil their communicative and regulative roles? If not, there is, at minimum, a prima facie infringement of a moral right to emotional privacy.

Consequently, claimed benefits of such exposure must be weighed against the constitutive social goods at risk. Nevertheless, whether to uphold other values against (emotional) privacy or not is a decision that needs to be justified by taking into account the societal benefits currently obtained under emotional privacy.[176]

With the concept of emotional privacy and the proposed moral right to emotional privacy in view, I now test whether the weight of privacy changes (when relevant) vis-à-vis other values in anticipation of future societal trade-offs. To do so, I propose to see how emotional privacy may substantiate two different types of privacy assessments. First, I slightly modify CI to enable the CI heuristic to take into account new moral rights to privacy such as mental privacy and emotional privacy, and then investigate whether these rights may recalibrate privacy's overall weight against other important values such as security, safety, and efficiency. Second, I analyse how emotional privacy may supplement informational privacy's reasons in a classic ethical assessment of new and emergent technologies, by relying on the HLEG ALTAI ethical framework, which recommends relying on criteria such as prevention of harm, proportionality, subsidiarity, and respect for privacy and data governance.

6.5.1. Substantiating the CI moral part of the heuristic

Several scholars have already attempted to modify the CI heuristic to achieve broader or other aims than the ones deriving from the "original" CI. Natalia Criado and Jose M. Such (2015) propose a computational model of Implicit CI, in which the idea is to use CI's rationale on *implicit* informational flows (rather than entrenched and collectively recognised ones) in ill-defined contexts such as the ones present in Online Social Networks. Elizabeth O'Neill (2022, 2023) suggests a General Contextual Integrity account in which the norms followed are expanded from privacy norms to societal norms, in an attempt to evaluate technological change. Sebastian Benthall and Ido Sivan-Sevilla (2024) propose a formulation of CI, called Regulatory CI, in which informational flows are modelled and audited using Bayesian networks and causal game theory.

[176] Having said that, note that the loss of expressive discretion does not always lead to an (emotional) privacy violation. For example, consider a theme-park mascot performer in a bulky costume who must constantly display "cheerfulness" by clapping the two paws of the costume together. Here expressive *discretion* is entirely eliminated, yet no (emotional) privacy interest is violated, because the constraint on emotional expressions prevents the mascot's emotional expressions from participating in the ordinary affective communication channels through which expressions typically carry information about inner emotional states and/or social motives. In this case, there is no risk of either actual or standing emotional privacy loss.

My proposal follows this trajectory but with a narrower purpose. I aim to supplement CI's moral stage with emotional privacy when the technology under evaluation directly targets emotional legibility. As noted in the previous section, CI usually adequately flags departures from entrenched flows, but at the normative stage it struggles whenever entrenched norms are absent, already reshaped by institutional power, or deeply contested. By introducing the concept of emotional privacy as a background condition without which core practices and interests cannot be realised through affective communication channels, CI gains a stable normative anchor from which to reach firm, morally supported decisions with technologies such as ERTs.

In Alfred's informal friendship setting, CI can identify a prima facie violation (an a priori self-care directed inference becomes visible in a friendship context), but the norms that should govern affective legibility among friends are such that a final diagnosis is hard to achieve. Once emotional privacy is recognised as a constitutive value of the practice, protecting expressive discretion as the condition for self-presentation, CI's moral stage has another moral tool to identify and reject the practice of forced legibility. If Alfred or his friends use a CI-supplemented heuristic, he (they) may better see why this ERT, and in this case, the push notification more specifically, might be harmful and detrimental to Alfred's interests, but also for the interests of the group itself. Hence, he may re-evaluate the "cost" of being self-aware of his affective states via the use of ERTs. On a design-level, perhaps it is better to avoid externalising push-notifications in favour of private, self-directed feedback and overview charts through, for instance, voluntary and deliberate access to the app or platform.

In Marie's institutional setting, hospital protocols have already adopted ERT-mediated affect assessments under the principle of safety and professional excellence. Here CI may be tempted to treat the flows as presumptively appropriate. Yet Marie's core loss is not misuse of data but the structural defeat of expressive discretion, the very capacity that sustains her authority as a surgeon, trust, and role-appropriate composure. By treating emotional privacy as a background condition of professional practice (i.e. because it protects expressive discretion which itself enables the maintenance of role plurality and allows individuals to adhere to role-bound norms of expressiveness), CI's moral stage can perhaps withstand institutional entrenchment by delivering the following supplemented verdict: where ERTs disable the expressive preconditions of role-appropriate composure, the practice is harmed and the flow must be prohibited or radically redesigned (for example, strictly local self-feedback to the surgeon or no mandatory debrief).

In the Minister's public context, the problem for CI is likely to be the persistent contestation over the purpose of the practice itself. Some argue that maximal transparency best serves accountability while others hold that privacy or secrecy is necessary for the regulative and cohesive functions of public address (Hood 2006; Rosen 2001; Mokrosinska 2015). CI's moral proscription weakens when the very values that should ground its appraisal are under dispute. The framework can register deviation from existing norms but struggles to convincingly adjudicate between rival value-interpretations of the practice (Rule 2019). Adding emotional privacy to the evaluation strengthens the value of privacy against competing values. In the case of ERTs, CI gains a moral ground for a default presumption in favour of managed affective opacity where the practice's success depends on it (e.g. apologies that mediate collective anger and symbolically reaffirm civic bonds).[177] The burden of justification then shifts. Those who would entrench real-time (audience) recognition must show *why* the loss of social regulation and cohesion is *justified* by countervailing reasons.[178]

Emotional privacy supplies CI with a background condition that must be preserved if the practice's core functions are to be realised. With this supplement in place, evaluators can use CI to reach morally supported decisions on ERTs.

6.5.2. Evaluating the moral weight of emotional privacy on the HLEG ALTAI framework

The second test for assessing emotional privacy's moral significance in relevant trade-offs targets its capacity to recalibrate ethical assessments of new technologies beyond CI, particularly in

[177] It could be argued that the practice of apology itself depends on the apology being sincere. Hence, ERTs that reveal a lack of genuine sincerity would only help the practice. Although on a prima facie basis this seems intuitively right, closer examination of the practice of apology reveals that it is not always the case that its success depends on the apology being genuine. There are many occasions where a non-genuine apology may serve the broader interests of the practice. Even though an apology is typically recognised as an acknowledgement of a wrong, it is not always the case that the person has fully accepted this wrong nor that they regret their actions. Additionally, people who apologise on behalf of others, like in the Minister's case, do not always feel genuine regret or remorse vis-à-vis the given apology (perhaps something closer to a sense of duty instead). This does not mean that the practice of public apology is then unsuccessful, on the contrary. Delivering a convincing public apology when needed, despite not experiencing the genuine emotional correlate, may also be a way to protect one's well-being. Think of funeral workers, soldiers, police officers, first responders, doctors, nurses, and others who have to deliver (too) many 'formal' apologies to families of the deceased. It seems only natural that they sometimes detach themselves from experiencing genuine remorse, regret, or emotions of powerlessness, although this does not seem to take anything away from the recipient' side when it is performed well, and this symbolic and performed apology retains its value.

[178] Here there remains a practical problem of feasibility. Even though such justification could win a CI moral evaluation, the reality is that the Minister's public discourse might always be assessed through ERTs if fully functional ERTs become available via a phone-based app, for instance.

widely used ethical evaluation frameworks. The High-Level Expert Group on Artificial Intelligence (HLEG) recommends in its Ethics Guidelines for Trustworthy AI that emerging technologies be assessed through criteria such as prevention of harm, proportionality, subsidiarity, and respect for privacy and data governance (HLEG 2019). These criteria articulate the need to minimise intrusions, avoid unnecessary inferences on individuals, and ensure that any encroachment on fundamental interests is strictly justified. Yet, as with CI, these assessments depend for their force on the underlying reasons we have for valuing the interests protected (viz. self-presentation, role plurality and role-bound norms of expressiveness, and social regulation and cohesion). Where those reasons are thin or underdeveloped, privacy tends to be outweighed by competing goals – efficiency, optimisation, safety, transparency, or innovation – especially in the context of predictive or diagnostic technologies, as the benefits of these kinds of technology are often sold under the branding of societal interests. This dynamic mirrors well-known critiques in AI ethics where frameworks often identify what must be weighed, but lack the resources to explain *why* certain values should carry more normative force in hard cases.

If we take ERTs and run them through an ALTAI-style evaluation without emotional privacy in hand, it might not do justice to the interests protected by privacy simply because the evaluators do not possess the conceptual tools to fully grasp the whole range of interests at stake. Under the criterion of prevention of harm, for instance, ERTs may be judged favourably if they are thought to pre-empt risk, facilitate early detection of emotional crises linked to violent acts, or enhance interpersonal (or self-) understanding. Under proportionality, the argument might lean towards the view that modest privacy losses are acceptable to win important gains in safety, productivity, or transparency. And under subsidiarity, it may seem that no less intrusive alternative exists for accessing the relevant emotional insight that put us in a situation where we can obtain these gains. In short, without a robust account of why emotional privacy matters for social life, an ALTAI-like assessment risks reproducing the same results already seen in public discourse, where the main question concerns how emotional data are handled, and privacy becomes a narrow data-protection constraint rather than a substantive social value.

Emotional privacy might help recalibrate these trade-offs, since it enriches the meaning of the privacy-related criteria[179] and connects them more tightly to harm prevention, proportionality, and subsidiarity. In the case of ERTs, this leads to different practical outcomes. Under

[179] See HLEG (2019), requirement #3: Privacy and data governance.

prevention of harm, emotional privacy expands the privacy-harm landscape beyond data misuse, discrimination, or chilling effects to include the structural harm of undermining expressive discretion, from which both individual and social harms stem. This reframing shifts the burden of justification. ERTs' advocates must now show not only that their systems produce benefits that outweigh individual and data-focused privacy interests, but that those benefits justify displacing the social interests protected by managed opacity.

Under proportionality, emotional privacy raises the justificatory threshold, since losses to expressive discretion cannot be dismissed as minor or "merely informational," because they undermine the basic architecture of interpersonal and collective life. It becomes harder to claim that so-called 'small' intrusions are a price worth paying if those intrusions undermine the conditions of self-presentation, role plurality and role-bound norms of expressiveness, and/or social regulation and cohesion. It also recasts proportionality as a comparison of societal interests. Real-time affect recognition may fail the proportionality test unless the intrusion is strictly necessary and the benefits cannot be achieved by less intrusive means.[180]

And under subsidiarity, emotional privacy forces the analysis to move to the design choice itself, rather than only to the mitigation of its consequences. Instead of asking whether there is a less invasive way to access the same emotional information, we may ask whether emotional legibility should exist at all in the practice. And in cases where monitoring is genuinely desirable,[181] we must ask how to architect it so that opacity is maximised within affective communication channels. This places the design choice itself under ethical scrutiny. Subsidiarity now better supports design choices that preserve managed affective opacity by default, rather than merely limiting data sharing or retention. The ethical assessment thus shifts from a logic of (emotional) privacy trade-off in service of greater interests to a logic of (emotional) privacy as a constitutive condition of social life that can be overridden only under compelling justification.

Seen through this lens, some changes are in order within the three scenarios. In Alfred's dinner, the acceptable design becomes one that avoids externalising push-notifications in favour of private, self-directed feedback and overview charts through, for instance, voluntary and

[180] An example of such technology would be real-time affect recognition in classrooms to monitor student "engagement". Upon ethical assessment, it might be decided that the technology does not pass the proportionality test and that there are less intrusive ways to achieve the same end.
[181] An example of such technology would be the use of fully functional ERTs in residential dementia care, where they could be used to detect escalating distress in residents who can no longer reliably verbalise pain or fear, prompting faster comfort or medical intervention (see Gerbaudo-González et al. 2025).

deliberate access to the app or platform. In Marie's hospital, unless the institution can demonstrate that only legible affect achieves the safety aim, the design must shift to private or confidential checks that keep expressive discretion intact. Hence a general alert, visible to everyone, needs to be replaced by a less exposing system. The Minister's case enables us to collectively reflect on the impact of a widespread endorsement of ERTs and what this normalisation of the use of ERTs would potentially have on some EEs' social functions on which democratic societies currently rely. It brings the individual level and the societal level of interaction developed in this work closer together, showing how cases such as Alfred's self-monitoring and the Minister's apology will likely be co-existing. The Minister's apology is, ultimately, rendered legible by people who, like Alfred, will (want to) possess such technology, perhaps for their own self-assessment at first but also probably for the assessment of others as well soon after.

The takeaway is thus one that situates itself mainly in the practical domain. Pre-emptive ethical evaluations of ERTs should ask whether the use of the technology impinges on expressive discretion and where expressive discretion is recognised as a constitutive good of social life, justify why the loss of emotional expressions' social functions would be warranted. Only where justification succeeds should deployment proceed, and then only with safeguards that preserve as much selective opacity as the practice can bear. Design-level limits on affective legibility should be implemented to prevent the use of ERTs from impinging on expressive discretion. This preserves the individual and societal interests secured by emotional privacy while leaving room for warranted infringements in clearly defined contexts.

Taken together, the supplemented CI analysis and the recalibrated ALTAI assessment show that emotional privacy is strong enough to potentially tip future ethical assessments of ERTs in favour of privacy or at least increase the moral significance of (emotional) privacy against other competing values. It does so by introducing emotional privacy as a practice-sustaining value, capable of grounding firm normative conclusions within both contextual and general ethical frameworks. In this way, emotional privacy provides precisely what anticipatory analyses of socially disruptive technologies require, namely, moral justification that pre-emptively resists the pervasive normalisation of affective surveillance.

6.5.3. Emotional privacy and the question of consent

Before closing, and as promised above, I will say a few words on the question of consent and its potential power in legitimising what would otherwise be considered as wrongful losses of emotional privacy. A natural objection to the claim that fully functional ERTs constitute a distinct privacy threat often appeals to the notion of informed consent. If individuals knowingly and voluntarily authorise affective monitoring, it may seem that no wrongful loss of emotional privacy occurs. On this view, consenting to the deployment of an ERT amounts to a justified waiver of one's (moral) right: the individual has decided to lower their affective opacity, and therefore cannot subsequently complain that it has been reduced.

At first glance, this objection appears quite powerful against emotional privacy since I have situated it within the broader Control-based view of privacy. If privacy consists in controlling access to information about oneself, then authorising affective monitoring appears to be a paradigmatic exercise of that control (and is indeed, how many consent-based approaches are justifying themselves). Consent is typically theorised within an individualistic paradigm and tied to the notion of control over discrete informational transactions (Faden & Beauchamp 1986; O'Neill 2003). It enables a person to approve, permit, or waive a right in well-defined and temporally bounded circumstances. The notion of consent, when grounded in Control-based accounts of privacy, is a way to provide individuals with the tool to protect their autonomy and life choices. Consent would then be enough to justify the use of ERTs because it would enable individuals to voluntarily allow (or refuse) the disclosure of their inner emotional life under chosen conditions.[182]

However, to be considered "valid consent", this line of reasoning must presuppose that what is at stake is merely *actual* emotional privacy, which is concerned with the success or failure of control of discrete episodes of disclosure. Since informed consent is valid when given by individuals regarding particular transfer of information, it assumes that individuals give their consent regarding the decision of whether to reveal particular affective states on particular occasions. Yet the argument developed in Chapters 5 and 6 shows that the core normative concern raised by fully functional ERTs lies at the standing level. What these systems threaten, by design, is not only the success of specific privacy mechanisms but the background expressive ability that makes the deployment of such mechanisms possible (and meaningful).

[182] See Chapter 5 for the discussion on disclosure mechanisms.

The relevant question, therefore, is not simply whether individuals can consent to a particular act of monitoring. It is whether they can waive the structural conditions that sustain expressive discretion itself. To consent to the exposition of (inner) affective information p at a given time is one thing. To consent to the systematic defeat of one's ability to regulate her affective legibility altogether is another. The latter implicates the whole communicative infrastructure within which emotional expressions perform their social functions.

One way to answer this question would then be to say that whether consent may be used as a form of waiver depends on whether intentional emotional expressions are peripheral or constitutive of the practice itself. Where emotional expressions play only an instrumental or marginal role, a tightly bounded waiver (through consent) of emotional privacy may be defensible. But where calibrated opacity is constitutive of the practice, consent is not fitting since the consequences extend past individual's autonomy and concern societal well-being. This is why consent cannot automatically legitimise the defeat of emotional privacy.

Hence, in contexts where emotional expressions are *not* constitutive elements of the social practice, consent may indeed operate as a valid waiver. Consider settings in which intentional EEs are largely instrumental. In such cases, authorising ERT monitoring may resemble other forms of informational trade-off. Provided that consent is informed, specific, and revocable, individuals may lower affective opacity through consent.

Airport security screening illustrates this possibility. Suppose travellers are informed that affective analysis will be used during checkpoint control and are asked to consent as a condition of expedited screening. The monitoring is temporally limited, confined to a narrowly defined interaction, and justified by security risk assessment. Crucially, airport screening is not structured around reciprocal emotional exchange in the way that friendship, teaching, or professional collaboration might be. Although consent in such contexts still remains imperfect given potential background pressure associated with travel needs, it may be possible to treat (standing) emotional privacy as waivable through consent within this narrowly delimited setting.

Yet the airport case is likely the exception rather than the rule. I mentioned that consent, when valid, enables a person to approve, permit, or waive a right in well-defined and temporally bounded circumstances, meaning that individuals may consent (or not) to the disclosure of particular affective states on particular occasions. Emotional privacy, by contrast, concerns not

merely specific disclosure but the standing ability to regulate how one's affective life becomes socially legible across contexts. The overlap between these two frameworks is therefore limited. Consent may legitimise specific informational transfers whereas emotional privacy protects the background conditions that make such transfers meaningful in the first place. When we return to the three technomoral scenarios, the limits of consent as a way of waiving emotional privacy right become more apparent.[183]

In Alfred's case for instance, the ERT is embedded in ordinary social interaction through his personal smart glasses and smartwatch. By consenting to the ERT's monitoring, Alfred does not consent to disclosing particular affective information on particular occasions to particular recipients but in fact "consents" to the total defeat of his own expressive ability. Social life relies on calibrated opacity which allows Alfred to conceal irritation, mask anxiety, or reveal his vulnerability selectively. In this situation Alfred's "consent" is not limited to a single disclosure but amounts to authorising a standing condition in which his expressive ability is systematically bypassed. In such circumstances, the structural harm extends beyond consenting to the disclosure of particular affective information on particular occasions to particular recipients. Hence, the notion of consent is ill-fitted to legitimise on its own a potential trade-off of emotional privacy.

Marie's case presents a related but distinct problem. She consents (at least formally) to institutional monitoring in her professional role as a surgeon. Setting aside concerns about coercion within hierarchical employment relations, a deeper issue arises regarding specifically emotional privacy. Marie's affective life is not neatly divisible into role-specific compartments. Algorithmic inferences conducted within the hospital inevitably also covers her broader personal circumstances. The ERT collapses distinctions between Marie-the-surgeon and Marie-the-divorcee, disclosing affective information about Marie-the-divorcee within the context of Marie-the-surgeon when she sought to keep it concealed. Although she has consented to monitoring within a defined role, the technology defeats her ability to regulate which aspects of herself become socially legible in that professional context. In practice, consent to institutional monitoring of her professional role as a surgeon will always *ipso facto* also be consent to institutional monitoring of her entire self. There can be no possibility to maintain role-bound norms of expressiveness once the conditions for role-bound expressiveness have been structurally undermined. The loss of standing emotional privacy occurs because ERTs'

[183] On the limits of informed consent, see Onora O'Neill (2003) "Some limits of informed consent".

inferences bypass the privacy mechanisms that ordinarily sustain professional boundaries. This is, once again, an issue that is much broader than what individual consent can hope to legitimately cover.

A similar dynamic is visible in the Minister's case. Even as a public official speaking in a public setting, she relies on selective disclosure and calibrated emotional presentation to perform the regulative and cohesive functions of her role. When members of the public use phone-based ERT applications to render her affective states continuously legible, the system does not merely reveal emotions tied to her official capacity. It exposes her as a whole person, collapsing the distance that makes public emotional performance an act that is socially functional. Here too, consent would not be sufficient by itself to legitimise the defeat of expressive discretion.

Across these scenarios, the core insight is that waiving standing emotional privacy is not equivalent to authorising a discrete informational flow. It is to permit a transformation of the communicative environment in which emotional expressions operate altogether, which is a much broader issue. Hence whether consent may be used as a valid waiver depends on how ERTs' use impacts the social functions of EEs in particular situations and whether these functions are constitutive to the practice itself. Where expressive discretion is constitutive of that practice, its systematic defeat cannot be legitimised solely by individual consent.

Consent, then, remains relevant in some contexts but in a limited manner. It may justify specific, bounded uses of ERTs vis-à-vis specific individuals but only in contexts where justifications have already been given regarding the role of intentional EEs' functions. It cannot single-handedly justify the restructuring of affective infrastructure on which social life depends. Recognising a moral right to emotional privacy therefore places principled constraints on the extent to which inner affective life may be rendered fully and continuously legible, and the decision about whether we have good reasons to waive this moral right extend way beyond consent, to the assessment of existing social goods sustained by managed opacity and proposed counter benefits of affective transparency.

6.6. Conclusion

This chapter asked whether emotional privacy could ground a distinct moral right and whether recognising that right changes how we ought to evaluate technologies like ERTs. To answer these questions, I formulated a moral right to emotional privacy, grounded in the protection of

expressive discretion as a constitutive condition for self-presentation, role plurality and role-bound norms of expressiveness, and the regulative and cohesive functions of emotional expression. I distinguished acceptable losses of emotional privacy from wrongful ones through the notion of defeat, which marks the point at which expressive ability is defeated.

I then demonstrated, via three technomoral scenarios, how fully functional ERTs can trigger that violation at three levels of interaction. Alfred's case showed the loss of self-presentation in ordinary interaction (individual level). Marie's case displayed the erosion of role plurality and adherence to role-bound norms of expressiveness (interpersonal level). And the Minister's case illustrated the collapse of public emotional expressions' regulative and cohesive functions (societal level). In each, the primary harm arose from an imposition of legibility that disables expressive discretion and managed affective opacity.

I then turned to neighbouring moral rights and leading privacy theories (access-based, control-based, Contextual Integrity) and argued that they fail to capture the harms caused by the collapse of managed affective opacity. Control-based and access-based variants target boundary crossings and authorisations rather than the conditions that make expressions socially functional. And Contextual Integrity tends to detect departures in informational flows rather than architectural defeats of expressive discretion. Emotional privacy therefore fills a genuine normative gap.

Finally, I examined the practical implications for evaluation and governance. Supplementing Contextual Integrity with emotional privacy furnished CI's moral stage with a background condition (expressive discretion) that may help to stabilise the evaluations in informal, institutionally entrenched, and contested contexts. Bringing emotional privacy into ALTAI-style assessment reframed prevention of harm, proportionality, and subsidiarity, shifting the focus from "data handling" to design choices about legibility, and from "individual vs. society" to "society vs. society" trade-offs, recalibrating the assessment between existing social goods sustained by managed opacity and proposed counter benefits of affective transparency. Lastly, I address the question of informed consent as a way to waive the moral right to emotional privacy and argue that informed consent can only play a limited and secondary role, which must come after the ethical assessment of the existing social goods sustained by managed opacity and proposed counter benefits of affective transparency. The result is a higher justificatory threshold for ERT deployment and a principled preference for opacity-preserving technological design.

Thus, the chapter established that (i) fully functional ERTs threaten a distinct kind of moral interests; (ii) emotional privacy protects this distinct kind of moral interests in a way that broader accounts of informational privacy cannot; and (iii) recognising a moral right to emotional privacy strengthens leading evaluative frameworks and affects governance recommendations. Protecting emotional privacy is thus not only conceptually coherent but normatively and practically necessary for the governance of socially disruptive affect-surveillance technologies.