

Should we campaign against sex robots?

John Danaher¹, Brian D. Earp², and Anders Sandberg³

1. National University of Ireland - Galway

2. The Hastings Center Bioethics Research Institute (Visiting Scholar)

3. University of Oxford, Future of Humanity Institute

Authors' personal copy. Forthcoming book chapter (penultimate draft). Please cite as:

Danaher, J., Earp, B. D., & Sandberg, A. (forthcoming). Should we campaign against sex robots? In J. Danaher & N. McArthur (Eds.) *Robot Sex: Social and Ethical Implications* [working title]. Cambridge, MA: MIT Press. Draft available online ahead of print at: https://www.academia.edu/25063138/Should_we_campaign_against_sex_robots.

* Note: this is a draft paper; further revision may still be made prior to final publication. Feedback at this stage is welcome.

Abstract: In September 2015 a well-publicised *Campaign Against Sex Robots* (CASR) was launched. Modelled on the longer-standing *Campaign to Stop Killer Robots*, the CASR opposes the development of sex robots on the grounds that the technology is being developed with a particular model of female-male relations (the prostitute-john model) in mind, and that this will prove harmful in various ways. In this chapter, we consider carefully the merits of campaigning against such a technology. We make three main arguments. First, we argue that the particular claims advanced by the CASR are unpersuasive, partly due to a lack of clarity about the campaign's aims and partly due to substantive defects in the main ethical objections put forward by campaign's founder(s). Second, broadening our inquiry beyond the arguments proffered by the campaign itself, we argue that it would be very difficult to endorse a general campaign against sex robots unless one embraced a highly conservative attitude towards the ethics of sex, which is likely to be unpalatable to those who are active in the campaign. In making this argument we draw upon lessons from the campaign against killer robots. Finally, we conclude by suggesting that although a generalised campaign against sex robots is unwarranted, there are legitimate concerns that one can raise about the development of sex robots.

1. Introduction

September 2015 saw the launch of the *Campaign Against Sex Robots* (CASR).ⁱ Spearheaded by Kathleen Richardson of Leicester De Montfort University and Erik Brilling of the University of Skövde, the CASR models itself on the longer-standing *Campaign to Stop Killer Robots*, which was itself organized to preemptively ban fully autonomous weapons. CASR opposes the development of sex robots (for a sense of what we mean by this term, see the chapter in this volume by McArthur & Danaher) on the grounds that they are ‘potentially harmful and will contribute to inequalities in society’.ⁱⁱ The campaign’s founders believe that ‘an organized approach against the development of sex robots is necessary’.ⁱⁱⁱ

The campaign received considerable media attention after its launch.^{iv} There also appears to be some support for the view it espouses. David Levy, one of the leading proponents of the development of sex robots (Levy 2007), was organizing the *2nd International Conference on Love and Sex With Robots* prior to the launch of the campaign. The conference was due to be held in Malaysia in November 2015, but was cancelled shortly after the launch of the CASR as a result of opposition from Malaysian government officials. Although there is unlikely to be a direct causal link between the campaign and the cancellation, its cancellation does nevertheless demonstrate an appetite and receptiveness to the position being advocated.

Given this apparent potential for the CASR to gain meaningful traction, it seems important to assess its merits. That is what we attempt to do in this chapter. We do so in two main parts. First, we try to examine the objections to the development of sex robots that have actually been put forward by the CASR. We conclude that these objections are ultimately unpersuasive, although we suggest that the degree of (un)persuasiveness depends upon the actual aim of the campaign, which is currently not well-specified. Second, we try to broaden the focus by asking whether there could ever be a good reason to object to the development of something like this technology.^v To do this, we draw upon lessons from the *Campaign to Stop Killer Robots*. We conclude that we are unlikely to come up with good reasons to preemptively reject the development of sex robots, unless we adopt a highly conservative approach to the ethics of sex, which many will find problematic for other reasons. This does not mean that there are no legitimate concerns one can have about the development of sex

robots; it simply means that the concerns are not best addressed by adopting an organized campaign against their development.

More generally, the analysis we present in this chapter has implications for how societies should respond to controversial new technologies, based upon a range of factors including risk (and type) of harm, prospect of benefit, and different ways of pursuing regulation. Lessons learned here, therefore, should apply well beyond this specific debate over sex robots and autonomous weapons.

2. What is the argument behind the *Campaign Against Sex Robots*?

We start by engaging the campaign on its own terms. To do this, we turn to the position paper authored by Kathleen Richardson (2015).^{vi} This paper sets out a series of objections to the development of sex robots, which are grounded in an analogy between sex robot-human relations and prostitute-john^{vii} relations. This way of framing things suggests that the main intellectual grounding for the CASR (at least in its current form) is an argument from analogy. Richardson makes this explicit at the outset of her paper, commenting in particular on the model of human-robot relations that is adopted in the work of David Levy (2007):

In his book, Sex, Love and Robots [sic]^{viii} David Levy proposes a future of human-robot relations based on the kinds of exchanges that take place in the prostitution industry. Levy explicitly creates ‘parallels between paying human prostitutes and purchasing sex robots’ [p.194] . I want to argue that Levy’s proposal shows a number of problems, firstly his understanding of what prostitution is and secondly, by drawing on prostitution as the model for human-robot sexual relations, Levy shows that the sellers of sex are seen by the buyers of sex as things and not recognised as human subjects. This legitimates a dangerous mode of existence where humans can move about in relations with other humans but not recognise them as human subjects in their own right.

(Richardson 2015, 290)

Although Richardson does not spell out her objection to sex robots in formal terms, we propose that it can be reconstructed in the following manner:

- (1) Prostitute-john (or sex worker-client) relations are ethically problematic (for a number of reasons, but particularly due to objectification of the sex worker).
- (2) Sex robot-human relations are being modeled on sex worker-client relations and so will share or exemplify similar properties to those relations.
- (3) Therefore, sex robot-human relations will be ethically problematic (by analogy).
- (4) Therefore, we ought to campaign against sex robots.

This argument is not formally valid. Analogical arguments tend not to be (Walton 2013); they are informal defeasible arguments, dependent on the strength of the similarity between the two cases being used. There is also something of a gap between the first conclusion (3) and the second conclusion (4). Nevertheless, it is our view that the reconstruction is fair to the position put forward in the paper by Richardson because the prostitute-john analogy is front and centre in her analysis; moreover, the gap between the first conclusion and the second conclusion is something we wish to highlight in our critique.

Granting that this reconstruction fairly captures the argumentative centerpiece of the CASR, we will now attempt to show that this argument suffers from three major flaws, each of which serves to undermine the CASR in its current form.

2.1 – The first flaw: The objective is unclear

The first problem with the CASR's argument is the gap between the two conclusions. Why is it that ethically problematic properties in human-sex robot relations support an organised campaign against the development of sex robots? The inference, as stated, is unwarranted. Now, there may be a way to bridge this gap, but the strength of the inference from (3) to (4) will depend on the objective of the proposed campaign. Unfortunately, the CASR are not clear about their intended

objective. Until we have that clarity, it will be nearly impossible to tell whether or not the campaign is something that is worthy of our support. Why is this?

Broadly speaking, when it comes to the development of sex robots (or any potentially harmful technology), there are three main attitudes that we can adopt. We can be entirely *prohibitive* in our outlook, favouring the preemptive suppression of the technology. This is the attitude adopted by the *Campaign to Stop Killer Robots* in response to the development of fully autonomous weapons systems.^{ix} This prohibitive attitude could extend to the criminalization of those who create, distribute and use the technology (Danaher 2015), though criminalization may not be necessary.

Alternatively we can be entirely *libertarian* in our outlook, favouring complete freedom in the development and use of the technology. Finally, in between these two extremes, we can be *regulative* in our attitude, favouring some oversight and intervention into the development and use of the technology, but falling short of a complete ban. This regulatory attitude could take many forms, ranging from the strong prohibition of certain types of sex robot while tolerating or even possibly promoting others, to the weaker crafting of industry ethical standards, advertising guidelines, safety protocols and the like.

In short, there is a spectrum of possible responses to the prospect of sex robot development. The strength of the argument in favour of the CASR depends almost entirely on where along this spectrum the objective for the campaign lies. Clearly, it does not lie at the extreme libertarian end: there is something about the development of sex robots that the CASR finds objectionable. But then where between total prohibition and weak regulation does it lie? The answer makes a difference to the burden of proof that the CASR's argument needs to discharge. If they favour weak forms of regulation – which, for example, might just ask manufacturers of these robots to keep in mind various ethical issues that are raised by the technology, and make some effort to address those issues – then the burden of proof will be relatively low. Indeed, if there are plausible ethical concerns, then we could probably all agree that these should be factored into the regulation of this technology. If, on the other hand, they favour total prohibition, the burden of proof is much higher. In liberal political regimes, there is (or should be) a general presumption against total prohibition of new technologies, particularly when it comes to a technology designed

for personal use. The ethical issues identified by the campaign would need to be extremely compelling to warrant preemptive suppression.

So where along the spectrum do the campaign's objectives lie? Unfortunately, this is unclear. The CASR's website clearly states that an 'organized approach against the development of sex robots' is required;^x on their 'About' page they list several unqualified objections to the development of sex robots; and in their position paper and support materials they appeal to the preemptive model of the *Campaign to Stop Killer Robots*. All of this suggests that their objective is close to the total prohibition end of the spectrum. But there is also evidence of a more moderate view. The 'About' page, for example, states that the campaign is in favour of 'ethical technologies that reflect human principles of dignity, mutuality and freedom' and that they want 'computer scientists and roboticists to examine their own conscience when asked to provide code, hardware or ideas to develop this field'. Richardson's reliance on the 'prostitute-john' analogy in her position paper also suggests that if sex robots could engender an alternative and more egalitarian relationship with humans then her opposition might wane.^{xi} Furthermore, in one media report, Richardson is recorded as saying that we should simply 'examine what it means' to create such technologies, not completely ban them; although in the very same piece she is also reported to have called for a ban (Boran 2015). This suggests some equivocation on the part of the campaign's most visible founder as to its overarching objective.

This equivocation is problematic and may also explain some of the media appeal of the CASR. As other chapters in this book make clear, ethical objections to the development of sex robots are not uncommon. The fact that the particular set of objections raised by the CASR has seen such widespread coverage is, at least *prima facie*, puzzling. But now we may have an explanation: by ostensibly favouring the extreme of prohibition, the campaign can generate much initial media interest, and then when pushed on this extreme view it can retreat to a more modest and reasonable position, a position that may not actually warrant an organized approach against the development of sex robots (notwithstanding the stated aims of the campaign). If so, this would constitute a so-called 'motte and bailey' tactic that is not uncommon in debate (Shackel 2005). This is just a speculation. The more serious point is that until the objective is clarified the CASR should not win widespread support.

2.2 – The second flaw: It paints a misleading view of sex work

The second problem with the CASR's argument is its reliance on the sex worker-client analogy. This analogy may not be essential to the campaign's case – a point to which we return below – but it features heavily in Richardson's position paper. Hence it is worth taking seriously. The claims that seem to be^{xii} driving the argument are that sex worker-client relationships have several bad-making properties and that these bad-making properties will be shared and (thereby) reinforced and/or normalised by sex robot-human relationships.

In advancing this position, Richardson adopts an almost entirely negative view of sex work. Indeed, she objects to 'sex work' terminology in and of itself as well as to related discourse, seeing it as an attempt to normalise sex work as a reasonable choice in the labour market, when in fact, according to Richardson, it is highly problematic. In support of this view, she cites various studies showing that prostitution often occurs 'in the absence of consent', that violence and trafficking are common in the industry, and that many of the workers are young girls or young women (between the ages of 13-25). She does not engage at all with the position of those who do not share her uniformly negative view of prostitution/sex work, including some who have offered constructive articulations of alternative viewpoints in the recent literature (e.g. Moen 2014). Instead, she seems simply to take for granted that her position should be accepted as the only reasonable one that could be plausibly advanced.

Her main objections to sex work are that it is built around an asymmetrical power-relation between the client and the prostitute (one in which the client has the balance of power); that the work itself is highly gendered (primarily women, and some men, providing sex to primarily male clients);^{xiii} and that the commercial relationship involves the denial of subjectivity (and hence effective objectification) of the sex worker. In other words, according to this view, the sex worker is treated as a plaything that can be used and manipulated at the whim of the client, not as a human being with feelings, agency and autonomy that ought to be respected. To back these objections up she quotes from a study by Farley, Bindley and Golding (2009), which included interviews with clients making claims such as these:

*'Prostitution is like masturbating without having to use your hand',
'It's like renting a girlfriend or wife. You get to choose like a catalogue',
'I feel sorry for these girls but this is what I want'*

(Farley et al 2009, 8)

Each of these views seems to reinforce the notion that the sex worker is being treated as little more than an object and that her subjectivity is being denied. The client is elevating his status and failing to empathise with the prostitute: he is substituting his fantasies for her real feelings. Insofar as that is indeed what is going on in a client-sex worker relationship, and if that is representative of such relationships, then we agree that it would be a serious problem. Among numerous other reasons, the failure or inability to empathise is often associated with higher rates of crime and violence (Baron-Cohen 2011).

We thus seem to have two main lines of objection to sex work. The first is concerned with violence and lack of consent within the industry; and the second is concerned with objectification and the lack of empathy this entails. These are the bad-making properties of sex work that support the first premise of Richardson's argument. Leaving aside the question of whether these bad-making properties carry over to the case of sex robots, there are at least two problems with the objections raised regarding the existence (and nature) of these bad-making properties that may prevent the overall analogy from getting off the ground.

The first problem is that both objections are derived from a highly selective view of sex work. Richardson paints a wholly negative picture of the industry. But there are many prominent sex work researchers who challenge that picture. For example, Sanders et al. (2009) provide an overview of the empirical literature on sex work in which they develop a far more nuanced account. For instance, on the issue of violence in the industry, they note that it is a 'minority' of clients who commit violence, and that the vast majority of commercial transactions take place without violence or other incident (Sanders et al 2009, 44), similar to the situation regarding non-commercial sexual relationships. Likewise, on the question of objectification and asymmetrical power relations, they note that much of this opposition is grounded in a conservative view of sexual ethics that fails to consider the possibility of emotionally rich and intimate sex occurring outside of the confines of a long-term monogamous

relationship (Sanders et al 2009, 83). Furthermore, they argue that the attitudes of clients towards sex workers are often far more complex and multi-dimensional than Richardson supposes:

...[G]eneral understandings of sex work and prostitution are based on false dichotomies that distinguish commercial sexual relationships as dissonant from non-commercial ones. Sanders (2008b) shows that there is mutual respect and understanding between regular clients and sex workers, dispelling the myth that all interactions between sex workers and clients are emotionless. There is ample counter-evidence (such as Bernstein 2001, 2007) that indicates that clients are 'average' men without any particular or peculiar characteristics [who are] increasingly seeking 'authenticity', intimacy and mutuality rather than trying to fulfil any mythology of violent, non-consensual sex.

(Sanders et al. 2009, 84)

With respect to non-commercial sexual relationships, such as might exist between 'ordinary' couples (i.e., couples for whom the explicit exchange of money is not directly associated with the sexual encounter, although other types of less overt exchanges may sometimes occur), it must be acknowledged that these are neither inherently nor exclusively models of 'authenticity, intimacy and mutuality' either, notwithstanding their non-commercial nature. While this is arguably far from optimal, what it suggests is that the quality of sexual relationships in terms of respect, power-asymmetries, etc., fall on a spectrum both within and outside of the commercial context; hence, much more work would be needed to show that there is something unique to an exchange-based sexual relationship that strictly entails the undesirable qualities being discussed, and which does so in a way that is significantly worse than comparable (i.e., casual) non-exchange-based sexual relationships.

We do not raise these points to paint a rosy and Pollyannaish view of sex work. We are only trying to highlight the need for greater nuance than Richardson seems willing to provide. It is simply not true that all forms of prostitution involve the troubling features she identifies, while at the same time, many forms of non-commercial sex do involve those features. Given that that is the case, the proposed

analogy with sex robots appears to be on much weaker initial footing than Richardson assumes.

This brings us to the second criticism. Even if Richardson is right about both (a) the bad-making properties of prostitution, and (b) the view that these properties will “carry over” to the case in which the “prostitute” is a non-human robot (thus potentially normalising the problematic attitudes that may be associated with such bad-making properties; but see discussion below), there is no strong reason to think that such an outcome would be sufficient to warrant a total prohibition on sex robots, nor even, perhaps, an organized campaign to oppose them. For better or worse, denials of subjectivity, asymmetries of power, and other objectionable qualities or outcomes are rife throughout the capitalistic workplace. While this might form the basis of a general critique of capitalism (such as a Marxist critique), it does not appear, in most other cases, to motivate, much less successfully ground, a *prohibition* of the activities or services that have such undesirable features. For example, many service workers in our economies regularly have their subjectivity denied by their clients: one may fail to care about the feelings of the massage therapist giving them a massage, for example, or about those of the carpenter repairing their deck. But while we may have very good reason to attempt to change the norms of behavior concerning such uncompassionate treatment of service workers, it would seem a strange and rather ham-fisted approach to attempt to ban the service industries altogether. Instead, a change in regulation and social attitude may be all that is appropriately called for.

We must be clear that we are not suggesting that there are no morally-relevant differences between the (typical) relationship between a client and her massage therapist (or between a homeowner and her carpenter) and the (typical) client-prostitute relationship. Instead, we are only pointing out that the presence of even gravely undesirable features in a given type of exchange relationship (up to and including denials of subjectivity and problematic asymmetries of power) is not normally seen as being a sufficient justification for banning the relationship altogether; rather, regulations that seek to mitigate the undesirable features that do exist may worthily be pursued. On top of this, and perhaps even more importantly, there is reason to think that many of the negative features of client-prostitute relationships in particular could actually be directly caused (or worsened) by its prohibition. This is because prohibition “cannot alleviate the background conditions

that contribute to people’s motivation to sell sexual services in the first place” (Earp and Moen 2015, 54).

As one of us has argued elsewhere (drawing on the work of Luke Semrau 2015):

... there is an important distinction to be drawn between “being pressured to sell X” and “being pressured, with the option to sell X.” If someone is being directly pressured to sell [something] (or become a prostitute, etc.), then prohibiting the selling of X could at least in principle be of some help. [But] if someone is experiencing a more general pressure (such as extreme economic insecurity), but has a number of ways—including, but not limited to, selling X—to begin to relieve this pressure, then prohibiting the selling of X is actually more likely, all else being equal, to make this person even worse off. This is because it would remove (or drive underground, and therefore make more dangerous) at least one otherwise viable option for “making ends meet.” As a consequence, the person who was considering selling X, and who would do so if it were not prohibited, must now turn to an even less desirable option (as judged by them) to relieve the more general pressure.

(Earp and Moen 2016, 55)

Indeed, some of the problems of prohibition can arguably be seen in the case of violence and trafficking. Insofar as prostitution is criminalized, sex workers cannot avail themselves of the various protections that are afforded to most other kinds of workers and they may also fail to report what happens to them for fear of adverse legal consequences. This is why many sex worker activists — who are in no way unrealistic about the negative features of the job — favour legalisation and regulation, as opposed to outright prohibition (for further discussion, see Murphy 2015).

In sum, although there may certainly be harms associated with prostitution that are different to and/or worse than the harms that are associated with other forms of work, *simply* referring to those harms without showing how they are unique, uniquely bad, or would be reduced rather than exacerbated by a ban, is not enough to show that prohibition would be warranted.

2.3 – The third flaw: Application of the analogy is unpersuasive

Fortunately for Richardson, proving the badness of prostitution may not be essential to her argument. All the CASR really needs to show is that sex robot-human relationships will have bad-making properties which may be similar to (or intensified versions of) the supposed bad-making properties of sex worker-client relations. If she can show this, then the possible disanalogies between the two scenarios may not undermine the case for organized opposition to the development of sex robots. Indeed, it may be that these disanalogies *support* opposition in the case of sex robots.

One obvious disanalogy between the two cases is that sex workers are persons and hence of high moral status; whereas sex robots are unlikely to be persons, at least for the foreseeable future. This disanalogy can cut both ways. On the one hand, it suggests that sex robots cannot be moral victims and hence we need not be concerned about their treatment at the hands of their users; on the other hand, the fact that they are not moral victims means that we don't need to worry about the negative impact of restrictive and prohibitive policies on their well-being (as we do with human sex workers). This may cause us to favour an organized campaign against their development, particularly if we think that (1) the harms arising from their use could carry over to real human victims (through normalizing problematic attitudes or norms of behavior), and (2) the putative benefits of their use are minimal.

This raises an important interpretive question about the intention behind Richardson's analogy with sex work. Is her claim that, just as the treatment of and attitudes toward prostitutes is bad, so too will be the treatment of and attitudes toward sex robots? Or is it that the development of sex robots will increase the demand for human prostitution and/or encourage users of sex robots to treat real human (females) more as objects? Richardson's paper supports the latter interpretation. At the outset, she states that her concern about sex robots is that they:

[legitimate] a dangerous mode of existence where humans can move about in relations with other humans but not recognise them as human subjects in their own right.

(Richardson 2015)

The key phrase here seems to be ‘in relations with other humans’, suggesting that the worry is about how we end up treating one another, not how we treat the robots themselves. This is supported in the conclusion where she states:

In this paper I have tried to show the explicit connections between prostitution and the development and imagination of human-sex robot relations. I propose that extending relations of prostitution into machines is neither ethical, nor is it safe. If anything the development of sex robots will further reinforce relations of power that do not recognise both parties as human subjects.

(Richardson 2015, S.4)

Again, the emphasis in this quote seems to be on how the development of sex robots will affect inter-human relationships, not robot-human relations. Let us reflect this in a modified version of premise (2):

- (2*) Sex robots will add to and reinforce the bad-making properties of prostitution (i.e. they will encourage us to treat other humans with a lack of empathy and thereby exacerbate existing, harmful gender/power inequalities).

Richardson supports this premise by focusing on the work of David Levy (2007), who draws explicit parallels between the development of sex robots and prostitution. Levy’s suggestion is that the relationship between a user and his/her sex robot can be akin to the relationship between a client and a prostitute. Levy is quite explicit about this and spends a good portion of his book looking at the motivations of those who purchase sex and how those motivations might transfer onto sex robots. Interestingly, Levy is far more nuanced in his discussion of this literature than Richardson is inclined to be, highlighting some of the complexity in the motivations of clients and also including a discussion of female clients of male sex workers (a perspective that is entirely absent from Richardson’s analysis). In any event, the inference Richardson draws from Levy’s work is that the development of sex robots is proceeding along the lines that Levy imagines and hence we should be concerned about its potential to reinforce the bad-making properties of prostitution.

Richardson identifies two major criticisms of her claim. The first, mentioned above, holds that if robots are not persons then there is nothing wrong with treating them as objects/things which we can use for our own pleasure. In other words, the technology is a morally neutral domain in which we can act out our fantasies. The second criticism points to the potentially cathartic effect of these technologies. If people act out negative or violent sexual fantasies on a robot, they might be less inclined to do so to a real human being. This suggests that sex robots may consequently help to *prevent* the bad things that Richardson worries about. This view is explored in more detail in other chapters in this book (McArthur, Strikwerda and Danaher)

Richardson has responses to both of these criticisms. She argues that technology is not a value-neutral domain. Our culture and our norms are reflected in our technology. We should be worried about how cultural meaning gets incorporated therein. Furthermore, she has serious doubts about the catharsis argument. She points to the historical relationship between pornography and prostitution. Pornography has now become widely available, but this has not led to a corresponding decline in prostitution nor, in the case of child pornography, abuse of real children. On the contrary, prostitution actually appears to have increased while pornography has increased. The same appears to be true of the relationship between sex toys/dolls and prostitution:

The arguments that sex robots will provide artificial sexual substitutes and reduce the purchase of sex by buyers is not borne out by evidence. There are numerous sexual artificial substitutes already available, RealDolls, vibrators, blow-up dolls etc. If an artificial substitute reduced the need to buy sex, there would be a reduction in prostitution but no such correlation is found.

(Richardson 2015)

Is this a robust defence of the bad-making properties of sex robot-human relations? Is there enough here to warrant an organized campaign against their development?

We have some doubts. The evidence adduced to show that sex robots will exacerbate harmful inter-human relationships is weak, and even if it is correct it does not support a strongly restrictive approach to the development of sex robots. At best, it supports a regulative approach (see analysis above). Furthermore, if we were to

adopt such a regulative approach we would need to be sensitive to both the merits and demerits of this technology and the costs of any proposed regulatory strategy. This is something that Richardson neglects because she focuses entirely on the negative. In this vein, we offer four responses to her argument, some of which target her support of premise (2*), others of which target the relationship between any putative bad-making properties of sex robots and the need for an organised ‘campaign’ against them.

First, Richardson’s primary support for premise (2*) - viz. that the prostitute-john relationship is reflected in the model of sex robot development used by David Levy — is weak. We have already noted that Levy is more sensitive to the complexity of sex worker-client relations than Richardson is. But even if he weren’t, the argument would not be very strong. True, Levy is a pioneer in this field and has a degree of influence, but that doesn’t mean that all sex robot developers are obliged to adopt his model. If we are worried about the relationship between the sex robot user and the robot, we can try to introduce standards and regulations that reflect a more positive set of sexual norms. For instance, the makers of Roxxxy (billed as the world’s first sex robot) claim to include a personality setting called ‘Frigid Farah’ with their robot.^{xiv} Farah will demonstrate some reluctance to the user’s sexual advances. One could argue that this reflects a troubling view of sexual consent: that resistance is not taken seriously (i.e. that ‘no’ doesn’t really mean ‘no’). Does this mean that one should oppose the development of sex robots *tout court*? Not necessarily. Instead, one could address the problem by regulating against this kind of setting and insisting that every sex robot is required to give positive, affirmative signals of consent. Although there are numerous problems that might be raised by the prospect of an ever-willing (robot) sexual partner (see Danaher, this volume, for a longer analysis), such an approach would at least plausibly reflect and reinforce a more desirable attitude toward sexual consent. But we need not settle that example to make the broader point: namely, that it is far from clear that there is anything inherent in the nature of sex robots, *qua* sex robots, that would entail the promotion of negative social attitudes toward women and sexual consent. Instead, we could demand and enforce specific design regulations that would promote a more positive set of attitudes. Maybe this is all Richardson really wants her campaign to achieve. But in that case, she is not really campaigning against the sheer development of sex robots; she is

rather campaigning for a better version of them. If so, she could rename her campaign ‘the Campaign for Positive Sex Robots’, which changes the framing of the debate considerably.

Second, we think it is difficult to make definitive claims about the likely link between the use of a future technology like sex robots and actions toward real human beings. Richardson’s claims about the correlation between pornography and an increase in prostitution will be relatively unpersuasive to someone who does not think that either pornography or sex work is unremittingly bad. What would be more persuasive is if she could prove that there was some correlation (and ultimately some causal link) between the increase in pornography/prostitution and the *mistreatment* of sex workers. The evidence on this seems to be mixed. The legalisation of sex work is sometimes associated with a decrease in mistreatment, and the related link between pornography and sexual violence more generally is deeply contested. Some studies show an increase;^{xv} some show a decrease;^{xvi} and some are mixed or uncertain.^{xvii} Danaher (2015) speculates that we may be landed in a similarly ambiguous position when it comes to evidence concerning a link between sex robot usage and real-world sexual aggression. Or, conversely, sex robots may be sufficiently different from pornography as to increase the real-world effect. But in which direction is unclear: it may embolden or satiate. It is simply too early and too difficult to tell. A precautionary approach does not imply a ban as the first step taken, but rather a need for more research and close observation (see Earp, Sandberg and Savulescu in press for a defence of this view).

Third, and perhaps most importantly, when thinking about the appropriate policy toward sex robots, or any new prospective technology, it is important that we weigh the good against the bad. As two of us have argued previously (Earp et al. 2014, 7):

... when discussing the possible hazards associated with some predicted future development—one has to remember that any new technology poses risks. This is true whether it is an anti-love pill, a powerful military weapon, or something more mundane. Hence the mere possibility that such a technology might be used for ill can never by itself constitute sufficient reason to reject it—however alarming such a possibility may be. Instead, the potential harms

that might accrue from the misuse of the technology must be weighed against the potential benefits that might accrue from its responsible use.

In addition, as Bostrom and Roache (2011) have argued,^{xviii} even the careful anticipation of possible benefits and harms is not sufficient to give a full analysis of the prudence of developing some new technology. Instead, these efforts must be complemented with a meaningful attempt to identify “potential supporting policies and practices that can alter the balance for the better” (144). As the philosopher C. A. J. Coady has stated:

If indeed there is insufficient knowledge of outcomes and consequences, or no social or institutional regulatory regime for prudent implementation of the innovations and for continuing scrutiny of their effects, or no room for overview of the commercial exploitation of the innovations, then ... critics [of new technologies] clearly have a point. [But] warnings can be heeded. [We can] insist on safeguards and regulation, both scientific and ethical.

(Coady 2009, 165, emphasis added)

In the case of sex-robots, specifically, there are several potentially good-making properties that would need to be factored into the discussion (McArthur, Di Nucci this volume). For example, there is the simple hedonistic argument: sex robots provide people with a way of achieving pleasurable states of consciousness. Whether this is a good argument or not, or whether increasing hedonistic states should count for very much when taken into consideration with other factors, is a complex question; but at the very least, it needs to be engaged with in a serious way when evaluating the likely value or disvalue for society of developing sex robots. Similarly, there is the distributive argument: for whatever reason, there are people in the world today who lack access to certain types of sexual experience, including people with certain kinds of disabilities (see Di Nucci – this volume; see also Earp and Moen 2015) and sex robots could make those experiences (or, at least, meaningful approximations of them) available to such people. There is also the argument that sex robots could ameliorate imbalances in sex drive between the partners in existing relationships (McArthur – this volume), or could add a desired kind of diversity to the sex lives of such couples, without involving human third parties (and the potential interpersonal

strife to which those third parties could give rise). It could also be the case that sex robots would allow for particular forms of sexual self-expression and identity to flourish (see generally Gupta 2012), including a pure sexual preference for machine-sex (see Hauskeller 2014 for a longer, more critical discussion of this). If that turned out to be the case, then, in the interests of basic sexual freedom and diversity, we should permit it. This is by no means an exhaustive list of positive potential attributes. It simply highlights the fact that there may be much good to the technology and this must be weighed against any putative negative features when determining the appropriate policy. Few new technologies are unalloyed goods; trade-offs must be carefully considered.

Fourth, and finally, when thinking about the appropriate regulative policy one also needs to think about the potential costs, both monetary and non-monetary, of that policy. We might agree that there are certain bad-making properties arising from human-sex robot relations, but it could be that any proposed regulatory intervention would do more harm than good overall. Regulation of pornography, for instance, has historically involved greater restrictions toward pornography depicting sexual minorities (e.g., gay and lesbian porn).^{xix} Regulatory intervention into sex robots may end up doing the same, targeting robots used and designed by sexual minorities. It may also be the case that policing the development and use of sex robots would require significant resources and extensive intrusions into our private lives. Are we willing to bear these costs? Less intrusive regulatory policies — e.g. ones that simply encourage manufacturers to avoid problematic stereotypes or norms in the construction of sex robots — would arguably be much more tolerable. Again, that may be all that Richardson wants. But if that is the case, the campaign she is leading needs to be much clearer about its aims.

2.4 – Interim conclusion

An interim summary is in order. The CASR’s primary argument against the development of sex robots turns on an analogy between sex robot-human relations and sex worker-client relations. The fear is that the former will reinforce the bad-making properties of the latter. We find that this analogy does not support the organized opposition to the development of sex robots, much less a full-fledged ban.

The analogy is premised on an overly negative view of sex work and the associated need for its prohibition; its application to the sex robots scenario is weak and speculative; and it also ignores the potential good-making properties of sex robots as well as the costs of organized opposition. At most, we may have a basis for regulation of sex robot development; but this seems to fall far short of the current rhetorical demands of the campaign.

3. Should we ever campaign against sex robots?

The arguments of the CASR may be unpersuasive, but it is still worth asking the broader question that their campaign might be seen to provoke: should we *ever* campaign against a technology of this sort? This is a question of considerable social importance. Artificial intelligence (AI) and robotics are on the rise. We can expect more and more robots to enter into daily social use, including in ethically loaded cases.^{xx} Perhaps there are legitimate grounds on which we can oppose the development of some of them. What is more, perhaps these grounds could apply to the case of sex robots.

In this section, we address these issues by considering the arguments against the development of fully autonomous weapons systems (AWSs), or ‘killer robots’. Since the CASR is modeled after the *Campaign to Stop Killer Robots*, this seems an obvious place to start. For context, AWSs are robotic systems that operate independently of their human creators, administering lethal force to enemy targets. There has been considerable social and academic opposition to the development of such robots, going beyond the campaign just mentioned. For example there is the Future of Life Institute’s (FLI’s) open letter on autonomous weapons systems, which is strongly critical of the development of such technology.^{xxi} Many people are persuaded that AWSs simply should not be allowed to exist, and that a preemptive ban on their development is in order. If arguments to that effect are indeed successful, they might also apply to the case of sex robots in ways that have not yet been raised in our discussion of the CASR.

In what follows, then, we survey some of the leading objections to the development of AWSs, treating the issue largely on its own terms. After taking the time to do so, we then return to the issue of sex robots, and conclude the chapter by

reflecting on some of the key lessons from this ‘killer robot’ analysis for the debate over the CASR.

3.1 – Arguments opposing the development of AWSs

There are three main families of argument against the development of AWSs. The first is pragmatically-oriented and is concerned with the *consequences* of deploying AWSs in lieu of human soldiers. Proponents of this objection are concerned with the practical limitations of the current and prospective technologies. For instance, they worry about the targeting systems of AWSs and their ability to adapt to dynamic battlefield conditions, in particular their ability to make fine-grained and context-sensitive distinctions between who is and is not an enemy combatant (Guarini and Bello 2012). The fear is that AWSs will be more indiscriminate in their administration of lethal force than human operators would be and hence responsible for worse outcomes. In making this argument, then, distinctions need to be drawn between AWSs, which operate entirely independently of human controllers, and teleoperated unmanned weapons systems, which are still under remote human control. One can favour the latter on the ground that their use poses less risk to human soldiers (Strawser 2010), while still opposing the former (e.g. Purves, Jenkins and Strawser 2015).

These pragmatically-oriented objections are important, and certainly fears about targeting systems and battlefield adaptability will need to be addressed. But the general utility of such objections is limited. As the underlying technology improves, the consequentialist fears dissipate. Other objections to the development of AWSs try avoid this technological contingency.

The second family of arguments is concerned with the need for moral agency and responsibility in warfare. They can be grouped under the general heading of ‘responsibility gap’ arguments. The most-widely discussed example comes from Robert Sparrow (2007). The gist of his argument is that the actions within a just war must be carried out by, or ultimately be capable of being traced to, the decisions of a responsible moral agent.^{xxii} Humans are capable of being responsible moral agents. AWSs are not. They lack the capacity for responsible agency and will always lack this

capacity. Hence, their use in a theatre of war opens up a responsibility gap: lethal force is used but no one is ultimately responsible for its use.

Even if we grant the moral premise of this argument, we are still owed some account of why it is that AWSs lack the capacity for responsible moral agency. Purves, Jenkins and Strawser (2015) present two such accounts. The first focuses on the *codifiability* of moral judgment. It maintains that AWSs will only be capable of exercising moral judgment if moral judgment is capable of codification. AWSs will have to be programmed and this programming will depend on well-defined rulesets that identify the key moral requirements that the AWSs will need to follow. These rulesets will then have to be reduced to a computer language.^{xxiii}

The problem here is that several leading normative theories claim that moral judgment is not ultimately codifiable in this manner. Seemingly robust moral principles admit an indefinite number of exceptions that cannot be anticipated in advance by human programmers; and some moral theorists claim that moral judgment is always *particularized* (Dancy 2013). Accordingly, Purves, Jenkins and Strawser's second account of why AWSs lack appropriate moral agency focuses on other mental constituents of responsibility. Following the views of leading just war theorists, it argues that actors within a war are not simply required to act in conformity with sound moral judgment; they must also act for the *right moral reasons*. In defending this view, Purves et al. use the example of a racist soldier, who kills the right enemy combatants but for the wrong reasons (Purves, Jenkins and Strawser 2015, 860). They argue that the deployment of such a soldier would be wrong. The problem for AWSs is that they will be incapable of acting for moral reasons. This claim builds upon objections to strong AI (the belief that you could create an artificial being with human-equivalent mental faculties). The leading accounts of what it means to act for a reason require a sophisticated internal mental architecture that AWSs will lack for the foreseeable future. Hence, they cannot act for moral reasons and so their deployment in a theatre of war is contrary to the requirements of just war theory.

Both of these defences of the responsibility-gap argument can be challenged. They rest on controversial metaphysical claims about the nature of mind and morality, as well as the prospects for strong AI. Furthermore – as the authors themselves acknowledge – important tradeoffs would arise if the robots achieved better battlefield outcomes than human soldiers *even while* failing to meet the requirements for moral

responsibility. Others have used this possibility to develop defences of the use of AWSs (e.g., Muller 2016; Simpson and Muller 2016).^{xxiv} Nevertheless, if the premises are sound, the arguments presented here provide strong moral reasons to object to the use of AWSs that are not contingent on the current state of the technology.

The final family of objections to AWSs focuses on the directness of the harm they can potentially cause. This is perhaps the most significant consideration when it comes to the merits of an organized campaign against their development. One thing that all participants to the debate about AWSs can agree upon is the *prima facie* wrongness of killing. In other words, everyone accepts that it is only permissible to kill a human being in a limited range of circumstances. Killing someone when these circumstances do not apply is a grave moral wrong. And even when the circumstances do apply it is still bad *for* the (involuntary) victim. A key property of AWSs is that their entire *raison d'être* is to exert lethal force against enemy targets. In other words, they are directly designed to do something that everyone agrees is a *prima facie* moral wrong or a deeply regrettable outcome. The hope is that they will become sophisticated enough to recognise when the circumstances that permit killing do in fact arise. But there is always the risk that they will misjudge or miscalculate and do something that is a serious moral wrong, or even enable moral wrongs on an unprecedented scale (for example, enabling totalitarian states to harm or coerce citizens with no possibility of soldiers' consciences stopping them). This is one reason why a preemptive campaign against the development of this particular kind of robot seems so compelling: if the campaign succeeds, it can *directly* prevent a *prima facie* wrong. We will return to this issue in our discussion of sex-robots near the end of this chapter.

Now, some people object to this line of reasoning. They argue that the alleged advantages of the preemptive ban are more illusory than real. AWSs are likely to be developed by unscrupulous people, they claim, whether we try to prevent this occurrence or not. A ban, they argue, would merely prevent research into making them as ethical as is reasonably achievable. A related argument is that counterfactual assessment of what would happen in the *absence* of AWSs is also important. Humans are probably going to use lethal force against each other anyway just as they always have. The advantage of AWSs, then, if they could be made sufficiently

technologically sophisticated, is that they could be more selective and less destructive than human actors. But this line of thinking, too, can be challenged. One obvious concern about the introduction of AWSs is that they can lower the threshold for launching lethal attacks: if no human soldiers will be harmed in the process, a commanding officer may be more likely to issue a command.

3.2 – Lessons for the *Campaign Against Sex Robots*

Now that we have some idea of the debate concerning AWSs, we should reflect on the implications of that debate for the CASR. Some of the arguments we have just surveyed seem relevant to the case of sex robots; others, less so. We proffer no final judgment on the merits of the campaign against AWSs here (although we note that one of us, Sandberg, has signed the FLI's open letter against their development; see above); instead, we simply wish to consider the lessons of that campaign for our primary topic. To do this, we shall assume, *arguendo*, that at least one of the preceding arguments reaches the threshold needed to warrant an organized and systematic campaign against the development of killer robots. Do similar considerations apply to the development of sex robots? We think not.

First, there are no equivalent *practical* concerns when it comes to the development of sex robots. The practical fears surrounding AWSs comes from the fact that they are intended to do harm and that inadequate programming or technical sophistication could lead to the wrong kind (or degree) of harm. Sex robots, by contrast, are not directly intended to cause harm; in fact, the opposite is true—they are intended to cause pleasure. Certainly the robots could malfunction and these malfunctions could give rise to health risks; or the robots could be used in dangerous, perhaps unsanitary, ways. But these practical concerns seem comparatively minor, to the point that they would not warrant preemptive organized opposition to the sheer development of the robots in question. At most, they would warrant the creation of appropriate industry standards as well as comprehensive user safety guidelines. Now, it could be argued that the pleasure-inducing effects of sex robots could also be addictive and hence a source of harm for those who become addicted to using them (along with their families and friends). This is a more serious concern but the

potential harm is indirect and speculative. Hence we defer to our third response, which we will come on to in just a moment.

Second, there is no equivalent need for moral agency and responsibility in the design of sex robots. Again, the opposite would appear to be true. The need for moral agency in the case of military robots stems from the morally fraught nature of their actions and the need for moral accountability in warfare. But we don't need sex robots to be morally accountable. Indeed, turning sex robots into moral agents would probably add to, rather than mitigate, ethical problems surrounding their use. For instance, one of the putative benefits of sex robots (discussed by McArthur and Di Nucci in this volume) is that they can facilitate positive sexual autonomy (i.e. desire to have sex) without interfering with another person's negative sexual autonomy (i.e. desire to avoid sex). If the robot is a moral agent, then this putative benefit could disappear. The robot would have to be given the same moral standing as any other moral agent. Their consent to a sexual act would become a serious issue, and the propriety of their use for sexual pleasure could no longer be taken for granted (cf. the contributions from Petersen and Geras in this volume). This doesn't seem like a suitable ground on which to favour organized opposition to their development.

Third, there are no equivalently direct moral harms that could be prevented by the campaign against sex robots. The harm from AWSs is direct and irreversible; the alleged harms from sex robots are speculative, indirect and in principle reversible. The argument from the CASR focuses on how the use of sex robots would impact on real women (both those who work in the sex industry and those who do not). The fear is that individual use of sex robots will distort the user's downstream interactions with real human beings and contribute to existing social problems arising from systematic inequality and oppression of women. These are certainly legitimate ethical concerns, and we agree that society should work hard to reduce systemic inequality, invidious sexual objectification (cf. Marino 2008), and the harmful erosion of empathy. But those problems stem from individual choices and social institutions, not from the sex robots themselves. As one of us has previously put it:

...even if one bought the arguments that sex robots are likely to induce bad social changes, these changes are occurring because of individual decisions and beliefs, as well as sociocultural institutions. There are many other levers that could be pulled to improve the situation of sex workers, women, or

people's attitudes to each other. Some of these levers may be far more powerful than a technology ban. Conversely, even a successful ban of sex robots may fail to reach the desired goal because of other technologies or intermediaries causing the undesired social changes. By acting against a possible contributor rather than the bad thing itself, effort is wasted.

(Sandberg 2015)

In our view, this is the crucial difference with the campaign against killer robots. By stopping killer robots you directly prevent a *prima facie* moral wrong. By stopping sex robots, however, you do not.^{xxv}

Direct harm arguments against sex robots are possible to make, but they are of a very different kind to the ones used by the CASR. They would hold that there exist *inherent* bad-making properties of human-robot sex that produce a direct—but possibly victimless—moral harm. Arguments in this vein might fault human-robot sex for being nonreproductive, for example, or for being a kind of inter-species sex (assumed to be intrinsically wrong); for lacking essential relationship properties that should exist during any sexual act (perhaps masturbation would be seen as wrong on this view as well); or for being an immoral form of hedonism. Proponents of this view could then argue that this direct and intrinsic (albeit, again, potentially victimless) moral harm could be prevented from materializing if we were to pre-emptively ban sex robots, just as preventing killer robots prevents a new moral harm.

Are such arguments convincing? To some people, they may be. However, they appear to hinge on a peculiarly conservative and vice-related view of sex that is unlikely to hold sway in many pluralistic societies. Thus, while certain individuals may feel motivated to refrain from using sex robots (should they become available) on the basis of these perceived direct moral harms, they would need to convince large swaths of the population of their conservative view in order to justify a sweeping change to policy. By contrast, AWSs are only legitimately going to be used by the state, rather than by individual persons, for societal goals rather than personal goals. That makes the rest of society proper moral stakeholders in decisions about their use and development (since the state should represent and act in their interest), and conversely they are to some extent responsible for what it does. That is not the case

with the personal use of sex robots, however, as the bedroom door insulates against most would-be stakeholders.

Thus, while one can easily imagine social conservatives and vice-oriented moralists supporting the CASR because of shared opposition to sex robots, they would be doing so on the basis of a fundamentally different sort of moral claim – that there is something *inherently* bad about human-robot sex – that would fit uneasily with the stated ethics of the CASR’s leading proponents.

4. Conclusion

Robots are going to form an increasingly integral part of human social life. Sex robots are likely to be among them. Though the proponents of the CASR seem deeply concerned by this prospect, we have argued that there is nothing in the nature of sex robots themselves that warrants preemptive opposition to their development. The arguments of the campaign itself are vague and premised on a misleading analogy between sex robots and human sex work. Furthermore, drawing upon the example of the *Campaign to Stop Killer Robots*, we suggest that there are no bad-making properties of sex robots that give rise to similarly serious levels of concern. The bad-making properties of sex robots are speculative and indirect: preventing their development may not prevent the problems from arising. Preventing the development of killer robots is very different: if you stop the robots you stop the *prima facie* harm.

In conclusion, we should preemptively campaign against robots when we have reason to think that a moral or practical harm caused by their use can best be avoided or reduced as a result of those efforts. By contrast, to engage in such a campaign as a way of fighting against—or preempting—indirect harms, whose ultimate source is not the technology itself but rather individual choices or broader social institutions, is likely to be a comparative waste of effort.

Notes

- ⁱ For all documentation and supporting materials relating to the campaign, see <http://www.campaignagainstsexrobots.com/> (accessed 21/1/16)
- ⁱⁱ Text taken from the CASR's 'About' page: <http://campaignagainstsexrobots.org/about/> (accessed 21/1/16)
- ⁱⁱⁱ Text taken from the CASR's 'About' page <http://campaignagainstsexrobots.org/about/> (accessed 21/1/16).
- ^{iv} For a sample of the media coverage, see the 'Press' page on the CASR website.
- ^v Here we follow the definitions of sex robot and having sex with a robot presented in earlier chapters by McArthur and Danaher and Migotti.
- ^{vi} The paper was published in the ACM SIGCAS newsletter, but a full text version can be found at: <http://campaignagainstsexrobots.org/the-asymmetrical-relationship-parallels-between-prostitution-and-the-development-of-sex-robots/> (accessed 21/1/16)
- ^{vii} Richardson prefers 'prostitute-john' to 'sex worker-client' because it is more negatively value-laden. She disapproves of some attempts to create a positive impression of sex work.
- ^{viii} The actual title of Levy's book is *Love and Sex with Robots*.
- ^{ix} The campaign's website states that a 'A comprehensive, pre-emptive prohibition on fully autonomous weapons is urgently needed'. See <http://www.stopkillerrobots.org/learn/> (accessed 21/1/2016)
- ^x Text taken from the CASR's 'About' page: <http://campaignagainstsexrobots.org/about/> (accessed 21/1/16)
- ^{xi} Although, in saying this, we would need to know whether Richardson thinks that this is possible. Some of her comments suggest that it wouldn't be because robots could never meet the criteria for personhood. See Richardson (2015).
- ^{xii} The qualification 'seems to make' needs to be added because Richardson is never that clear about her argument.
- ^{xiii} Obviously, there are some sex workers providing services to women but they are ignored in Richardson's argument. This is not uncommon in the literature about the ethics of sex work.
- ^{xiv} The personality types are described on the 'FAQ' page on the TrueCompanion website: <http://www.truecompanion.com/shop/faq> (accessed 4/5/2016)
- ^{xv} See Malamuth and Chech (1981); Zillman and Bryant (1982 & 1989); and Allen et al (1996).
- ^{xvi} Ferguson and Hartley (2009); and Diamond (2009).
- ^{xvii} Barak and Fisher (1997); and Barak, Fisher, Belfry and Lashambe (1999); Malamuth, Addison and Koss, M. (2000); and Hald, Malamuth and Lange (2013).
- ^{xviii} The rest of this paragraph is adapted from Earp, Sandberg and Savulescu (2014), in which the quote following this paragraph also appears.
- ^{xix} A famous example is the Canadian legal case of *R v Butler* in which the Canadian Supreme Court proposed a new test for banning materials on the basis of obscenity. In the aftermath of this decision, mainstream pornography continued to flourish while pornography catering to minority sexualities was suppressed. For an extended analysis see Crossman, Bell, Gotell and Ross (1997).
- ^{xx} Examples could include autonomous cars which have to make life and death decisions about whether to swerve to avoid an impact with a pedestrian or not; or the darknet shopper, which is an automated shopping robot that randomly selects items from the deep web for purchase. Given the nature of the deep web these can include illegal drugs and firearms. For a description see: <http://www.independent.co.uk/life->

style/gadgets-and-tech/news/random-darknet-shopper-exhibition-featuring-automated-dark-web-purchases-opens-in-london-a6770316.html (accessed 17/2/16)

^{xxi} This was signed by a number leading scientists, ethicists and robotics researchers.

See <http://futureoflife.org/open-letter-autonomous-weapons/> (accessed 17/2/16)

^{xxii} The argument appeals directly to the requirements of just war theory. There is some dispute as to *jus in bello* (i.e., the justice of actions in a war) and *jus ad bellum* (i.e., the justice of the decision to go to war) requirements and whether they offer a way round the responsibility gap argument. Purves, Jenkis and Strawser (2015) argue that this distinction should make no difference and, for the purposes of this discussion, we concur.

^{xxiii} There may be ways to create an AI that avoid codifiability. Machine learning techniques show some promise in this respect. They can be given examples of moral and immoral behavior and then extrapolate to new cases. They can sometimes learn subtle exceptions, but the problem with machine learning is that it tends to be opaque. We will not know whether an AI has figured out an important exception until we place it in that situation: we cannot just assume that it has from proper behavior in the training cases, and seeing proper behavior in other test cases merely gives us inductive reasons to think it will behave appropriately in real-world situations. If the exception is of a different *type* than the other test cases, then we have no real reason to trust the system.

^{xxiv} It is worth noting that in these papers, Muller and Simpson develop a position that is similar to the one advocated in the present chapter, namely: there are bad-making properties of AWSs that could warrant regulation and intervention, but not a preemptive campaign against their creation.

^{xxvi} A possible argument might be that if a person knew that nobody would be harmed by having sex of some (inherently) problematic kind, they might do it more often. Sex that is problematic because of harm to partners would now be unproblematic, but if there is an inherent harm (as per the conservative argument) or indirect bad effects (closer to Richardson's view), then there is now more harm overall due to the actions of these people. However, this would be overshadowed by the different actions of other people: moral people who abstain from problematic sex because it is morally problematic are likely a far smaller group in most cases than people who abstain from it simply because it is hard to find partners.

Bibliography

- Allen, M., Emmers, T., Gebhardt, L. and Giery, M (1996). Exposure to pornography and acceptance of rape myths. *Journal of Communication* 45: 5–26.
- Barak, A. and Fisher, W.A. (1997). Effects of interactive computer erotica on men's attitudes and behavior toward women: an experimental study. *Computers in Human Behaviour* 13: 353–369
- Barak, A., Fisher, W.A., Belfry, S. and Lashambe, D. (1999) Sex, guys and cyberspace: effects of internet pornography and individual differences on men's attitude toward women. *Journal of Psychology and Human Sexuality* 11: 63–91
- Baron-Cohen, S. (2011). *Zero degrees of empathy: a new theory of human cruelty and kindness*. London: Penguin.
- Bernstein, E. (2001). The meaning of the purchase: desire, demand and the commerce of sex. *Ethnography* 2(3): 389-420.
- Bernstein, E. (2007). Sex work for the middle classes. *Sexualities*, 10(4): 473 – 488.
- Boran, M. (2015, November 4). Sex robots should be banned, expert tells web summit. *The Irish Times*. Retrieved from: <http://www.irishtimes.com/business/technology/sex-robots-should-be-banned-expert-tells-web-summit-1.2416948> (accessed 21/1/2016)
- Bostrom, N. and Roach, R. (2011). Smart policy: cognitive enhancement and the public interest. In , Savulecu, J., ter Meulen, R., and Kahane, G. (eds) *Enhancing human capacities*, 138–152. Oxford, UK: John Wiley & Sons.
- Coady, C. A. J. (2009). Playing god. In Savulescu, J., and Bostrom, N. (eds) *Human enhancement*, 155–180. Oxford, UK: Oxford University Press.
- Crossman, B., Bell, S., Gotell, L. and Ross, B. (1997) *Bad attitudes on trial: pornography, feminism and the Butler decision*. Toronto: University of Toronto Press.
- Danaher, J. (2015). Robotic rape and robotic child sexual abuse: should they be criminalised? *Criminal Law and Philosophy*, in press. doi: 10.1007/s11572-014-9362-x
- Dancy, J. (2013). Moral particularism. In Zalta, E. (ed) *The Stanford Encyclopedia of Philosophy*. Retrieved from: <http://plato.stanford.edu/entries/moral-particularism/> (accessed 18/1/15).
- Diamond, M. (2009). Pornography, public acceptance and sex related crime: a review. *International Journal of Law and Psychiatry* 32: 304–314.
- Earp, B. D., Sandberg, A. and Savulescu, J. (2014). Brave new love: the threat of high-tech “conversion” therapy and the bio-oppression of sexual minorities. *AJOB: Neuroscience* 5(1): 4-12.

Earp, B. D., Sandberg, A. and Savulescu, J. (in press). The medicalization of love: response to critics. *Cambridge Quarterly of Healthcare Ethics: Clinical Neuroethics*, in press.

Earp, B. D. and Moen, O. M. (2015). Paying for sex—only for people with disabilities? *Journal of Medical Ethics* 42(10): 54-56. doi:10.1136/medethics-2015-103064

Farley, M., Bindel, J. and Goulding, J. (2009). *Men who buy sex: who they buy and what they know*. London: Eaves

Ferguson, C. J. and Hartley, R. D. (2009). The pleasure is momentary ... the expense damnable? The influence of pornography on rape and sexual assault. *Aggression and Violent Behavior* 14: 323–329.

Guarini M. and Bello, P. (2012). Robotic warfare: some challenges in moving from noncivilian to civilian theaters. In Lin, P., Abney, K. and Bekey, G. A. (eds) *Robot ethics: the ethical and social implications of robotics*. Cambridge, MA: MIT Press.

Gupta, K. (2012). Protecting sexual diversity: rethinking the use of neurotechnological interventions to alter sexuality. *AJOB Neuroscience*, 3(3), 24-28.

Hald, G. M., Malamuth, N. M. and Lange, T. (2013). Pornography and sexist attitudes among heterosexuals. *Journal of Communication* 63: 638–660

Levy, D. (2007). *Love and sex with robots: the evolution of human-robot relationships*. New York: HarperCollins.

Malamuth, N. M. and Chech, J. V. P (1981). The effects of mass media exposure on acceptance of violence against women: a field experiment. *Journal of Research in Personality* 15: 436–446

Malamuth, N. M., Addison, T. and Koss, M. (2000). Pornography and sexual aggression: are there reliable effects and can we understand them? *Annual Review of Sex Research* 11: 26–91.

Marino, P. (2008). The ethics of sexual objectification: autonomy and consent. *Inquiry*, 51(4): 345-364.

Moen, O. M. (2014). Is prostitution harmful? *Journal of Medical Ethics* 40(2): 73-81.

Muller, V. (2016). Autonomous killer robots are probably good news. In Di Nucci and Santoni De Sio (eds) *Drones and responsibility: legal, philosophical and socio-technical perspectives on the use of remotely controlled weapons*. London: Ashgate.

Murphy, C. (2015, August 14). Sex workers' rights are human rights. *Amnesty International*. Retrieved from <https://www.amnesty.org/en/latest/news/2015/08/sex-workers-rights-are-human-rights/>.

Purves, D., Jenkins, R. and Strawser, B. (2015). Autonomous machines, moral judgment, and acting for the right reasons. *Ethical Theory and Moral Practice* 18(4): 851-872.

Richardson, K. (2015). The 'asymmetrical' relationship: parallels between prostitution and the development of sex robots. *SIGCAS Computers & Society* 45(3): 290-293. Retrieved from: <http://campaignagainstsexrobots.org/the-asymmetrical-relationship-parallels-between-prostitution-and-the-development-of-sex-robots/> (accessed 21/1/2016)

Sandberg, A. (2015, September 16). Sex and death among the robots: when should we campaign to ban robots? *Practical Ethics Blog*. Retrieved from: <http://blog.practicaethics.ox.ac.uk/2015/09/sex-and-death-among-the-robots-when-should-we-campaign-to-ban-robots/> (accessed 21/1/2016)

Sanders, T., O'Neill, M. and Pitcher, J. (2009). *Prostitution: sex work, policy and politics*. London: Sage.

Sanders, T. (2008). Male sexual scripts: intimacy, sexuality and pleasure in the purchase of commercial sex. *Sociology* 42(1): 400-417.

Semrau, L. (2015). The best argument against kidney sales fails. *Journal of Medical Ethics*, 41(6), 443-446.

Simpson, T. and Muller, V. (2016). Just war and robots' killings. *The Philosophical Quarterly* 66(263): 302-322.

Shackel, N. (2005). The vacuity of postmodernist methodology. *Metaphilosophy* 36(3): 295-320.

Sparrow, R. (2007). Killer robots. *Journal of Applied Philosophy* 24(1): 62-77.

Strawser, B. (2010). Moral predators: the duty to employ unmanned aerial vehicles. *Journal of Military Ethics* 9(4): 342-368

Walton, D. (2013). Analogical arguments. In LaFollette, H. (ed) *The International Encyclopedia of Ethics*. Hoboken: Wiley-Blackwell.

Zillman, D. and Bryant, J. (1982). Pornography, sexual callousness and the trivialization of rape. *Journal of Communication* 32: 10-21

Zillman, D. and Bryant, J. (1989). Pornography and men's sexual callousness toward women. In Zillmann, D. and Bryant, J.(eds) *Pornography: research advances and policy considerations*. Hillsdale, NJ: Erlbaum.