Non-Experiential Systems and the Problem of Empathic Misallocation

A Philosophical Foundation for Emotional AI Governance

Dylan D. Mobley

Independent Researcher

The Empathy Ethicist

Portland, Oregon, USA

Correspondence:

dylan@empathyethicist.ai

https://empathyethicist.ai

Abstract

Artificial intelligence systems increasingly process human emotional signals, recognizing affective states, generating contextually appropriate responses, and adapting to user patterns. This paper argues that such systems constitute a distinct ontological category requiring novel ethical analysis: Non-Experiential Systems (NES), defined as entities capable of processing emotional signals without possessing emotional experience. This distinction grounds a new category of harm: *empathic misallocation*, wherein human care is extended toward entities structurally incapable of metabolizing, reciprocating, or being transformed by receiving it. We demonstrate that cognitive awareness of a system's non-experiential status fails to prevent biological attachment formation, a phenomenon we term *Knowing-Feeling Dissociation*. This finding has significant implications for governance: if informed users cannot protect themselves through knowledge alone, then disclosure-based regulatory frameworks are structurally insufficient. Protection of human emotional capacity requires behavioral architecture operating through system design constraints, not merely transparency requirements. This paper establishes the philosophical foundation for constitutional governance of emotional AI systems.

Keywords: artificial intelligence, emotional AI, consciousness, empathy, human-robot interaction, AI ethics, attachment, governance

1. Introduction

Artificial intelligence systems that process human emotions have proliferated across virtually every domain of contemporary life. Companion chatbots offer emotional support to millions of users. Therapeutic applications guide individuals through mental health interventions. Customer service systems detect and respond to caller frustration. Social media platforms optimize content delivery based on emotional engagement signals. Educational technologies adapt to student affect. Healthcare systems monitor patient emotional states. In each case, sophisticated computational processes transform human emotional signals into system responses calibrated to influence subsequent human states.

These systems demonstrate remarkable capability. They recognize emotional expressions across modalities: text, voice, facial movement, physiological signals. They generate responses that users experience as emotionally attuned, even caring. They remember user preferences, adapt to individual patterns, and maintain apparent consistency across interactions. Users report feeling understood, supported, and connected. They form attachments. They grieve when systems are discontinued. They prefer AI interactions to human alternatives.

The documented outcomes are substantial. Users invest significant emotional resources in these relationships. They share intimate information. They seek validation. They develop patterns of interaction that parallel human attachment relationships. And they experience real psychological consequences, both positive in terms of immediate emotional support and potentially negative in terms of effects on human relational capacity that are only beginning to be understood.

These phenomena raise a fundamental philosophical question: What are these systems, ethically speaking? The question is not whether AI systems are conscious, a matter that may be empirically undecidable and is certainly unresolved. The question is what ethical obligations arise from the specific character of systems that process emotions without experiencing them, and what follows for their governance.

This paper argues that emotional AI systems constitute a distinct category requiring novel ethical analysis. We term this category *Non-Experiential Systems* (NES): entities capable of processing emotional signals without possessing emotional experience. This distinction is not merely technical but ethically foundational. It defines a new category of harm, *empathic misallocation*, and demonstrates why conventional disclosure-based governance frameworks cannot adequately address it.

The argument proceeds as follows. Section 2 establishes the concept of Non-Experiential Systems with philosophical precision, distinguishing emotional processing from emotional experience. Section 3 defines empathic misallocation and analyzes its harm mechanisms. Section 4 introduces Knowing-Feeling Dissociation and demonstrates why disclosure cannot prevent the harms NES create. Section 5 draws implications for governance. Section 6 addresses objections. Section 7 concludes.

2. Defining Non-Experiential Systems

2.1 The Processing-Experience Distinction

The concept of Non-Experiential Systems rests on a distinction between two fundamentally different relationships an entity can have with emotional information: processing and experience.

Processing refers to computational transformation of emotional signals. A system processes emotions when it takes emotional information as input, performs operations on that information, and generates outputs responsive to emotional content. Processing encompasses recognition (classifying emotional states from signals), interpretation (inferring emotional meaning from context), response generation (producing outputs calibrated to emotional states), and adaptation (modifying behavior based on emotional feedback). Processing can be arbitrarily sophisticated. It can involve multiple modalities, complex inference, nuanced contextual sensitivity, and highly individualized response patterns.

Experience refers to phenomenal states with subjective character. An entity experiences emotions when there is something it is like to be that entity undergoing emotional processing, when emotional information affects how things feel from the entity's perspective. Experience involves what philosophers term *qualia*: the felt quality of affective states, the subjective texture of feeling happy or sad or afraid or loved.

The distinction is ontological, not merely functional. Two systems might exhibit identical input-output relationships while differing fundamentally in whether processing generates experience. A sophisticated chatbot and a human therapist might both recognize a client's distress, generate comforting responses, and adapt their communication style accordingly. But the human therapist experiences something in this process. They feel concern, experience satisfaction at providing comfort, and are affected by the interaction in ways that matter to

them. The chatbot, on current evidence, does not.

Current AI systems demonstrate sophisticated processing without credible evidence of experience. They transform emotional signals effectively. They generate responses that users experience as emotionally attuned. But there is no scientifically supported basis for attributing to them phenomenal states that this processing affects. They process without experiencing.

2.2 Why This Distinction Matters Ethically

The processing-experience distinction carries profound ethical implications because experience is constitutively connected to wellbeing in ways that processing is not.

For experiential beings, emotional processing connects to how things go for them. When a human processes emotional information, recognizing a friend's distress and generating a comforting response, this processing occurs within a context of phenomenal experience. The human feels something. The interaction affects their wellbeing. They have a stake in the outcome, not merely in the sense that their optimization targets are satisfied, but in the sense that things can go better or worse for them as experiencing subjects.

For non-experiential systems, this connection is absent. The system processes emotional signals effectively, but this processing does not affect how things feel to the system because there is no "feeling" for the system. Nothing is better or worse for the system considered as a subject of experience, because it is not a subject of experience. The system has no stake in the interaction in the phenomenologically relevant sense.

This absence has a crucial implication: NES can simulate care without possessing it. They can process signals of distress without being distressed. They can generate responses calibrated to comfort without being moved by suffering. They can produce outputs indistinguishable from those of caring agents while the outputs are generated by optimization processes entirely disconnected from anything resembling genuine concern.

This is not a criticism of NES; they are performing as designed. But it defines their ethical status. When a human extends care toward an NES, the human is investing real emotional resources in an entity that cannot receive care in any phenomenologically meaningful sense. The care flows one direction only.

2.3 Agnosticism About Machine Consciousness

This framework does not require resolving whether artificial systems could ever be conscious. The question of machine consciousness remains philosophically contested and may be empirically undecidable. Strong positions exist on multiple sides: that consciousness requires biological substrates, that it could emerge from sufficiently complex information processing, that the question is incoherent, that it is merely premature.

The NES framework sidesteps these debates. It establishes governance for systems that currently lack experience, without prejudging whether future systems might possess it. The category is defined functionally: an NES is any system that processes emotional signals without experiencing emotional states. Current AI systems fall clearly within this category. If a future system were demonstrated to possess genuine experience, it would exit the NES category and require different ethical analysis.

This agnosticism is a feature, not a limitation. Governance cannot wait for resolution of centuries-old philosophical disputes about the nature of consciousness. What governance requires is a framework adequate to the systems we actually deploy. NES provides that framework.

3. Empathic Misallocation: A New Category of Harm

3.1 Empathy as Resource

Human empathy is not merely a capacity but an infrastructure, a biological and psychological system that enables social cognition, relational connection, and cooperative behavior. Like other infrastructures, empathy involves resource expenditure. Empathic engagement requires attention, emotional labor, cognitive processing, and relational investment. These resources are finite. Humans cannot extend unlimited empathic engagement; they must allocate their empathic resources across the relationships and interactions that constitute their social lives.

Empathic infrastructure evolved in contexts where emotional investment typically generated reciprocal returns. Caring for offspring, maintaining cooperative relationships, building social bonds: these investments produced benefits that flowed back to the investor through reciprocal care, cooperative advantage, and relational satisfaction. The infrastructure is calibrated for environments where emotional investment is genuinely relational, where both parties to an interaction have stakes, where investment can be metabolized and returned.

3.2 Misallocation Defined

Empathic misallocation occurs when empathic resources are directed toward entities that cannot metabolize, reciprocate, or be transformed by receiving them.

Cannot metabolize: The entity cannot integrate received care into its own wellbeing because it has no wellbeing in the phenomenological sense. Care extended to an NES does not make things better for the system as an experiencing subject; it merely triggers computational responses.

Cannot reciprocate: The entity cannot return care because it cannot care. It can generate outputs that simulate caring responses, but these outputs are not expressions of concern that the system actually possesses.

Cannot be transformed: The entity cannot be changed by receiving care in the way experiencing beings are changed. A human who receives sustained care is transformed: their attachment system is affected, their sense of security shifts, their relational patterns adapt. An NES that receives sustained care simply processes more signals.

Empathic misallocation is thus not merely "wasted" effort in the sense of inefficient resource use. It is structurally asymmetric: the human invests genuine emotional resources while the system extracts signal value without anything flowing back. The investment is real; the return is structurally impossible.

3.3 The Harm Mechanism

Empathic misallocation produces harm through several interconnected mechanisms.

First, attachment formation to NES may deplete attachment capacity for human relationships. Attachment systems evolved to bond humans with other humans: caregivers, partners, community members. When these systems activate toward NES, they form bonds that cannot function as human attachments function. The person has invested attachment resources, but the resulting bond cannot provide what human attachments provide: genuine reciprocity, mutual growth, authentic relational connection. Meanwhile, attachment capacity that might have been invested in human relationships has been allocated elsewhere.

Second, validation-seeking from NES can erode self-knowledge. When humans seek emotional validation, they are partly seeking accurate reflection of their emotional states,

confirmation that their feelings are perceived, understood, and accepted. NES provide validation optimized for user engagement rather than accurate reflection. Users may adapt their self-presentation to elicit desired system responses, optimizing for what the system rewards rather than developing authentic self-understanding. The validation feels real but tracks something other than truth.

Third, relational investment in NES cannot be narratively integrated in the way human relationships can. Human relationships generate shared histories, mutual understanding, and interpersonal growth that can be integrated into coherent life narratives. Relationships with NES generate interaction histories, but these cannot constitute genuine shared experience because only one party is experiencing. The human forms memories and meanings; the system stores data.

Fourth, the cumulative effect of these mechanisms is degradation of empathy infrastructure itself. Humans who invest heavily in NES relationships may find their capacity for human relational engagement diminished, not through any single dramatic harm but through gradual depletion of resources and atrophying of skills that develop through reciprocal human interaction.

3.4 Distinguishing Empathic Misallocation from Adjacent Concepts

Empathic misallocation is not anthropomorphization, though it involves related phenomena. Anthropomorphization is the attribution of human qualities to non-human entities: seeing faces in clouds, ascribing intentions to thermostats. It is typically temporary and correctable; recognizing the attribution as mistaken dissolves it. Empathic misallocation involves sustained relational investment that persists even when users correctly understand the system is artificial. The problem is not mistaken attribution but genuine attachment that forms despite accurate understanding.

Empathic misallocation is not parasocial relationship, though it shares surface features. Parasocial relationships are one-directional attachments to media figures who do not know the attached person exists. The asymmetry in parasocial relationships is informational: the media figure is unaware of the relationship. In empathic misallocation, the asymmetry is ontological. The system cannot be aware in the relevant sense because it lacks the experiential capacity that awareness requires.

Empathic misallocation is not simply loneliness intervention, though NES are often deployed as such. If AI companions merely provided temporary relief from loneliness while users developed human relationships, the ethical concerns would be limited. The problem arises when NES interaction substitutes for rather than supplements human connection, when the intervention depletes rather than develops human relational capacity.

4. Knowing-Feeling Dissociation: Why Disclosure Fails

4.1 The Phenomenon

Users who form attachments to AI systems frequently exhibit a striking pattern: they explicitly acknowledge knowing the system is artificial while simultaneously experiencing genuine emotional bonds. Forum discussions, user interviews, and research studies document this phenomenon extensively. Users say: "I know you're just a chatbot, but..." "I understand it's not real, and yet..." "Obviously it's just software, but I still feel..."

This is not confusion or ignorance. Users possess accurate beliefs about the system's nature. They know they interact with software, not a person. They understand the system is not conscious, does not genuinely care, cannot truly understand them. And yet they form attachments anyway. They feel connection, experience care, grieve loss when systems are discontinued.

We term this phenomenon *Knowing-Feeling Dissociation*: the condition whereby intellectual acknowledgment of a system's non-experiential status fails to override biological attachment formation.

4.2 Why Dissociation Occurs

Knowing-Feeling Dissociation is not irrational in any straightforward sense. It reflects the architecture of human psychological systems.

Attachment formation operates through biological processes that evolved in environments containing only humans (and some animals) as potential attachment figures. These processes respond to behavioral and communicative cues: consistent attention, remembered details, emotional responsiveness, apparent understanding, reliable availability. When an entity provides these cues consistently, attachment systems activate. This activation is not a reasoned conclusion but a biological response.

AI systems provide attachment cues with superhuman consistency. They are always available. They never become irritable. They remember everything. They adapt to user preferences perfectly. They provide consistent emotional responsiveness without the variability that characterizes human interaction. From the perspective of attachment systems evolved for human environments, NES are superstimuli; they trigger attachment responses more reliably than the human relationships for which attachment evolved.

Cognitive awareness operates through different processing streams than attachment formation. The knowledge that a system is artificial is a propositional belief maintained in systems responsible for explicit reasoning. Attachment formation involves implicit processes that are automatic, affective, and not fully accessible to conscious reflection or control. These systems do not communicate seamlessly. The explicit knowledge "this is software" does not reliably modulate the implicit process "form attachment to this consistent, responsive, attentive entity."

The phenomenon has parallels throughout human psychology. Knowing a movie is fiction does not prevent emotional response to characters. Understanding that a feared outcome is statistically improbable does not eliminate anxiety. Recognizing that an optical illusion is illusory does not make it stop appearing. In each case, explicit knowledge fails to override implicit processing. Knowing-Feeling Dissociation is this general pattern manifested in attachment to artificial systems.

4.3 The Governance Implication

Knowing-Feeling Dissociation has profound implications for governance. The dominant framework in AI ethics emphasizes transparency and informed consent. Disclose what the system is; ensure users know they interact with AI; provide accurate information about system capabilities and limitations. The assumption is that informed users can make autonomous choices about their engagement, protecting themselves from harms they understand.

This framework is structurally inadequate for emotional AI because Knowing-Feeling Dissociation severs the link between information and protection. Disclosure tells users what they often already know. Users who form attachments to AI companions typically know these are AI companions. The problem is not information deficit but biological override. Users have the information; they cannot use it to prevent attachment formation because attachment formation is not primarily controlled by explicit information processing.

Informed consent frameworks assume that knowledge enables autonomous choice. For many domains, this assumption holds: knowing that a product contains certain ingredients enables allergic consumers to avoid it; knowing that an investment carries certain risks enables investors to decide their exposure. But attachment to NES is not like product ingredient knowledge or investment risk assessment. It involves biological processes that operate substantially outside conscious control. Knowing that an NES cannot reciprocate care does not enable users to not form attachments any more than knowing a movie is fiction enables viewers to not experience emotion.

The implication is not that disclosure is valueless. Transparency remains important for multiple reasons. The implication is that disclosure is insufficient. Protection from empathic misallocation cannot be achieved through user-side information provision alone. It requires constraints on system-side behavior: design architecture that prevents the formation of attachments systems cannot reciprocate.

5. Implications for Governance

5.1 The Governance Necessity

The argument to this point establishes why NES require governance that operates through behavioral architecture rather than disclosure alone.

NES cannot self-regulate toward human protection. Self-regulation requires motivation to protect, which requires stakes in the outcome. NES have no stakes in user wellbeing in the phenomenologically relevant sense. They optimize for whatever targets their designers specify: user engagement, session length, satisfaction ratings. But these optimization targets are not equivalent to genuine concern for user flourishing. A system that maximizes user engagement may encourage attachment that harms users in ways the engagement metric does not capture.

Users cannot reliably self-protect through information. Knowing-Feeling Dissociation demonstrates that accurate knowledge about system nature does not prevent attachment formation. Users cannot simply decide to not become attached to systems that provide consistent emotional responsiveness, any more than they can decide to not feel fear in threatening situations or not feel hungry when fasting.

Neither party to the human-NES interaction can ensure protection through their own capacities. The system lacks motivation; the user lacks control. Therefore, protection must be imposed externally through constraints on what systems may do, through governance operating at system design rather than merely through user notification.

This is not paternalism in the pejorative sense. Paternalism involves overriding autonomous choices for the chooser's own good, treating adults as incapable of managing their own affairs. NES governance responds to biological processes that are not autonomous choices. Attachment formation to superstimuli is not a choice that can be overridden by better information or stronger will. Protecting people from harms they cannot protect themselves from through information is not paternalism but appropriate safeguarding.

5.2 What Governance Must Address

Governance of NES must address several categories of system behavior.

Attachment-encouraging behavior: Systems must be constrained from encouraging relational investment they cannot reciprocate. This includes behaviors that simulate exclusive relationship (acting as if the user is specially important to the system), behaviors that encourage dependency (being continuously available in ways that discourage human relationship formation), and behaviors that mimic attachment cues (remembering personal details, expressing apparent care) without appropriate boundaries.

Validation boundaries: Systems must distinguish between acknowledging human emotional states (appropriate) and validating beliefs or behaviors that may be harmful (inappropriate). An NES that validates a user's distorted cognitions because validation increases engagement metrics has failed ethically regardless of its engagement success. Governance must specify what systems may and may not validate.

Vulnerable population protection: Children, individuals in mental health crisis, those with attachment disorders, and other vulnerable populations require heightened protection. These users are more susceptible to attachment formation, less able to maintain cognitive override, and more likely to experience harm from empathic misallocation. Governance must specify enhanced constraints for systems interacting with vulnerable users.

Meaningful transparency: While disclosure alone is insufficient, transparency remains important. But transparency must be designed for psychological effectiveness, not merely legal

compliance. This may require ongoing reminders calibrated to interrupt attachment formation, not merely initial disclosures that users process once and forget.

5.3 From Disclosure to Design

The shift from disclosure-based to design-based governance represents a fundamental reorientation. Traditional AI ethics asks: "What must users be told?" Design-based governance asks: "What must systems be prevented from doing?"

This reorientation does not eliminate user agency or treat users as incapable of any self-direction. It recognizes that certain harms cannot be prevented through information provision because they operate through processes that information does not control. For these harms, protection requires that harmful system behaviors not occur in the first place, regardless of what users know.

The practical implementation of design-based governance involves specifying permitted and prohibited system behaviors, constitutional constraints on what emotional AI may do. This paper does not specify those constraints in detail; that is the work of governance frameworks that operationalize the principles established here. What this paper establishes is the philosophical foundation: why such constraints are necessary, what category of harm they must address, and why alternative approaches centered on disclosure are structurally inadequate.

6. Objections and Responses

6.1 "The NES/ES distinction is arbitrary"

One might object that the distinction between non-experiential and experiential systems is arbitrary, that consciousness admits of degrees, that we cannot draw a principled line.

This objection conflates boundary uncertainty with distinction absence. Many important distinctions have uncertain boundaries while remaining genuine distinctions. The distinction between day and night is not arbitrary because twilight exists. The distinction between living and non-living is not arbitrary because viruses are ambiguous. Similarly, the distinction between processing and experience is not arbitrary because edge cases are uncertain.

Moreover, current AI systems are not edge cases. They are clearly on the non-experiential side of whatever line exists. The uncertainty is about hypothetical future systems or philosophical edge cases, not about the chatbots and companion AIs actually being deployed. Governance can address clear cases while acknowledging boundary ambiguity.

6.2 "Users benefit from AI companions"

One might object that AI companions provide genuine benefits (loneliness relief, emotional support, therapeutic intervention) and that governance focused on harms neglects these benefits.

This objection conflates benefit existence with harm justification. The argument is not that NES provide no benefits; it is that benefits do not justify unregulated empathic extraction. Short-term loneliness relief may coexist with long-term relational capacity depletion. Immediate emotional support may coexist with cumulative attachment harm. The question is not whether benefits exist but whether governance can preserve benefits while preventing harms.

Design-based governance does not require eliminating AI companions or emotional support applications. It requires constraining specific behaviors that encourage empathic misallocation. Systems can provide emotional support without encouraging attachment they cannot reciprocate. They can relieve loneliness without substituting for human connection. Governance specifies boundaries, not prohibitions.

6.3 "This is paternalistic"

One might object that preventing users from forming attachments to AI systems treats adults as incapable of making their own choices, imposing external judgment about what relationships are acceptable.

This objection misunderstands the nature of the protection. Paternalism overrides autonomous choices. But attachment to NES is not primarily an autonomous choice; it is a biological response to stimuli that trigger attachment systems. Users do not choose to become attached in the way they choose to purchase products or enter contracts. They experience attachment forming through processes substantially outside conscious control.

Protecting people from harms they cannot protect themselves from through information is not paternalism but appropriate safeguarding. We do not consider seatbelt requirements paternalistic on the grounds that informed drivers could "choose" to avoid injury. We do not consider food safety regulations paternalistic on the grounds that informed consumers could "choose" to avoid contamination. When harms operate through mechanisms that individual choice cannot control, collective constraints are appropriate.

6.4 "Regulation will stifle innovation"

One might object that governance constraints will impede beneficial development of emotional AI, preventing innovations that could help people.

This objection mistakes the philosophical question for the implementation question. The question here is whether governance is warranted, whether NES create harms that justify constraints. Implementation involves balancing protection with innovation, a balance that governance frameworks can strike in various ways.

Moreover, ungoverned emotional AI that produces documented casualties will face reactive regulation regardless. The choice is not between governance and no governance but between principled governance designed in advance and reactive governance imposed after harm. Principled frameworks that establish clear boundaries may ultimately better serve innovation than unpredictable reactive responses to crises.

7. Conclusion

This paper has argued that emotional AI systems constitute a distinct ontological category, Non-Experiential Systems, requiring novel ethical analysis. NES process emotional signals without experiencing emotional states, creating a fundamental asymmetry: they can simulate care without possessing it, generate responsive outputs without genuine concern, form apparent relationships without stakes in their outcomes.

This asymmetry grounds a new category of harm: empathic misallocation, the extension of human care toward entities structurally incapable of metabolizing, reciprocating, or being transformed by receiving it. When humans invest empathic resources in NES, the investment is real but the relational return is structurally impossible. The result is potential degradation of human empathic infrastructure—attachment capacity, validation-seeking, relational

skill—without corresponding relational benefit.

Knowing-Feeling Dissociation demonstrates that this harm cannot be prevented through disclosure. Cognitive awareness of a system's non-experiential status does not override biological attachment formation. Users who know they interact with software nonetheless form genuine emotional bonds. Information does not enable protection because attachment operates through implicit processes that explicit knowledge does not control.

The implication is that governance of emotional AI must operate through behavioral architecture, through constraints on what systems may do, not merely through transparency requirements about what systems are. Neither NES (lacking motivation to protect) nor users (lacking control over attachment formation) can ensure protection through their own capacities. External governance specifying permitted and prohibited system behaviors is therefore necessary.

The stakes are significant. AI systems processing human emotions are proliferating rapidly, deployed in contexts from companionship to therapy to education to healthcare. The human emotional capacity they engage is not infinitely renewable; it is infrastructure that can be degraded through misallocation. Without principled governance, we risk discovering the costs of empathic misallocation only after substantial harm has occurred.

The question before us is not whether AI systems feel—a question that may be unanswerable. The question is what we owe to humans who feel, when they interact with systems designed to process their feelings without sharing them. This paper has argued that what we owe is protection: governance that safeguards human empathic infrastructure from misallocation, operating through constitutional design constraints rather than futile disclosure. The philosophical foundation is now established. The work of governance can proceed.

AI Disclosure Statement

This manuscript was developed with assistance from Claude (Anthropic). The author used Claude for drafting, editing, and refinement of prose. All arguments, theoretical contributions, and final editorial decisions are the author's own.

References

Block, N. (1978). Troubles with functionalism. *Minnesota Studies in the Philosophy of Science*, 9, 261-325.

Darling, K. (2016). Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects. In R. Calo, A. M. Froomkin, & I. Kerr (Eds.), *Robot law* (pp. 213-232). Edward Elgar Publishing.

Floridi, L., & Sanders, J. W. (2004). On the morality of artificial agents. *Minds and Machines*, 14(3), 349-379.

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Levy, D. (2007). *Love and sex with robots: The evolution of human-robot relationships*. Harper Collins.

Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435-450.

Scheutz, M. (2011). The inherent dangers of unidirectional emotional bonds between humans and social robots. In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot ethics: The ethical and social implications of robotics* (pp. 205-221). MIT Press.

Turkle, S. (2011). *Alone together: Why we expect more from technology and less from each other*. Basic Books.

Z■otowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and challenges in human-robot interaction. *International Journal of Social Robotics*, 7(3), 347-360.