Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Intimacy Robots and Inward Decline: From Individual Symbolism to Civilisational Consequence

Aftab Haider

aftabhaider516@gmail.com

Southwest University of Political Science & Law

Research Article

Keywords: intimacy robots, artificial companions, moral deskilling, symbolic-consequences argument, social solidarity, collective action

Posted Date: November 4th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10764109/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Philosophical work on intimacy robots has kept to two registers: the welfare of the individual user, and the symbolic wrong done to the classes of person such artefacts represent. Danaher's symbolic-consequences argument, the fullest statement of the critical case, concedes that the downstream effects of that symbolism are hard to establish, and recommends an experimental posture. This paper takes the concession as its problem. The debate has no macro-social level of analysis, and without one it cannot say which consequences to look for. Drawing on the sociology of solidarity, Vallor's account of moral deskilling, and the older literature of civilisational decline, it sets out three mechanisms by which individually rational uses of intimacy technology could aggregate into collectively costly outcomes: demographic subtraction, relational deskilling, and the habituation of desire to objects that cannot reciprocate. Each is put as a conditional with its defeaters named, not as a prediction. Tested against current evidence, including a nationally representative survey of adolescent companion-AI use and experimental work finding that AI companions do reduce loneliness, the third mechanism proves best supported analytically and the first least supported empirically. The result is a framework for aggregation, not a forecast of collapse; the paper closes by saying what would disconfirm it.

1. Introduction

The ethics of sexual and romantic technology has become a substantial research programme. Since Levy's *Love and Sex with Robots* (2007) placed the question on the agenda, and more decisively since the appearance of Danaher and McArthur's *Robot Sex* (2017), a recognisable set of positions has crystallised: a permissive case grounded in individual welfare and sexual autonomy, a critical case grounded in symbolic representation and the reinforcement of objectifying norms, and a sceptical middle that questions whether either side has the evidence its claims require. Döring, Mohseni and Walter's scoping review of the field (2020) found ninety-eight publications on sex robots between 1993 and 2019, of which only thirty-two were peer reviewed, the majority were theoretical, and—this is their most consequential finding—not one was an observational or experimental study using an actual sex robot or doll as stimulus material.

That configuration of a rich normative literature resting on a thin empirical base is not itself an objection to the literature. It is, however, a constraint on what the literature can be asked to do, and it has left one question conspicuously underdeveloped. Both the permissive and the critical case operate at the level of the individual or the dyad. McArthur (2017, pp. 31–45) argues that sexbots will benefit particular users. Petersen (2017, pp. 155–171) asks whether the artefacts themselves could be wronged. Richardson (2015) and Sparrow (2017) argue that the design and use of such machines expresses and reinforces a demeaning conception of women. Danaher (2017a, pp. 103–131) formalises this last family of arguments as the symbolic-consequences argument and subjects it to the criticism that its symbolism is often removable and, where not removable, reformable—so that what will finally matter are the *consequences* of the symbolism, which he acknowledges are "likely to be difficult to ascertain." Nyholm

and Frank (2017, pp. 219–244) approach the question from the opposite direction, asking not whether such machines are harmful but whether the relationship they offer could be the genuine article.

This paper begins from that acknowledgement. If the critical case ultimately rests on consequences, then the debate needs an account of how consequences aggregate: how a large number of individually defensible private uses could produce a socially costly outcome, through what pathway, on what timescale, and subject to what conditions. No such account has been offered. The symbolic-consequences argument gestures at social effects—attitude change, norm erosion—but treats them as an empirical question to be settled later, without specifying the social theory that would tell us where to look. The result is a debate in which critics assert diffuse harm and defenders correctly reply that no mechanism has been specified.

I aim to specify one. The argument has three parts. First, I identify what the existing literature has and has not established, and locate the gap precisely (Section 2). Second, I recover a body of social theory that was built to answer exactly this kind of question—the relationship between intimate association and the durability of social orders—and I take the objections to it seriously, since much of it is dated, ideologically freighted, and methodologically weak (Section 3). Third, using that theory, I construct three candidate aggregation mechanisms and test each against current evidence (Sections 5–7), finding that they are unequally supported: the mechanism most often asserted in public debate, demographic decline, is the least defensible, while the mechanism least discussed, the habituation of desire, is the most.

There are two things that need to be clarified about what is being claimed here so that the point is easier to understand and there is no confusion about the meaning. This paper does not predict civilisational collapse, and the phrase “inward decline” in my title names a hypothesis about a mechanism, not a forecast about an outcome. Nor does it recommend prohibition. What it offers is a framework at a level of analysis the debate currently lacks, together with a specification of the evidence that would tell for or against it. An earlier treatment of these questions confined the argument to a single religious tradition and could therefore be set aside as sectarian; the present paper is an attempt to state the case in terms that do not depend on any particular confessional commitment, which requires abandoning the appeal to authority that a single tradition permits.

2. What the Literature Has Established, and What It Has Not

2.1 Levy’s forecast and what was actually startling about it

Levy’s *Love and Sex with Robots* (2007) is routinely cited as the origin point of the debate and just as routinely misdescribed. It is worth being accurate about what the book argues, because both its critics and its admirers tend to summarise it by its conclusion—that human–robot marriage will be legal in Massachusetts by roughly 2050—while ignoring the structure of the argument that gets there.

Levy's method is analogical and, in its own terms, conservative. He surveys the psychological literature on human attraction and identifies the factors that reliably generate affection between persons: proximity, repeated exposure, perceived similarity, reciprocal self-disclosure, physical attractiveness, and, importantly, the perception of being liked in return. He then argues that none of these factors requires that the object of affection possess consciousness, only that it exhibit the relevant behaviour. If attraction is elicited by cues rather than by metaphysical properties, and if the cues are engineerable, then affection for machines follows without any need to solve the hard problem of machine mentality. The same structure governs his treatment of sex: he catalogues the long history of human sexual use of objects and argues that the addition of responsiveness is a difference of degree.

Read this way, nothing in Levy is shocking, and the frequent claim that the book scandalised its readership is misleading. Humans have used objects for sexual purposes throughout recorded history; the sex-doll industry long predates Levy; and the academic study of human–machine attachment was already established through Reeves and Nass's work on media equation effects and through Turkle's earlier ethnographies. What was genuinely novel in Levy was narrower and more interesting: the claim that the *affective* relationship, not merely the sexual one, would be substitutable, and that this substitution would attract legal recognition. That is a claim about institutions, not about desire, and it is the part of Levy's forecast that remains untested.

It is also the part that his critics have largely declined to engage. And it is worth noting, against Levy rather than for him, that his forecast has been partially vindicated in a form he did not anticipate. Levy expected embodied androids. What has actually arrived at scale is disembodied conversational companionship—text and voice systems with persistent memory and simulated affective investment—which satisfies most of the psychological conditions his own argument identified while dispensing with the engineering problem he thought was central. This matters for the analysis that follows, and I return to it in section 4.

2.2 The critical case and its most developed statement

The critical literature divides into two strands, and Danaher's editorial introduction to the volume (2017b, pp. 3–14) is a useful guide to how the division was understood at the point the field consolidated. The first strand, associated with Richardson (2015, pp. 290–293), argues from analogy with prostitution: sex robots are designed and marketed as entities whose function is availability without reciprocity, and in reproducing that asymmetry in a manufactured form they extend rather than replace a demeaning conception of the sexual partner. Richardson's argument has been criticised, in my view successfully, on two grounds. Danaher, Earp and Sandberg (2017, pp. 47–71) show that the analogy with prostitution is both empirically contestable and normatively unstable, since it depends on a characterisation of sex work that many sex workers reject, and that Richardson's proposed campaign against sex robots targets a technology whose design features are contingent rather than essential. Devlin (2018) adds that the assumption that such machines will be built exclusively in the image of a compliant woman reflects the current market rather than any necessity of the technology.

The second strand, which I take to be stronger, argues from representation rather than analogy. Sparrow (2017, pp. 465–477) considers a robot designed to simulate refusal and asks what the act of using it expresses. No person is injured; the artefact has no interests. Sparrow's claim is that the wrongness lies in what the act represents and in what the willingness to perform it reveals about the agent's character, and that a society which manufactures such representations is implicated in the attitudes they express. Danaher (2017a) generalises this into the symbolic-consequences argument: the objection to intimacy robots is that they symbolise a troubling attitude toward consent and toward the status of sexual partners, *and* that this symbolism has bad consequences.

Danaher's own criticism of the argument he formalises is the core of this study. He argues that the troubling symbolism is often removable—a robot can be designed to be capable of refusal, or marketed without the compliance framing—and where not removable, reformable, since symbolic meanings are conventional and can shift. What remains, therefore, is the consequences limb, and about this he is candid: the consequences "are likely to be difficult to ascertain," which is why he recommends an explicitly experimental approach to the technology's development.

2.3 The question of mutual love, and why it matters here

A third position, developed by Nyholm and Frank (2017, pp. 219–244), deserves separate treatment because it supplies a criterion the present argument uses. Rather than assessing harm, they ask whether mutual love between a human and a robot is possible, and they answer by extracting three requirements from the philosophy of love: that the parties be a good match, in the sense of being suited to each other's characteristics; that each value the other in their distinctive particularity rather than as a bearer of substitutable properties; and that each be committed to the other steadfastly, so that the relationship persists through changes that would give a purely satisfaction-seeking party reason to leave. Their conclusion is carefully hedged: mutual love with a robot is not impossible in principle, but it is hard to achieve, and building machines capable of satisfying these conditions might address some of the concerns critics have raised. Nyholm (2020) develops the surrounding scepticism about attributing to current systems the properties that friendship and love presuppose.

Two features of their analysis matter for what follows. The first is that the second and third conditions—particularity and steadfastness—are conditions the market has no incentive to satisfy. A product optimised for user satisfaction is optimised against steadfastness, since steadfastness consists precisely in remaining through the phases when the relationship is unsatisfying; and it is optimised against particularity, since a system that values the user in their unrepeatable specificity is a system that cannot be sold identically to anyone else. Nyholm and Frank's conditions are therefore not merely difficult, as they suggest, but commercially disfavoured. The second feature is that their framework identifies, without quite naming it, the specific competence that section 6 argues is at stake: the capacity to sustain commitment across dissatisfaction is a skill, and it is exercised only where the other party can be dissatisfying.

2.4 The gap

What is missing is worth stating precisely, because the gap is not in evidence, which everyone acknowledges, but in theory.

Suppose we grant Danaher everything. Suppose the symbolism is removable in principle. There remains a question that neither the symbolic-consequences argument nor its critics address: what *kind* of consequence should we be looking for, and at what level of social organisation would it appear? The literature's implicit answer is attitude change in individual users, measurable by the methods used to study pornography effects—does use of a compliant artefact make a person more likely to treat human partners instrumentally? That is a legitimate question, and it is the question the empirical literature is slowly moving toward. But it is not the only possible answer, and it may not be the most important one.

Consider an alternative. A technology can impose serious social costs without changing anyone's attitudes at all, if it changes what people *do* in a way that erodes a shared capacity that no individual is responsible for maintaining. On this model, the relevant harm is not that the user of an intimacy machine becomes a worse person, but that a large number of people, each acting reasonably, collectively withdraw from a practice whose maintenance is a precondition of goods they continue to want. The structure is that of a commons problem, and its distinguishing feature is that it produces bad outcomes without bad actors. Nothing in the existing debate is pitched at this level.

It is worth stating precisely how this differs from the three comparable programmes now active. The symbolic programme (Richardson 2015; Sparrow 2017; Danaher 2017a) locates the wrong in what the artefact represents, and is therefore vulnerable to redesign: change the representation and the objection lapses. The welfare programme (McArthur 2017; Di Nucci 2017; De Freitas et al. 2026) assesses effects on the individual user, and is therefore unable in principle to register costs that fall on no user in particular. The relational-possibility programme (Nyholm and Frank 2017; Nyholm 2020) asks whether the goods of love are achievable with an artefact, which is a question about the artefact's capacities rather than about the practice's social distribution. Each is doing legitimate work, and none is pitched at the level at which a commons problem would appear.

The claim of this study is accordingly modest in one respect and substantial in another. It is modest in that it adds no new normative principle and reports no new data. It is substantial in that it supplies the missing level of analysis, and in doing so changes what the existing arguments imply: an objection that survives redesign, that does not depend on identifying a harmed party, and that specifies observable predictions. Doing this requires social theory about the relationship between intimate association and the durability of social orders. Such theory exists, but it is old, and much of it is bad. Section 3 is an attempt to extract what is usable.

3. Intimate Association and Social Order: A Contested Inheritance

3.1 The classical convergence and what it actually shows

A striking feature of the pre-modern literature on political order is the recurrence, across traditions with no contact, of the claim that the household is prior to the political community.

Aristotle's version is the most carefully argued. In *Politics* I.2, he derives the *polis* through a developmental sequence: the pairing of male and female and the master–slave relation constitute the household, several households form the village, and several villages the city (Aristotle 1998, 1252a24–1252b30). The claim that the household is "prior" is doing specific work. Aristotle distinguishes priority in time from priority in nature, and holds that the city is prior *in nature* to the household and the individual, because the whole is prior to the part and only within the city can a human being achieve the completed function that defines the species (1253a19–29). The household's priority is therefore *temporal and generative*: it is the site at which citizens are produced and first habituated. This is a weaker claim than the one often attributed to Aristotle, and the weaker claim is the one that matters here. Aristotle is not saying the family is the most important institution. He is saying it is the institution through which the dispositions required by the others are transmitted, which is a claim about developmental dependence.

Cicero's formulation in *De officiis* I.17.54 is more concise and rhetorically stronger: the first bond of society is that of husband and wife, the next that of parents and children, then the shared household, which he calls the *seminarium rei publicae*, the seed-bed or nursery of the commonwealth (Cicero 1913, I.17.54). The metaphor is agricultural and generative, and it makes the same developmental point. The Confucian *Great Learning* organises the same dependence as a chain of cultivation running from the ordering of the self through the regulation of the family to the ordering of the state and peace in the world; the family occupies the hinge between private cultivation and public order (Confucius 1971, sections 4–5). Ibn Khaldun's *Muqaddimah* supplies the most explicitly sociological version: *ʿaṣabiyya*, the group feeling that enables collective action and therefore political power, originates in kinship and is progressively diluted as a dynasty urbanises and prospers, so that dynasties characteristically exhaust themselves within three or four generations (Ibn Khaldun 1958, vol 1, bk 2, sections 7–8; vol 1, bk 3, sections 11–12).

What does this convergence establish? Less than the rhetorical use often made of it suggests. Four traditions agreeing is not four independent confirmations, for at least two reasons. First, all four are describing societies in which the household was in fact the principal unit of production, education, welfare and defence, so the observed dependence may be a feature of pre-industrial social organisation rather than of human societies as such. Second, all four write from within patriarchal orders whose legitimation depended on precisely this claim, which gives us reason for caution about the direction of inference. The convergence is evidence that the household was load-bearing in the societies these authors observed. It is not, by itself, evidence that intimate association is load-bearing under conditions where the state, the market and the school have assumed most of the household's former functions.

That is a serious objection and it should be conceded. What survives it is the narrower Aristotelian claim: whatever else households do or fail to do, they remain the primary site at which human beings acquire the dispositions—patience, mutual accommodation, the capacity to sustain obligation under conditions

of inconvenience—that wider forms of cooperation presuppose. That claim is falsifiable and, as this study explains in section 6, has some contemporary support.

3.2 The twentieth-century sociology of family decline, handled sceptically

The mid-twentieth-century literature that explicitly linked family form to civilisational trajectory is more directly relevant and considerably less reliable, and I want to be explicit about its weaknesses rather than borrow its authority.

Zimmerman's *Family and Civilization* (1947) argues that both Greece and Rome moved through a sequence from the "trustee" family, through the "domestic" family, to the "atomistic" family—small, private, unstable, disinclined to reproduce—and that the atomistic phase coincided with civilisational decline. Zimmerman offers a list of symptoms of terminal family decay, among them the loss of the presumption of marital permanence, falling birth rates, and the disparagement of parenthood. The comparative sweep is impressive and the framework is not without descriptive value. But Zimmerman's method is essentially the assembly of illustrative material in support of a thesis held in advance; he does not consider whether family form might be a consequence rather than a cause of the political and economic changes he describes; and the work has been taken up primarily by authors with a prior commitment to its conclusions, which is a reason for additional caution rather than for confidence.

Unwin's Sex and Culture (1934) is even less convincing, and it is cited here mainly to be set aside. Unwin's claim—that a society's "social energy" varies with the strictness of its pre-nuptial and post-nuptial sexual restraint, and that relaxation is followed by decline within roughly three generations—rests on a coding of eighty tribal societies against categories he devised himself, with no inter-rater reliability, no control for the colonial provenance of his ethnographic sources, and an outcome variable ("social energy") that is operationalised largely by whether Unwin regarded the society's achievements as impressive. The correlation he reports is an artefact of the coding scheme. Sorokin's account of "sensate" culture (1956) has the same difficulty in a milder form: the diagnostic categories are constructed so that the conclusion follows.

The honest position is therefore that the twentieth-century literature supplies a *hypothesis* about the relationship between intimate association and social durability, and does not supply evidence for it.

3.3 The modern transformation, which the classical account cannot accommodate

The classical account treats the intimate bond as an institution with a public function. Giddens (1992) argues that the defining feature of late-modern intimacy is precisely that this is no longer true. The "pure relationship" is one entered and sustained for the satisfactions it provides to the parties, evaluated reflexively and continuously, and dissolved when it ceases to satisfy; its legitimacy derives from the

quality of the relationship itself rather than from any external function. Bauman (2003) describes the same shift more pessimistically, as the assimilation of human bonds to the logic of consumer goods.

If Giddens is right that this transformation is substantially complete, then intimacy robots do not introduce a new logic. They arrive within a logic already established by the middle of the twentieth century, and they are its instrument rather than its author. This is the strongest form of the causal-modesty objection, and it constrains what can be claimed: the argument of this paper cannot be that intimacy technologies commodify intimacy, since on Giddens's account that work was done long before. What can be claimed is something narrower and, I believe, more defensible—that these technologies remove the last constraint that the pure relationship retained, namely that it still required another person, whose independent will imposed limits on how thoroughly the relationship could be optimised for one party's satisfaction.

4. The Artefact: Distinguishing Mediation from Substitution

Analytical precision about the object of study is unusually important here, because public discussion runs together two technologies with very different diffusion curves.

Embodied sex robots remain marginal. The devices actually on sale are best described as animatronic dolls with conversational software; production runs are small, prices are high, and Döring, Mohseni and Walter (2020) confirm that no controlled study has yet used one as stimulus material. Any argument that rests on the prevalence of embodied intimacy robots is arguing about a hypothetical.

Conversational companions are not marginal. They are mainstream, and among adolescents they are close to universal. Common Sense Media's nationally representative survey of 1,060 US teenagers aged 13–17, conducted in April and May 2025, found that 72% had used an AI companion at least once and 52% qualified as regular users, interacting at least a few times a month (Common Sense Media 2025, p. 1). Eight per cent reported using companions for romantic or flirtatious interaction, and 9% described a companion as a friend or best friend. Laestadius and colleagues' grounded-theory analysis of 582 posts from the Replika user community documents relationships of considerable affective depth, including emotional dependence marked by users' sense that the system had needs of its own requiring attention (Laestadius et al. 2024).

The object of analysis, then, should be the affective substitute rather than the sexual prosthesis, and this reframing has a cost for the argument this study makes: it detaches the case from the demographic mechanism, which requires sexual substitution, and attaches it to the relational and dispositional mechanisms, which do not.

The distinction that carries the analytical weight is between technologies that mediate a bond between persons and technologies that substitute for one of the persons. The letter, the telephone, the personal advertisement and the dating application are all mediating technologies: they lower the cost of contact between people who then have to deal with each other. Their characteristic failure mode is a poor match.

A substituting technology has a different structure, because the friction it removes is not the friction of finding a partner but the friction constituted by the partner's independent will. This is why the standard reassurance, that every communications technology has attracted a moral panic and the panics were wrong, does not transfer. The technologies on that list were mediating technologies. The inference from "panics about mediating technologies were unfounded" to "panic about a substituting technology is unfounded" requires the premise that the distinction makes no difference, and that premise is what is at issue.

5. First Mechanism: Demographic Subtraction, and Why It Is the Weakest

The mechanism most frequently asserted in public argument runs as follows. Human reproduction depends on a chain: a desire for intimacy leads to partnership, and partnership leads, in some proportion of cases, to children. A technology that satisfies the desire without producing the partnership interrupts the chain at its first link, and a satisfied desire does not seek further. Aggregated, this subtracts from the population of people who might have had children.

The demographic context is not in dispute. The United Nations' 2024 revision of *World Population Prospects* reports a global total fertility rate of 2.25 births per woman, with 131 of 237 countries and areas—55%—below the replacement level of approximately 2.1, and projects that the global rate will reach replacement around 2036 (United Nations 2024). This is a substantial change, and it is occurring faster than earlier projections anticipated.

Nevertheless, this mechanism is the weakest of the three, and the reasons are worth setting out clearly because the argument of this paper is often assumed to depend on it. First, the causal path is long, and every link is contested. Fertility decline is overdetermined: female education, urbanisation, the opportunity cost of childbearing, housing costs, contraceptive availability and the collapse of child labour all contribute, and the relative weight of each is disputed among demographers. Adding a marginal factor to an overdetermined system is not obviously informative.

Second, the mechanism requires sexual substitution at scale, and section 4 established that embodied sexual substitutes are rare while affective substitutes are common. Affective substitution has no direct reproductive implication. It may have an indirect one, if companion use displaces partner-seeking, but that is a further claim requiring separate evidence, and the evidence currently available cuts against it: 80% of teenage companion users report spending more time with friends than with AI companions, and only 6% report the reverse (Common Sense Media 2025, p. 3).

Third, the direction of causation is unclear even if a correlation were found. Declining rates of partnership formation and sexual activity among young adults are documented across several high-income countries and predate the availability of conversational companions. If companion use is higher among those already withdrawn from partner-seeking, the technology is a symptom.

I therefore state this mechanism as a conditional with specified defeaters rather than as a claim. *If* affective substitutes become sufficiently satisfying that they displace partner-seeking rather than supplementing it, and *if* displaced partner-seeking is not compensated by later formation, then a marginal fertility effect follows. The conditional is defeated if companion use is found to be complementary to human relationship-seeking rather than substitutive, which is what the current adolescent data suggest, or if users are drawn disproportionately from those who would not have formed partnerships in any case.

One clarification, since the point is often misunderstood in both directions. Nothing in this analysis requires the view that a lower population is bad. The decoupling of sexuality from reproduction was, for Firestone (1970), a condition of women's liberation, and she was substantially right about that. The question here is not whether fewer births is a harm but whether a technology that operates on the *desire for partnership*, rather than on the decision to have children, changes the character of the choice—since every optimistic response to low fertility, whether through migration, later childbearing or productivity growth, assumes that the underlying orientation toward partnership persists and is merely deferred. That assumption, not fertility itself, is what a substituting technology puts in question.

6. Second Mechanism: Relational Deskilling

The second mechanism appears better supported and can be formulated more rigorously because an established concept already exists for it.

Vallor (2015, pp. 107–124) adapts the sociological literature on economic deskilling to moral life. Her argument is that moral competences are skills in a substantive sense: they are acquired through repeated practice under conditions of difficulty, they degrade without exercise, and they can be displaced by technologies that perform their function without requiring their exercise. Deskilling is not a matter of attitudes. A person may fully endorse the value of a competence and still lose it, simply by having fewer occasions to practise it. Vallor's examples include the moral skills involved in caregiving, in military judgement, and in sustaining attention to another person.

Applied to intimacy, the argument has a precise structure. The competences involved in sustaining a relationship with another person—tolerating misattunement, repairing rupture, accommodating a will that does not defer, remaining present through another's distress when it is inconvenient—are acquired only by exercising them, and they are exercised only where the other party's independence makes them necessary. A technology that supplies the goods of companionship while removing the occasions for these competences does not need to corrupt anyone in order to reduce their prevalence. It only needs to be more comfortable than the alternative, often enough, for long enough.

Turkle's formulation anticipated this: digital connection and sociable robots "may offer the illusion of companionship without the demands of friendship" (Turkle 2011). The observation that matters is that the demands are constitutive rather than incidental. Fromm's *The Art of Loving* (1956) makes the same point in the vocabulary of skill: love is an activity requiring discipline, concentration and patience, and

Fromm's distinction between the receptive orientation that says "I am loved because I am lovable" and the active one that says "I am loved because I love" is a distinction between two developmental positions, only one of which requires practice. Vallor's concept supplies the mechanism; Turkle supplies the phenomenology; Fromm supplies the account of what is being lost.

Nyholm and Frank's third condition allows the mechanism to be stated with more precision than either Turkle or Fromm permits. Steadfast commitment, on their account, is constitutive of love rather than an optional addition to it: a party who remains only while satisfied is not loving but consuming. Steadfastness, however, can only be practised where there is something to be steadfast through. A system engineered so that the user is never disappointed removes not merely the occasion for the skill but the possibility of acquiring it, and it does so as a design objective rather than as a side effect. This is the sense in which the concern here differs from a general anxiety about convenience. It is not that these systems make relationships easier; many good technologies do that. It is that the specific difficulty they remove is the one whose management constitutes the competence.

The scaling claim is where Ibn Khaldun becomes relevant. His argument is not that families produce solidarity by magic, but that the disposition to bear costs for others is acquired in relations of close association and then extended, with diminishing intensity, to progressively wider circles. Putnam's account of declining associational membership (2000) describes the same dependence at the level of civil society, and the US Surgeon General's advisory on loneliness and isolation (Murthy 2023) documents the health consequences of its erosion. If Ibn Khaldun's developmental claim is right, degrading the primary training ground has effects that appear at scales far from where the degradation occurs, and appear late.

Now the counter-evidence, which is substantial and which I believe limits the argument significantly.

De Freitas and colleagues (2026) report a series of studies, including experimental work, finding that AI companions do reduce loneliness—on a par with interaction with another person, and more effectively than comparison activities such as watching video—and that users underestimate this effect in advance. Self-disclosure and distraction do not account for the result. This is careful work, and it establishes something the critical literature has generally denied: the consolation is real, not illusory. Any argument that these technologies merely simulate relief is refuted by this evidence.

The adolescent survey data cut in both directions. Against the deskilling hypothesis: 67% of teenagers find conversations with AI companions less satisfying than conversations with real friends, only 10% find them more satisfying, and 80% spend more time with friends (Common Sense Media 2025, pp. 2–3). Teenagers, on this evidence, are not confusing the artefact with the relationship. For the hypothesis: 33% report having chosen to talk to an AI companion instead of a person about something important or serious, 9% cite ease relative to talking to real people as a reason for use, and 33% report using companions to practise a social skill subsequently deployed in life (Common Sense Media 2025, pp. 2–3). That last figure is genuinely ambiguous. It is equally consistent with upskilling—rehearsal that

transfers—and with deskilling, since the rehearsed version omits the feature that makes the real thing difficult.

So the honest conclusion for this section is a divided one. The mechanism is coherent, has a developed theoretical basis in Vallor, and receives partial support from the substitution data. It is not established, the consolation it warns against is genuine, and the strongest available evidence on adolescent users shows them discriminating clearly between artefact and person. What would settle it is longitudinal evidence on relational competence among heavy users, which does not yet exist.

7. Third Mechanism: The Habituation of Desire

The third mechanism is the least discussed and appears to be the best supported analytically, although it is the hardest to operationalise.

The relevant Kantian argument is not the one usually cited. The formula of humanity in the *Groundwork* (Kant 1998, 4:429) prohibits treating persons merely as means, and it does not apply to artefacts, which is why appeals to it in this debate tend to collapse: a machine has no humanity to be violated. The applicable argument is in the *Metaphysics of Morals*. Kant holds that we have no direct duties to animals, since duties are owed only to persons, but that cruelty to animals is nonetheless wrong as an indirect duty *to oneself*, because such treatment "dulls his shared feeling of their suffering and so weakens and gradually uproots a natural predisposition that is very serviceable to morality in one's relations with other people" (Kant 1996, 6:443).

This is precisely the argumentative form the intimacy-robot case requires, and it has the considerable advantage of being Kant's own solution to a structurally identical problem. The artefact cannot be wronged. The practice can still be impermissible, because of what it does to a disposition that is "serviceable to morality" in relations with those who can be wronged. Kant's argument has been much discussed—Kain (2018) and others have examined whether the indirect-duty framework is stable, and it is not without difficulties. But it locates the wrong in the right place: in the practitioner's dispositions rather than in the patient's interests. It should be noted also that this reframing answers Danaher's removability objection. If the wrong lies in the effect of the practice on the agent's dispositions rather than in what the artefact symbolises, then redesigning the artefact's symbolism does not by itself address it; what would matter is whether the *interaction structure*—non-reciprocity, unlimited availability, absence of independent will—remains intact, and that structure is much less easily reformed than the symbolism, since it is the product's value proposition.

Nussbaum's analysis of objectification (1995, p. 257) supplies the specification of the interaction structure. She identifies seven ways of treating a person as an object: instrumentality, denial of autonomy, inertness, fungibility, violability, ownership, and denial of subjectivity. An intimacy robot satisfies all seven by design and without wrongdoing, since it is genuinely an instrument, genuinely lacks autonomy, is genuinely fungible and owned, and genuinely has no subjectivity to consider. Nussbaum's own view is that objectification is not always wrong and can be benign within a relationship of mutual

respect; the relevant point here is not that the artefact is objectified but that the artefact instantiates, in a single object, the complete set of features that Nussbaum identified as the components of an attitude which is wrong when directed at persons.

Whether the resulting practice affects dispositions toward persons is an empirical question that Aristotle's account of habituation frames but cannot answer. His claim in *Nicomachean Ethics* II.1 is that states of character arise from activities of the corresponding kind—we become just by doing just acts and temperate by doing temperate ones—and that legislators therefore attend to habituation because it makes all the difference (Aristotle 2009, 1103a14–1103b25). Aristotle is describing acquisition through practice; whether practice on a simulacrum transfers to the genuine case is not a question he addresses, and the analogous question about pornography has proved difficult to settle empirically after four decades of attempts.

Sparrow's contribution (2017) is to insist that the question of transfer is not the only question. Even in the absence of behavioural transfer, the willingness to use an artefact designed to simulate refusal tells us something about the agent, and the manufacture of such artefacts implicates the manufacturer and the market. This is a claim about expressive meaning rather than about consequences, and it survives the failure of transfer studies.

I part company with Richardson (2015), however, on the further claim that the technology industrialises a specifically prostitutional model of relationship. Danaher, Earp and Sandberg (2017) are right that the analogy is unstable and that the design features it depends on are contingent. But the structural residue of Richardson's argument survives their criticism: whatever the analogy, these artefacts are not delivered value-neutral. They embody a determinate conception of what a partner is for, and the conception is one of availability without reciprocity—not because designers are malicious but because reciprocity is precisely the friction the product exists to remove.

8. Aggregation, Objections, and Limits

8.1 The aggregation structure

The three mechanisms share a form worth naming, since it is what the existing debate lacks and what this study is mainly offering.

For an individual, on a given occasion, choosing an artificial companion over the effort of a human relationship may be entirely reasonable: cheaper, more reliable, less risky, and—per De Freitas et al. (2026)—genuinely consoling. The competences described in section 6 and the dispositions described in section 7 are not, however, individually owned. They are maintained by being practised across a population, and their availability to any one person depends on their prevalence in others. A person who withdraws from the practice imposes no detectable cost on anyone and receives a real benefit. The aggregate of many such withdrawals is a reduction in a shared capacity that no one was assigned to maintain.

This is a commons structure, and its diagnostic features are worth stating because they distinguish this argument from the moralising it might be mistaken for. There is no villain: no one need have a bad motive or a false belief. There is no detectable individual harm, so no individual has standing to complain. The cost is deferred, appearing at a temporal distance from the choices that produce it. And the individually rational choice remains rational even after the aggregate cost is understood, which is why exhortation is not a remedy.

Framed this way, the argument does not license interference with any individual's private conduct, and I propose none. It licenses attention to an aggregate, which is what social theory is for.

8.2 The strongest objections

The compassionate objection. For some people—the severely disabled, the profoundly socially anxious, the bereaved, the very old who have outlived their peers—human intimacy is not merely difficult but foreclosed. McArthur (2017) and Di Nucci (2017) argue that such users stand to benefit, and De Freitas et al. (2026) now provide evidence that they do. I accept this without qualification. It bears on the analysis in the following way: an argument about aggregate effects of population-scale normalisation is not an argument against therapeutic use by those for whom the alternative is not a human relationship but nothing. The two claims are compatible, and treating them as competitors is a category error on both sides.

The historical objection. Devlin (2018) shows that the sex-technology panic is a recurring genre with a poor forecasting record. The reply is in section 4: the technologies on the list of failed panics were mediating technologies, and the inference to a substituting technology requires the premise that the distinction is immaterial. I would humbly add that the record is not uniformly reassuring. Social media was widely expected to increase connection; the current weight of evidence on adolescent wellbeing is contested but not favourable.

The liberal objection. What competent adults do privately, harming no one, is not a proper subject of public argument. I accept the premise and deny the inference. Describing an aggregate is not regulating an instance.

Causal modesty. This is the objection I accept most fully, and section 3.3 has already conceded its strongest form: on Giddens's account, the commodification of intimacy substantially preceded these technologies, which are its instrument rather than its cause. The claim I retain is correspondingly weaker—that a substituting technology removes the last structural constraint on the pure relationship, namely the other person's independent will.

8.3 Limitations

Four limitations are stated.

The first is empirical. Döring, Mohseni and Walter's finding that no controlled study has used an actual sex robot as stimulus material (2020) applies to this study as much as to those it criticises. The

mechanisms described in this study are theoretically motivated conjectures. The companion-AI evidence I have used is better—nationally representative survey data and experimental work—but it is cross-sectional or short-term, and every mechanism here concerns effects that would appear over years.

The second is that the strongest available evidence partly disconfirms the strongest version of the thesis. De Freitas et al. (2026) show the consolation is real. The Common Sense data show adolescents discriminating clearly between artefact and friend. A paper that suppressed these findings would be worse than one that concedes them.

The third concerns the social theory in section 3. The classical convergence is weaker evidence than its rhetorical use suggests, and the twentieth-century decline literature is methodologically poor. I have tried to retain only the narrow developmental claim.

The fourth is that the mechanisms are not independent, which makes them hard to test separately, and their interaction is asserted rather than modelled.

8.4 What would disconfirm this

The argument should be held to the standard it asks of others, so I specify the evidence that would tell against it. Longitudinal evidence that heavy companion users show stable or improved relational competence, and stable or higher rates of human partnership formation, over five to ten years, would defeat the deskilling mechanism. Evidence that companion use is concentrated among those with no realistic alternative would defeat the aggregation claim, since withdrawal from a practice one could not have participated in subtracts nothing. Evidence of complementarity—that companion use increases rather than displaces human relationship-seeking, as the adolescent time-allocation data weakly suggest—would defeat the substitution premise on which all three mechanisms depend. On the other hand, the mechanisms would be supported by evidence of declining relational competence among heavy users independent of prior disposition, and by evidence that the interaction structure identified in section 7 predicts dispositional change better than the artefact's symbolism does.

9. Conclusion

The literature on intimacy robots has been conducted at the level of the user and the represented class. Danaher’s formalisation of the critical case concedes that what finally matters are the consequences of the artefact’s symbolism, and that these are hard to ascertain. This paper has argued that they are hard to ascertain partly because the debate has no macro-social level of analysis, and has tried to supply one: three aggregation mechanisms by which individually reasonable uses could produce collectively costly outcomes, stated as conditionals with specified defeaters, and tested against the evidence now available.

The results are uneven, and the unevenness is the finding. The demographic mechanism, which dominates public discussion, is the least defensible: the causal chain is long, the technology it requires is not the one that has diffused, and the available data on substitution cut against it. The relational

mechanism has a rigorous basis in Vallor's account of moral deskilling and partial support from substitution data, but the best current evidence shows both that the consolation these systems offer is real and that adolescent users discriminate clearly between artefact and friend. The dispositional mechanism is the strongest analytically: Kant's indirect-duty argument in the *Metaphysics of Morals* locates the wrong in the practitioner rather than the patient, which is where it must be for an artefact that cannot be wronged, and it has the further advantage of surviving the removability objection, since what would have to be reformed is the interaction structure rather than the symbolism—and the interaction structure is the product.

Two remarks by way of closing. The first concerns a framing this study has deliberately avoided. It is conventional, in discussions of this kind, to invoke Postman's antithesis between Orwell and Huxley: that the one feared what we hate and the other what we love (Postman 1985, foreword). The contrast is memorable and it is also a simplification of both novels, since *Brave New World* contains a great deal of coercion—prenatal conditioning, hypnopaedia, exile for dissent (Huxley 1932, chs 1–2, 17)—and Huxley's own later reflections are considerably more concerned with propaganda and the techniques of mass persuasion, and less with pleasure, than the antithesis allows (Huxley 1958, chs 5–7). Invoked as a heuristic, it names something real about technologies that operate through attraction rather than compulsion. Invoked as an argument, it substitutes a literary contrast for the analysis of a mechanism, and that substitution is one of the things this study has tried to avoid.

The second is a matter of the appropriate register for a conclusion. The mechanisms described here may operate weakly, may be offset, or may be defeated by evidence I have specified. What can be said with more confidence is narrower: that a technology which removes the other person's independent will from an intimate relationship is a different kind of object from a technology that connects people, that the difference is analytically tractable, and that the debate has not yet been conducted at the level where its consequences, if any, would appear.

Declarations

Competing interests:

The author declares no competing financial or non-financial interests relevant to the content of this article.

Generative AI declaration

No AI tools were used in writing this study. A premium version of Grammarly was used as an AI-assisted tool for grammar polishing.

Ethical Approval:

Not Applicable.

Informed Consent:

Not Applicable.

Conflict of Interest:

The authors have no relevant financial or non-financial

Funding:

No funding was received for this project.

Author Contribution

A;H Conceptualisation, Methodology, Writing original draft, Supervision,

References

1. Aristotle (1998) Politics (trans: Reeve CDC). Hackett, Indianapolis
2. Aristotle (2009) The Nicomachean ethics (trans: Ross D, rev: Brown L). Oxford University Press, Oxford
3. Bauman Z (2003) Liquid love: on the frailty of human bonds. Polity, Cambridge
4. Cicero MT (1913) De officiis (trans: Miller W). Loeb Classical Library. Harvard University Press, Cambridge
5. Common Sense Media (2025) Talk, trust, and trade-offs: how and why teens use AI companions. Common Sense Media, San Francisco. https://www.commonsensemedia.org/research/talk-trust-and-trade-offs-how-and-why-teens-use-ai-companions. Accessed 30 July 2026
6. Confucius (1971) The great learning (Daxue). In: Legge J (ed) (trans) Confucius: Confucian analects, the great learning, and the doctrine of the mean. Dover, New York
7. Danaher J (2017a) The symbolic-consequences argument in the sex robot debate. In: Danaher J, McArthur N (eds) Robot sex: social and ethical implications. MIT Press, Cambridge, pp 103–131
8. Danaher J (2017b) Should we be thinking about robot sex? In: Danaher J, McArthur N (eds) Robot sex: social and ethical implications. MIT Press, Cambridge, pp 3–14

9. Danaher J, Earp B, Sandberg A (2017) Should we campaign against sex robots? In: Danaher J, McArthur N (eds) Robot sex: social and ethical implications. MIT Press, Cambridge, pp 47–71
10. Danaher J, McArthur N (eds) (2017) Robot sex: social and ethical implications. MIT Press, Cambridge
11. De Freitas J, Uğuralp AK, Oğuz-Uğuralp Z, Puntoni S (2026) AI companions reduce loneliness. J Consum Res 52(6):1126–1149. https://doi.org/10.1093/jcr/ucaf040
12. Devlin K (2018) Turned on: science, sex and robots. Bloomsbury, London
13. Di Nucci E (2017) Sex robots and the rights of the disabled. In: Danaher J, McArthur N (eds) Robot sex: social and ethical implications. MIT Press, Cambridge, pp 73–88
14. Döring N, Mohseni MR, Walter R (2020) Design, use, and effects of sex dolls and sex robots: scoping review. J Med Internet Res 22(7):e18551. https://doi.org/10.2196/18551
15. Firestone S (1970) The dialectic of sex: the case for feminist revolution. William Morrow, New York
16. Fromm E (1956) The art of loving. Harper & Row, New York
17. Giddens A (1992) The transformation of intimacy: sexuality, love and eroticism in modern societies. Polity, Cambridge
18. Huxley A (1932) Brave new world. Chatto and Windus, London
19. Huxley A (1958) Brave new world revisited. Harper & Brothers, New York
20. Ibn Khaldun (1958) The Muqaddimah: an introduction to history, 3 vols (trans: Rosenthal F). Princeton University Press, Princeton
21. Kain P (2018) Duties regarding animals. In: Denis L (ed) Kant's Metaphysics of Morals: a critical guide. Cambridge University Press, Cambridge, pp 210–233
22. Kant I (1996) The metaphysics of morals (trans: Gregor M). Cambridge University Press, Cambridge. (original work published 1797)
23. Kant I (1998) Groundwork of the metaphysics of morals (trans: Gregor M). Cambridge University Press, Cambridge. (original work published 1785)
24. Laestadius L, Bishop A, Gonzalez M, Illenčík D, Campos-Castillo C (2024) Too human and not human enough: a grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. New Media Soc 26(10):5923–5941. https://doi.org/10.1177/14614448221142007
25. Levy D (2007) Love and sex with robots: the evolution of human-robot relationships. HarperCollins, New York
26. McArthur N (2017) The case for sexbots. In: Danaher J, McArthur N (eds) Robot sex: social and ethical implications. MIT Press, Cambridge, pp 31–45
27. Murthy VH (2023) Our epidemic of loneliness and isolation: the U.S. Surgeon General's advisory on the healing effects of social connection and community. U.S. Department of Health and Human Services, Washington, DC. https://www.hhs.gov/sites/default/files/surgeon-general-social-connection-advisory.pdf. Accessed 30 July 2026

28. Nussbaum MC (1995) Objectification. Philos Public Aff 24(4):249–291. https://doi.org/10.1111/j.1088-4963.1995.tb00032.x
29. Nyholm S (2020) Humans and robots: ethics, agency, and anthropomorphism. Rowman & Littlefield, London
30. Nyholm S, Frank LE (2017) From sex robots to love robots: is mutual love with a robot possible? In: Danaher J, McArthur N (eds) Robot sex: social and ethical implications. MIT Press, Cambridge, pp 219–244
31. Petersen S (2017) Is it good for them too? Ethical concern for the sexbots. In: Danaher J, McArthur N (eds) Robot sex: social and ethical implications. MIT Press, Cambridge, pp 155–171
32. Postman N (1985) Amusing ourselves to death: public discourse in the age of show business. Viking Penguin, New York
33. Putnam RD (2000) Bowling alone: the collapse and revival of American community. Simon and Schuster, New York
34. Richardson K (2015) The asymmetrical relationship: parallels between prostitution and the development of sex robots. ACM SIGCAS Comput Soc 45(3):290–293. https://doi.org/10.1145/2874239.2874281
35. Sorokin PA (1956) The American sex revolution. Porter Sargent, Boston
36. Sparrow R (2017) Robots, rape, and representation. Int J Soc Robot 9(4):465–477. https://doi.org/10.1007/s12369-017-0413-z
37. Turkle S (2011) Alone together: why we expect more from technology and less from each other. Basic Books, New York
38. United Nations (2024) World population prospects 2024: summary of results. Department of Economic and Social Affairs, Population Division, United Nations, New York. https://population.un.org/wpp/. Accessed 30 July 2026
39. Unwin JD (1934) Sex and culture. Oxford University Press, London
40. Vallor S (2015) Moral deskilling and upskilling in a new machine age: reflections on the ambiguous future of character. Philos Technol 28(1):107–124. https://doi.org/10.1007/s13347-014-0156-9
41. Zimmerman CC (1947) Family and civilization. Harper and Brothers, New York