Gamification, Side Effects, and Praise and Blame for Outcomes

Sven Nyholm[1]

Abstract

"Gamification" refers to adding game-like elements to non-game activities so as to encourage participation. Gamification is used in various contexts: apps on phones motivating people to exercise, employers trying to encourage their employees to work harder, social media companies trying to stimulate user engagement, and so on and so forth. Here, I focus on gamification with this property: the game-designer (a company or other organization) creates a "game" in order to encourage the players (the users) to bring about certain outcomes as a side effect of playing the game. The side effect might be good for the user (e.g., improving her health) and/or good for the company or organization behind the game (e.g., advertising their products, increasing their profits, etc.). The "players" of the game may or may not be aware of creating these side effects; and they may or may not approve of/endorse the creation of those side effects. The organizations behind the games, in contrast, are typically directly aiming to create games that have the side effects in question. These aspects of gamification are puzzling and interesting from the point of view of philosophical analyses of agency and responsibility for outcomes. In this paper, I relate these just-mentioned aspects of gamification to some philosophical discussions of responsibility gaps, the ethics of side effects (including the Knobe effect and the doctrine of double effect), and ideas about the relations among different parties' agency.

Keywords Gamificiation · Praise and blame · Side effects · The Knobe effect · The doctrine of double effect · Responsibility gaps

✉ Sven Nyholm
    s.nyholm@lmu.de

[1]   LMU Munich, Munich, Germany

1 Introduction

When I go running, I use a running app that tracks data about and gamifies my running. If I run more than usual during a week or a month, for example, a badge is displayed, resembling a prize for having won a competition. My wife and father-in-law also use the same app, and our running is compared on a leader board. This turns our running into not just a form of exercise, but also a competition, since we are ranked in terms of who runs the most. My wife, father-in-law, and I enjoy the friendly competition. This motivates us to run more and longer. As a "side effect", we improve our condition and thereby become healthier. Another side effect is that it is hard for us not to associate running with the shoe brand behind the running app. So, whenever we need to buy new running gear, the brand in question is the supplier that first comes to mind. Presumably, this is the main aim of the running app from the point of view of the company behind it. A further aim of theirs might be to collect data about people's exercise patterns, which can be useful for them in different ways. Thus, different agents are interacting, directly or indirectly, via this running app, and various different goals are achieved as side effects of the use of the app.

Gamification of activities can also motivate people even if the people in question do not actually welcome – or do perhaps not voluntarily choose – the gamification of what they do. In *Forbes* magazine, for example, there was a story a few years back about how Disneyland was ranking the performance of the staff, displaying the ranking of staff members' performance on a leaderboard on a screen in the staff room (Allen, 2011). This motivated the staff to work harder, but they didn't like it. Some of them called this an "electronic whip".

In a similar vein, many academic researchers do not like that data about their research publications is being tracked and gamified, so that getting the most reads, citations, etc. is turned into a competition and works as a form of motivation for people to become more productive (Nyholm, 2023: 110). Even if this motivates some of the people whose research is turned into a competition with rankings and so on, those same researchers don't always welcome this. Some feel like they're being manipulated or given the wrong kinds of incentives to do their research and publish their papers.

People sometimes don't even fully realize that significant forms of gamification are used to incentivize them to behave in certain ways. Social media websites, for example, that display the number of "likes" and other forms of reactions to people's social media posts are turning posting content onto social media websites into a game, where people compete against each other in terms of who gets the most likes or other reactions (Nguyen, 2020a). While many people feel almost addicted to the social media they use, many people also complain about the stress involved in having one's performance on social media be displayed for all to see. The point of having people constantly return to these social media sites is, from a business standpoint, to have them on the websites for long periods of time, so that the companies running the sites can sell advertisement space. So, a side effect of people's constantly going on these social media sites and competing for others' attention and positive responses is that the companies operating the sites can sell more advertisement space (Véliz, 2020).

In short, in many parts of life – and many more than those just mentioned above – gamification is increasingly being used to influence people's behavior and to motivate them to act in certain ways. Most apps on people's smartphones use gamification, and many websites – such as social media websites – do it as well. Attempts to influence and manipulate people's behavior and motivation using technological means is sometimes called "persuasive technology", and there exists a small, but growing philosophical literature about the ethics of persuasive technology, including the ethics of gamification (e.g., Fogg, 2003; Spahn, 2012; Kim & Werbach, 2016; Marczewski, 2017; Smids 2018; Lanzing, 2019). This paper contributes to that literature, by bringing up some issues that – as far as I know – have not been widely discussed in this literature.[1]

My topic is the philosophical issue of (a) how to analyze the agency of the different parties involved in gamification, given that it typically involves the production of certain outcomes as a form of side effect of behaviors that are motivated by the gamification of the activities in question, with (b) a focus on who can be held responsible for these outcomes created with the help of gamification. I will approach these issues from the point of view of the ethics of side effects for reasons that I will unpack throughout my discussion. And I will be concerned both with blame for what might be considered as bad outcomes and with praise for good outcomes.

Why this topic? Here are two motivations. First, as noted above, what I discuss below hasn't yet been discussed in the philosophical literature about the ethics of gamification. One argument in a short paper by Tae Wan Kim can – as I will argue later in the paper – be related to the main line of argument in this paper. But otherwise, the topics in this paper have not been covered in the literature on the ethics of gamification. So, there is a gap to fill here.[2] Second, some authors have argued that gamification can change the character of certain activities, so that the value of the activities and their products change. For example, Danaher et al. (2018) argue that gamifying behaviors associated with romantic relationships might change the value of – or the ways in which people value – such behaviors.[3] I take inspiration from that general line of argument here, and ask whether gamification might also be seen as changing the character of different people's decisions and behavior with respect to

---

[1] Regarding contributions to the philosophy of games literature that could potentially be related to my discussion here, Nguyen (2019: 425; 2020b) interestingly discusses how games "highlight our ability to substantially, voluntarily, and quickly manipulate aspects of our agency", and also argues that certain forms of games allow agency to become a kind of art. While this discussion of games and agency is related to my discussion of how to analyze the agency of participants of gamified activates on a certain level, since we are both interested in agency and gamification, my entry into this discussion is quite different from Nguyen's since my approach treats the ethics of gamification as part of the ethics of side effects and responsibility attribution.

[2] The literature on the ethics of gamification has covered, among other things, the following issues: whether gamification is exploitative, manipulative, intentionally or unintentionally harmful to the parties involved, has a bad effect on people's character, changes the way people value things in potentially problematic ways, is dominating, makes our lives more or less meaningful, or threatens or potentially sometimes boosts personal autonomy (Kim & Werbach, 2016; Marczewski, 2017; Danaher et al., 2018; Danaher 2019; Arora & Razavian, 2021; Gorin, 2021; Parmer, 2021; Nyholm, 2023: chapter five.

[3] In particular, Danaher et al. (2018) argue that the gamification of dating and relationships might lead people to value quantifiable aspects of dating and relationships more than qualitative aspects of these things, even though the latter might seem more important from a philosophical point of view.

what they can be praised or blamed for. As we will see, gamifying people's activities can potentially also be seen as influencing what exactly they can and should be interpreted as intentionally doing – at least from the point of view of what they might deserve praise or blame for. I will discuss both those who introduce game-like elements into activities and those who engage in those gamified activities.

The rest of the paper divides into the following sections. I'll start by briefly explaining the general idea of gamification and the specific form of gamification I am interested in. The type of gamification I am interested in is, in short, a method of motivating people to produce certain outcomes as side effects of what they are doing (Sect. 2). I will then introduce some general ideas from the philosophy of responsibility – especially as it relates to worries about potential gaps in or a weakening of responsibility in relation to modern technologies. One crucial idea is that people are most strongly responsible for what they themselves intentionally do or decide to do – an idea that I will here call "Williams' point" (Sect. 3). I will then introduce two familiar ideas from the philosophy of side effects – namely, the "Knobe effect" and the "doctrine of double effect" – that I think bear interestingly on how to think about the issue of what people are responsible for in relation to gamified activities in the sense of what they can plausibly be praised or blamed for (Sect. 4). The following section combines what I consider to be suitably interpreted versions of Williams' point, the Knobe effect, and the double effect principle. I will bring these three ideas to bear on the issue of what those whom I will call "game designers" and "game players" can be praised and blamed for in relation to gamified activities (Sect. 5). The final section is a brief concluding discussion (Sect. 6).

2 What is Gamification (of the Sort that this Paper Focuses on)?

In general, the expression "gamification" refers to the introduction of game-like elements into activities or practices that aren't necessarily games in themselves (Werbach, 2014). This is usually done in order to encourage or motivate people to engage in the activities in question. None of this must necessarily be done in a technological context. You could make something into a form of game without there being any need to involve any particular modern type of technology. However, adding game-like features to activities is often done in many forms of contemporary technologies. The game-like elements can be to let people get points for doing certain things, giving them different forms of awards or recognition for doing certain things, or letting them compete with others by ranking them relative to others (Danaher et al., 2018). This is often done in apps on smart phones or on websites, such as social media websites of different sorts.

The form of gamification I will focus on here can be explained in the following general terms:

The game designer (e.g., a company or organization) creates a "game" (i.e., they add game-like elements to some activity) so as to encourage certain behaviors on the part of the players (the participants, e.g., people using an app or

visiting a website) in order to bring about certain outcomes as a side effect of
the players' playing the game.

In other words, the game designer creates a game-like activity that motivates players
to act in certain ways with the help of points, rankings, or other game-like forms of
incentives. The goal or hope is that by behaving in the ways that makes them perform
well in the game, the players will bring about certain outcomes as side effects of these
behaviors. Those side effects are typically what the game designers are ultimately
aiming to bring about.

Many of the examples I mentioned in the introduction fit with this idea. In the run-
ning app example, for instance, the game that is designed is a competition whereby
people get points, reach certain levels, or are ranked according to who runs the most.
This encourages people to run more. A side effect for the runners is that they improve
their condition and become healthier. A side effect for the company behind the run-
ning app is that users of the app start associating running with this running app and
the company behind it, which might motivate them to buy the goods the company is
selling. Also, data about people's running is collected as a further side effect, and this
data might be used for all kinds of purposes by the company behind the app (Arora
& Razavian, 2021; Nyholm, 2023: chapter five). In other words, the introduction of
game-like elements to certain practices might produce various different outcomes
(benefiting different people) as designed-for side effects of the gamification of the
activities in question.

It is important to note that while gamification will often be designed to work over
a period of time as part of a continuous activity (e.g., in the use of an app), it could
also be a one-off matter. For example, Tae Wan Kim discusses the so-called ALS ice
bucket challenge from the mid 2010s as a form of gamification in a way that I think
illustrates the point that gamification can be a one-off matter. "ALS" is short for
amyotrophic lateral sclerosis, which is a "progressive neurodegenerative disease that
affects nerve cells in the brain and spinal chord."[4] This is a disease that many people
used to not be aware of. In the mid 2010s, however, there appeared lots of videos
online with people pouring ice water over their heads in order to "bring attention to
ALS", as people would put it. Many people became motivated to join this movement
and do the ice bucket challenge, and thereby supposedly helped to bring attention
to ALS. As Kim points out, however, some – or even many – of the people who did
this ice bucket challenge were very likely motivated primarily by the opportunity to
get likes and other forms of attention on social media. Accordingly, bringing ALS to
people's attention was a sort of side effect of the game of getting likes and attention
by pouring ice water over one's head and posting videos of this online. This was not
an on-going activity for the people who participated in this game, but a one-time
social media post, where many people were presumably motivated to join in because
they were tempted by the opportunity of getting attention and approval from others.
In sum, gamification can be used either in on-going activities or be a one-off trick to
motivate people to act in a certain way.

---

[4] https://www.als.org/understanding-als/what-is-als (accessed on 20 July 2023).

Now, another interesting aspect of the kind of gamification I am discussing is that some of the designed-for side effects may potentially be things that the people participating in the gamified activities may not be particularly intrinsically motivated to help to bring about. When users run and are motivated by the game-like elements of the running app, for example, they are presumably typically not directly aiming to sell more running shoes or generate data for the company. Indeed, some of the people participating in the just-mentioned ALS ice bucket challenge may not have been particularly intrinsically motivated to spread awareness of ALS (even if they may have had nothing against doing so). Some of them were more likely primarily motivated by the prospect of doing something that would get attention and likes on social media.

Similarly, when users on gamified forms of social media are trying to increase their numbers of followers or friends on social media, or when they are trying to post content that will get a lot of attention and lots of "likes" and so on, they are presumably very rarely directly motivated to bring about the designed-for side effect of this behavior. For example, they may rarely be motivated to bring about the side effects whereby they are making the social media platform into a more attractive advertisement space for companies who want to advertise their products to users of the platform. But that is the intention behind many of those features, from the social media companies' point of view. That is, the company behind the social media has added game-like features to their site as a way of motivating people to frequently return to the social media site, to make posts, and to stay long on the site once they are logged in (Nguyen, 2020a).

Of note here, then, is that people whose activities have been gamified may not be directly motivated to do certain things the game designers want them to do. They might even be averse to doing the things in question. They might, for example, be averse to the idea of helping tech companies to sell advertisement space. Yet, these same people may be strongly motivated by the game-like elements, such as doing well in comparison to their peers, getting different forms of recognition from their peers, and so on. And the game-like elements added to the technologies they use (such as a social media site) might motivate these people to work hard. This might then have as a side effect the thing – in this case, high levels of user engagement, which is good for business – that the game designers (as I call them) want them to produce.[5]

In short, what I am talking about in this paper is adding game-like elements to activities with the aim of motivating people to engage in the activity, so that certain outcomes are produced as side effects of the people playing the game. Importantly,

---

[5] The initiative to engage in gamified behaviors might of course also come from the side of the user. For example, you yourself might want to achieve a certain outcome – such as improving your condition and getting healthier. But you may know about yourself that you are not sufficiently intrinsically motivated to take the means to your end (e.g., exercising, taking your medicine, changing your eating habits, or whatever) unless the activity is made into a form of game (Spahn, 2012). However, you may know that turning the activity into a game – or using an app or other technology that turns the activity into a game – may help you to become more motivated to engage in the activity in question (e.g., exercise, etc.). As a result, improving your condition and health will then be a side effect of engaging in behavior driven by what more directly motivates you (namely, playing some game or engaging in some game-like activity).

the side effects may be good or bad, welcome or unwelcome, depending on your perspective. The question therefore arises of who is responsible for the production of these outcomes that are intended or unintended side effects of people's playing the game. In particular, how should we think about praise and blame in relation to the outcomes of gamified behaviors or activities?

3 Responsibility for Actions and Outcomes: General Considerations

As noted above, while gamification does not necessarily need to be related to modern technologies (such as apps or websites), it often is. So, when it comes to responsibility and gamification, it is relevant to relate this issue to recent discussions of responsibility within the ethics of technology, to see what we can learn from such discussions in this context. In this section, I will bring up some considerations related to responsibility within technology ethics as well as the philosophy of responsibility more generally. I am here particularly interested in responsibility in the sense of who, if anyone, can be praised or blamed when good or bad outcomes are brought about in different ways.

Whenever technologies – especially new and emerging technologies – are discussed from an ethical point of view, one of the first questions that always comes up is "who is responsible?" (Nyholm, 2023: chapter six). This is usually discussed in relation to bad outcomes related to technologies, such as people being hit and killed by self-driving cars or AI technologies causing various problems for people (Matthias, 2004; Sparrow, 2007; Hevelke & Nida-Rümelin, 2015). The question then is "who is to blame if and when something goes wrong?", and a common worry is that gaps in responsibility might open up. But we can also ask who, if anyone, deserves praise or credit when a good outcome is produced as a result of the use of a new form of technology, e.g., if some new form of medical AI might make a correct diagnosis of a patient's problems (Nyholm, 2023: chapter six). To give another example: who, if anyone, might be seen as responsible for good or bad outputs created by new forms of generative AI technologies (Porsdam Mann et al. 2023)?

These kinds of worries may concern indeterminacies or even gaps in responsibility. Importantly, even if there is not a complete gap in responsibility, there might still be reason to worry about what I am here calling a weakening of responsibility. People are potentially sometimes less responsible than they might be in other cases because of the involvement of certain technologies that make the people in question less directly responsible for the outcomes produced. Or if lots of people are doing something together as a big group, then this might make some or all of the individuals involved less strongly responsible for the outcome than they would be if they were acting alone or together with a smaller group (cf. Van de Poel et al. 2015; see also De Jong, 2020).

As I see things, then, we should not think of responsibility as a binary or an either/ or matter: as in, either somebody is responsible or not. We can also think in terms of degrees of responsibility. That is, if somebody is responsible for something, X, under normal circumstances, there could be cases in which something makes them less than fully responsible for X – as opposed to circumstances when they are not responsible

for X at all. When I speak about gaps in responsibility here, I will mean either complete gaps or significant forms of weakening of responsibility.

In normal cases in which people are most clearly responsible for something, this is usually because they are directly doing something, on their own initiative, and because they know what the possible outcomes (including possible side effects) of what their actions are (Coeckelbergh, 2020). Moreover, by common criteria, the more people are in control over what happens, the more things are the result of their plans, and the more things happen because of what they are directly intending to do or knowingly allowing to happen, the more they are considered as being fully responsible for the outcomes of their actions (Nihlen-Fahlquist, 2017; List, 2021).

Now, another thing of importance in the present context is that people are typically primarily seen as responsible for their own actions and their effects. Granted, people can also sometimes be seen as being (partly) responsible for what other people do. But this is then usually because they are in some role in which they have the authority to command others to act in certain ways, so that others act in certain ways because those in charge directly tell them to do so (Sparrow, 2007; Nyholm, 2020: chapter three). Moreover, sometimes people can also be seen as being jointly responsible together with other people for outcomes that they produce together with these other people, at least if they are acting together as a well-organized team, group, or organization (List & Pettit, 2011). Normally, however, people are seen as most clearly and most strongly responsible for what they themselves do on their own initiative or for the outcomes of their own actions.

In fact, in Bernard Williams' now-classic attempt to refute utilitarianism, one of his most well-known arguments against utilitarian ethics is precisely that it fails to respect the idea that we are primarily responsible for the effects of our own actions, whereas other agents are primarily responsible for the effects of their actions. Utilitarianism, Williams (1973) thinks, implausibly treats ethics as simply being about the creation of the best overall outcome, while regarding everyone as equally responsible for bringing about these optimal outcomes. This is a problem, Williams argues, because normally, we are primarily and most strongly responsible for the effects of our own actions and not the effects of other people's actions. In my discussion below, I shall take this idea onboard, and will later refer to it as "Williams' point".

Another issue related to responsibility in general that is relevant in the current context is that there are important asymmetries between the conditions for being worthy of praise (responsibility in a positive sense) and the conditions for being worthy of blame (responsibility in a negative sense). For example, whether we are primarily doing something *as a means to some other end* matters in different ways relative to whether we are praiseworthy or blameworthy for our actions. If we are doing something that is generally considered good (e.g., being nice to somebody or helping somebody), but we are primarily doing this as a means to the end of promoting our own self-interest, as opposed to doing it for its own sake, then this makes a big difference to whether we will be seen as praiseworthy for doing the good thing, by common criteria. In order to count as clearly being praiseworthy for doing something good (e.g., being nice to, or helping, somebody), it should be that we are not just doing this as a means to some other end (e.g., the end of promoting our own self-interest) or because it is convenient to us. It should be that we have a non-instru-

mental motivation or disposition to do the good thing (Pettit, 2015). In contrast, if we do something bad (e.g., if we're cruel to, or if we harm, somebody) as a means to promoting some other end (e.g., the end of promoting our own narrow self-interest), then the consideration that we did the bad thing as a means to some other end doesn't affect whether we can be seen as clearly being blameworthy for doing the bad thing. Instead, the very fact that we are willing and prepared to do something bad as a means of achieving our own personal ends can be exactly the thing that makes us blameworthy for our action.

In short, whether we're praiseworthy for doing something that can be considered good (e.g., helping somebody) might importantly depend on whether we're doing it as a means to some other end or for its own sake. In contrast, whether we're blameworthy for doing something that's considered bad (e.g., harming somebody) doesn't in the same way depend to the same degree on whether we are doing this as a means to some other end or for its own sake. Having made that general observation about how there can be asymmetries between the criteria for praiseworthiness and the criteria for blameworthiness, let us now return to the idea of cases in which (via gamification or for other reasons) people produce good or bad side effects via their actions. I will now bring up two ideas from the ethics of side effects – one more recent idea, and one older idea – that, as I will later argue, bear in interesting ways on how to think about whether people might be seen as praiseworthy or blameworthy for the side effects produced in gamified activities. One these two ideas also has to do with asymmetries between praise and blame.

4 The Knobe Effect and the Double Effect Principle: Two Familiar Ideas from the Philosophy of Side Effects

To recap, one important question related to gamification is whether gaps in, or at least a weakening of, responsibility might come about as a result of the gamification of many activities. If not, how do we and how should we determine who is responsible for the outcomes produced as side effects when people engage in gamified activities? These concerns about possible gaps in, or a weakening of, responsibility in relation to gamification might arise both with respect to potential blame for negative outcomes and with respect to praise for positive outcomes produced as side effects of people engaging in gamified activities. Yet, these worries might perhaps be particularly pressing when the outcomes of gamified activities are controversial or seen as negative from certain points of view.

Importantly, the people "playing" the games might deny that they are responsible for some of the relevant outcomes (e.g., helping tech companies to sell advertising space by spending a lot of time on social media sites, or more generally extending tech companies' influence over people's lives). At the same time, the companies behind the gamified technologies might deny that they are responsible for certain outcomes because other people (viz. the players) are the ones who are actively playing the games in question and thereby producing the outcomes.

In general, the worry here is that active agency is outsourced from the game designer to somebody else (the game players). The game players may be producing

certain outcomes as side effects of what they are doing. Yet the game players may not be very interested in producing these side effects and may even regret doing so, whereas the game designers may want them to produce these side effects, but may potentially deny responsibility, since the players are the active agents who are playing the game and producing the effects.[6]

What is perhaps most puzzling, in other words, when it comes to how to think about who is responsible for outcomes produced with the help of gamification is that the designed-for outcomes of the gamification are often side effects of people doing certain things, as opposed to outcomes that the active agents are directly striving to bring about. And the people who want to bring about the outcomes in question are often different people than the active agents.

So, when we think about responsibility for outcomes and gamification, it becomes important to reflect on the ethics of side effects. The ethics of side effects is something that many philosophers have discussed in interesting ways in other sub-areas of moral philosophy, though not much (as far as I know) in relation to the ethics of gamification. I will therefore now bring up some interesting older discussions about the ethics of side effects, focusing on two examples in particular, and later highlight their relevance to new questions related to side effects produced by gamification.

4.1 The Knobe Effect and a Normative Reading of it

Let us start with the so-called Knobe effect. Many readers may be familiar with this idea from the field of experimental philosophy. On an abstract level, the apparent effect is that people tend to see bad side effects that others knowingly produce as being intentionally produced by them but good side effects that others knowingly produce as not being intentionally produced by them. Knobe (2003, 2010), after whom the effect is named, likes to illustrate this finding (also called the "side effect effect") with the help of a pair of vignettes that a great number of people have been presented with, which reveal interesting patterns in people's responses to these and similar vignettes. Let us consider the two most well-known vignettes.

In the first variation, the CEO of a company is told by his advisor that a new policy would make the company a lot of money, but that it would also harm the environment. The CEO says, "I don't care about the environment, I just want to make a lot of money!", and the policy is implemented. Sure enough, the company both makes a lot of money and harms the environment. Presented with this first story, most people say that the CEO intentionally harmed the environment.

If the story is changed by substituting the word "helping" for "harming" and the story is presented to people, then usually the people respond in a different way. So, now the CEO – according to this second version of the story – is told by an advisor that a new policy would help the environment and make a lot of money. The CEO says, "I don't care about the environment, I just want to make a lot of money!", and

---

[6] One might here also worry about what Rubel et al. (2019) call "agency laundering": i.e., tech companies trying to make others or technologies do things they want, so that they can reap certain benefits while claiming that somebody else (or something else, viz. a technology or other people) did the thing that led to whatever outcome others might be trying to hold them responsible for.

again the policy is implemented. Sure enough, a lot of money is made by the company and the company helps the environment. Here, people tend to respond in the following way: when asked whether the CEO intentionally helped the environment, they tend to say "no".

Importantly, there is no controversy about the fact that people tend to respond to these cases according to this pattern. The controversy is instead about how to interpret these findings. I will first briefly mention Knobe's initial way of interpreting his own findings, and then formulate a reading that I find more plausible and – additionally – more relevant in the present context. But let us first consider Knobe's own initial way of understanding these findings.

Knobe suggested that the very concept of *intentionally doing something* is such that people's application of this concept is not just sensitive to their beliefs about other people's psychological states and the relation between them and their behaviors – rather, people's application of the concept of *intentionally doing something* is also directly sensitive to judgments about the moral status of other people's actions and their effects. On this way of understanding things, people make judgments about whether other people's actions and their effects are good or bad, and then after this – or partly influenced by this – they make a judgment about whether those other people produced certain effects intentionally or not. In other words, the very concept of intentional action has some sort of moral component baked into it. It is not a purely psychological concept. Or so Knobe (2003) has argued.

Another possible reading of what is going on – the one that I will favor here – is that the normative criteria for qualifying as performing a praiseworthy act of doing good differ in kind from the normative criteria for qualifying as performing a blameworthy bad act (O'Brien, 2015; Pettit, 2015). To qualify as performing a praiseworthy act of helping, for example, it might need to be the case that we are actively trying to produce a certain good effect, as our main aim in acting in some way, and that we exert some effort in this direction, and that we show talent, or that we make some sacrifice in order to produce the good outcome (Maslen et al. 2020). Only then do we qualify as intentionally producing the good outcome in a praiseworthy way. In contrast, to qualify as producing a bad outcome in a blameworthy way, our commonsense criteria only stipulate that we allow ourselves to produce the bad outcome even though it may have been avoidable, perhaps doing so as a side effect of pursuing some other aim of ours, such as promoting our own self-interest.

In what follows, I shall put things in the following way, and I will call this "the normative reading of the Knobe effect": to qualify as intentionally doing good in a clearly praiseworthy way, it typically cannot be that the good that we do is a mere foreseen side effect of what we are doing, whereas to qualify as intentionally doing something bad in a blameworthy way, it can often be sufficient that we bring about something bad as a foreseen side effect of what we are doing.

This normative way of reading what is going on fits with the findings from Knobe's extensive studies of people's intuitions about the kinds of vignettes he has been presenting ordinary people with. But it does not make any assumptions to the effect that people's concept of intentional action is somehow in itself a normative concept. Instead, this way of reading what's going on simply assumes that, as noted in the

previous section, the normative criteria for being praiseworthy are in certain ways different from the normative criteria for being blameworthy for our actions.

As we will see below, this has interesting implications for how to think about the side effects (good or bad) of gamified activities. But before we get to those implications, let us first introduce another well-known idea from the ethics of side effects.

4.2 The Doctrine of Double Effect

Let us now contrast what was just discussed with another well-known idea from the philosophy of side effects, namely, the so-called doctrine of double effect, which provides an interesting contrast to the Knobe effect. The doctrine of double effect is partly associated with certain arguments within military ethics/just war theory or the ethics of abortion and other parts of traditional bioethics, but it is also often presented as a core part of common-sense morality (Persson, 2013). Many philosophers have criticized this doctrine, but many of them have also at the same time recognized it as capturing a widely shared moral intuition within ordinary common-sense thinking (e.g., Persson & Savulescu, 2012; Scanlon, 2008). For example, in his book about the trolley problem[7], David Edmonds (2013) argues that the doctrine of double effect is likely to be a key explanation of why many people have the intuitions they have about the most well-known trolley problem cases.

What is the doctrine of double effect? According to one way of stating this doctrine, it contains the following two ideas. First, if somebody directly produces some harm or evil, then this can justifiably be judged very harshly from an ethical point of view. Second, however, if somebody produces some harm or evil as a mere foreseen side effect of what they are doing, then this production of the bad outcome can justifiably be judged less harshly from a moral point of view.

For example, if a person kills somebody as a hate crime or as a direct means of achieving some self-interested aim of theirs, this can be judged very harshly from a moral point of view. In contrast, if somebody is fighting back against an attacker as an act of self-defense, and they accidentally kill their attacker as a foreseen possible side effect of defending themselves but not as a directly intended goal, then this can be judged much less harshly. Or to use the standard trolley cases that Edmonds (2013) discusses: if we redirect a trolley that is about to hit and kill five people to a side track, with the foreseen but not directly intended side effect of killing one person on the sidetrack, then this is judged less harshly than if we directly push a person in front of the trolley as an intended means of stopping the train from killing the five (see also Kamm, 2015).

Sometimes when philosophers introduce the doctrine of double effect, they instead say that doing something bad can be permissible if it is a foreseen side effect of doing something good or of doing something morally neutral. It is important to note, I think, that that way of stating the idea is compatible with one's being open to some

---

[7] As readers will know, the trolley problem is a philosophical puzzle regarding people's conflicting intuitions about different cases related to whether it is permissible to kill a smaller number of people in order to save a greater number, often illustrated with the help of examples involving a runaway trolley about to crash into people. See Edmonds, 2013 and Kamm, 2015.

degree – perhaps even a significant degree – of criticism for knowingly bringing about the bad side effect in question. The idea is just that on balance, it might be okay to bring about the side effect – even if one is to some extent open to criticism for doing so – if the main effect of one's action is good enough to outweigh the fact that one is simultaneously bringing about something bad as a foreseen side effect. Accordingly, the way I am summarizing the double effect principle here – i.e., that directly intending something makes one open to stronger moral criticism than allowing something as a side effect of one's action does – is quite compatible with other ways of spelling out the doctrine of double effect.

With these ideas – viz. the Knobe effect and the double effect principle – having been introduced into the discussion, let us now return to gamification and praise and blame for side effects. We can relate the points raised in this section back to what I called "Williams' point" above, and ask what this all implies for praise and blame for the designed-for side effects of gamified activities.

5 Implications for Praise and Blame for Side Effects Produced within Gamified Activities

To remind the reader, the type of gamification in focus in this paper works in the following way: a game designer (e.g., a company or other organization) introduces game-like elements into activities in order to encourage and incentivize game players (e.g., users of apps or webpages) to behave in certain ways that will have side effects that the game designer aims, by design, to bring about. Those side effects might be judged to be either good or bad, depending on one's perspective. And so the question is which of the involved parties can be seen as being responsible (as in, who can rightly be praised or blamed) for those outcomes of the gamified activities. The analysis of this is – I am taking it – constrained and complicated by, among other things, the following three theses:

1: *Williams' point*: we are primarily and most strongly responsible for the effects of our own actions, rather than for the effects of other people's actions.

2: *The normative version of the Knobe effect*: to qualify as intentionally doing good in a clearly praiseworthy way when we produce certain effects, it must be that these are not merely foreseen side effects of our actions. It should be that these effects are directly aimed-for effects of our actions, which are what directly engages our efforts and drives us in our actions. In contrast, to qualify as intentionally doing bad in a blameworthy way when we produce certain effects, it can be enough that we allow these effects to come about as side effects of our actions. So, we can be responsible for foreseen negative side effects of our actions.

3: *The double effect principle*: however, when any negative effects of our actions are side effects of these actions – as opposed to directly aimed-for aspects or outcomes of our actions – we can be judged less harshly for bringing about these effects than if we are directly aiming to bring about these effects (either for their own sakes or as a directly intended means to some other end).

We can now try to bring everything together and reflect on what this means with respect to, on the one hand, game players' responsibility for the designed-for side effects of their gamified activities and, on the other hand, game designers' responsibility for the designed-for side effects of the gamified behavior of the activities that the game players engage in. Let us start with the former.

In general, we can see that players in gamified activities cannot deny responsibility for the effects of their activities by saying that it is other people who are bringing about these effects through their actions. After all, they are the active agents behind these effects. (This relates to Williams' point.) However, we can also see that players can potentially be seen as less clearly responsible for any good side effects that might be generated by their gamified behaviors, as opposed to any bad side effects that might be generated and that they are aware of producing. (This relates to the normative reading of the Knobe effect.)

In other words, it might be argued that if good things are generated as side effects of our engaging in gamified activities, then we can take less credit for those good effects than if we were directly aiming to bring them about. But if we realize that a consequence of our engaging in the gamified activity is the creation of a side effect that the game designer wants to bring about by creating this game, and this effect is bad (bad, that is, as seen from some critical point of view, which the game designer might wish to disagree with), we are not off the hook. We can then be blameworthy in a clear way for bringing about these side effects through our actions. (These points follow from the normative reading of the Knobe effect.) However, at the same time we would be less blameworthy than if we were ourselves directly trying to bring about the bad effects in question. (This last point follows from the double effect principle.)

The game designer might seem to be able to deny responsibility for the production of the outcomes that are designed-for side effects of the players' playing the game and that critics may see as bad effects. (This relates to Williams' point.) The game designer does indeed ultimately aim to bring about these effects – so they are not mere foreseen side effects. (Hence the double effect principle might seem to be relevant here.) However, the trick here is precisely that the game designers have planned things so that somebody else (viz., the game players) are the ones who are the active agents who bring about the effects as a result of their actions. (So, the Williams point might seem to give the game designer a way of denying responsibility.)

If we stop there, we might come to a rather frustrating conclusion: namely, that the players of the game are more clearly responsible than the game designers for any bad outcomes that might be the designed-for side effects of their gamified activities, even though those players might be averse to the side effects, and even though the game designers might be directly aiming to stimulate the players to act in ways that bring about these side effects. This might seem unfair, and it might seem that it would be better to be able to hold the game designers responsible.

One way of going here is to say that we should abandon one or more of Williams' point, the normative Knobe effect, or the double effect principle when assessing the level of responsibility for the actions and decisions of game designers who gamify certain activities with the aim of bringing about certain effects that might be considered bad/problematic. That would be one option, and we might perhaps motivate that

option by saying that these ideas are more well-suited for the assessment of individuals' own actions, and not well-suited for the assessment of the decisions and actions related to the design of gamified activities. However, there is also another way of thinking about the responsibility of the game designers that might get them back on the hook, so to speak, without our having to abandon any of these three ideas.

We can say the following: the outcome that the game designers are directly aiming for – and for which they can therefore be held fully responsible and potentially be blamed for – should be seen as a broader outcome than the side effects of the players' actions within the gamified activities taken on their own. What the game designers directly aim to achieve through their own actions is not, we could plausibly say, the outcomes that are side effects of the gamified activities. What game designers directly aim for is, rather, *that game players are led to act in ways that bring about certain outcomes as side effects of their participation in the gamified activity.*

Importantly, what is in italics in the previous sentence is not a mere foreseen side effect of the decisions and actions of the game designers. It is rather what they are directly aiming to achieve. So, if it is judged that there is something bad about the fact *that game players are led to act in ways that bring about certain outcomes as side effects of their participation in the gamified activity*, then the game designers can be seen as fully responsible for this. And this conclusion would be licensed by all of Williams' point, the normative Knobe effect, and the double effect doctrine. At least this applies to cases where there might be reason to judge the side effects to be bad or in other ways problematic.[8]

In contrast, if there is some good side effect of the gamified activity, and the game designer knows about this, then there is no guarantee in the same way that they are fully responsible in a praiseworthy way for bringing it about *that game players are led to act in ways that bring about certain outcomes as side effects of their participation in the gamified activity*. Why not? Because bringing it about that game players act so as to produce the good side effects may not be what they are ultimately aiming for: it could rather be a means for motivating the game players to participate in the game, where there are also certain other side effects (e.g., advertising their brand, collecting data, or whatever) which is the thing that they are really ultimately trying to achieve. It is primarily, in other words, if (1) game designers are really mainly trying to make people act in certain ways that have certain side effects that can be judged as good, and (2) there are no other side effects they are ultimately more interested in bringing about in the end, that they can be seen as responsible in a very clearly praiseworthy way for influencing people to act in these ways with these positive effects.

In other words, from the point of view of how to assess game designers' actions, the outcomes of their actions can be thought of in terms of how they are influencing – or perhaps even manipulating – people to act in certain ways, where those ways of acting are predicted to have certain side effects. If there are some such side effects that can be judged to be good, but that are either merely foreseen side effects or if they primarily function as means for tempting the players to behave in these ways –

---

[8] My thinking here is partly influenced by recent discussions about how to fill or dissolve responsibility gaps related to autonomously operating AI technologies. For an overview of recent discussions of that topic, see chapter six of Nyholm, 2023. For further discussion, see also chapter three of Nyholm, 2020.

whereas the main side effects the game designers are hoping to bring about are other ones, and those side effects can be judged as problematic – then the game designers are primarily responsible for bringing it about that other people act in ways that generate the problematic side effects.

Accordingly, if we only wish to assess responsibility for the outcomes that are side effects of the game players actions within the gamified activities, then it is primarily the players who can be held responsible for these effects. If the type of outcome we are assessing is instead the broader state of affairs *that the players are being incentivized to behave in ways that bring about these side effects*, then we can also assess – and should perhaps primarily assess – the responsibility that the game designers have for this. They are the ones who are introducing game-like elements into activities in order to motivate people to engage in those activities and thereby bring about certain outcomes as effects of the players' participation in the activities in question.

6 Concluding Discussion

How plausible are the conclusions about praise and blame for side effects produced in gamified activities that were derived from Williams' point, the normative reading of the Knobe effect, and the double effect principle in the section above? In these concluding remarks, I will specifically focus on the idea that if we bring about good outcomes as side effects of participating in gamified activities, this may not be a way of producing these outcomes in a clearly praiseworthy way.

One thing I would like to point out – to follow up on something mentioned in the introduction – is that this conclusion about praise (or lack thereof) for positive side effects of gamification fits with an argument that Tae Wan Kim has formulated in a short 2015 paper about the ethics of gamification. Using the ALS challenge as his example, Kim (2015:3–4) writes the following (and this is worth quoting at length):

> Suppose that Alan genuinely cares about ALS patients and hopes more people become aware of the disease. He knows that his friends Kent and Taylor like to be recognized by others on Facebook. So Alan takes the ice bucket challenge himself on his Facebook and then nominates Kent and Taylor. During the preparation for the challenge, Kent and Taylor realize that some filmed challenges are "liked" and "shared" many times […] while others are simply ignored. So they decide to film their pouring ice bucket with a certain funny idea. The mention ALS in passing, but do not mean it. Many "like" and "share" their video. Kent and Taylor are excited about the points and badges. Many who would not otherwise know about ALS are now aware of ALS. Alan is happy about the outcome. So, everyone becomes happier.

The question here is if everything is as good as it could be, from a moral perspective. Perhaps it is not. From the point of view of the normative reading of the Knobe effect, as discussed above, it might be argued that Kent and Taylor fail to intentionally promote ALS awareness in a clearly praiseworthy way. After all, what they really want to achieve are likes and shares on social media. And the promotion of ALS awareness

is merely a foreseen side effect, not something they would be independently motivated to pursue for its own sake.

Kim (2015:3–4) argues in a parallel fashion, when he writes the following:

> Participating in the ice bucket challenge is itself a desirable act, and the worthiness of an action in part depends on the desirability of the act itself. Nonetheless, two actions that are equal in moral desirability may be of different moral worth, because the worthiness of an action also significantly depends upon the extent to which one is motivated to perform it with reasons that make it desirable.

Kim also morally assesses Alan's involvement:

> I submit that Alan [in effect] attempts to get Kent and Taylor's decision-making to fall short of a certain important moral ideal […], due to the influence of game design elements, [whereby] a decision maker becomes detached from the reason that makes her action desirable, which can put the action at risk of significantly losing its moral worth.

If Kent and Taylor truly do not care one way or the other about promoting ALS awareness, this does seem to be the right conclusion. And the argument sketched above based on the normative reading of the Knobe effect also seems to get the right conclusion. That is, Kent and Taylor are not promoting ALS awareness in a clearly praiseworthy way. And their actions (desirable though they are in terms of some of their consequences) lack the moral worth of the actions of somebody who would be motivated to promote ALS awareness for its own sake, because they think that this is an important goal.

Regarding Alan in this example, we might think that his action of creating a gamified activity for Kent and Taylor that leads them to promote ALS awareness as a side effect of their social media activity is more in the direction of something that could be seen as a praiseworthy decision or set of actions. Kim, though, suggests in the quote above (and in the rest of his paper) that Alan's actions may also fall short of a moral ideal because he is in effect creating a situation in which Kent and Taylor are led to act in ways that lack moral worth. Ultimately, it would be better, from this point of view, if Alan could inspire Kent and Taylor to act in desirable ways, not because their activities have been gamified, but because Kent and Taylor can directly appreciate and be moved by the reasons that support acting in favor of this particular good cause. Kim goes so far as to conclude that Alan's actions are in a certain respect "manipulative".

A question, then, is whether gamification always has something problematic about it, even in cases where we gamify activities with the goal of incentivizing people to act in ways that bring about good outcomes. Do gamified actions and activities of the sort that this paper has been focusing on always fall short of being ways of producing good outcomes in clearly praiseworthy ways? Perhaps this depends on whether there is a complete lack of motivation to produce the good outcome on the part of the people whose activities are gamified.

Suppose that in a gamification version of the Knobe effect vignette from Sect. 4 above, we have the following scenario: the adviser comes to the CEO and says, "there is a competition between different companies organized by the environmental association, where whoever comes up with a new product that is good for the environment gets a prize – namely, a lot of money – and will be seen as being the best company". The CEO, who is very competitive, says, "I don't care about the environment at all, but I want to win a lot of money. More importantly, I really want our company to win any competition against our competitors. So, go ahead and tell the product development department to come up with some maximally environmentally friendly new product!" The product development department goes to work and comes up with a new product. It is environmentally friendly, and it is more environmentally friendly than the products the competing companies come up with. So, the company that is run by the CEO makes a lot of money and wins the competition.

Did the competitive CEO intentionally help the environment in a clearly praiseworthy way by participating in this game/competition? Here we might be inclined to say "no". But suppose now that the CEO says the following in another variation of the case, after having heard about this competition: "I do care about the environment, but I normally cannot bring myself to do anything that would help the environment. I am, however, very competitive. Now that there is a competition that can motivate me, let's go ahead and get our product development department to develop an environmentally friendly product. That way, I can do something for the environment. Admittedly, I lack sufficient motivation to do that for its own sake alone, but the extra boost that the competition gives me helps to move me to decide to do this." Again, the product development department succeeds in developing an environmentally superior product. The product helps the environment, and the CEO's company wins the competition.

In this case, does the CEO deserve at least a little bit of praise? Suppose further that Kent and Taylor are also like the CEO in their motivations: they care a little bit about ALS awareness, but not enough to be moved by this idea for its own sake. However, the prospect of getting attention on social media moves them to do something (participate in the ice bucket challenge) that helps to spread awareness about ALS. Does their using this opportunity to participate in a game help them to deserve some amount of praise and do they now qualify as bringing attention to ALS in a partly praiseworthy way?

I think that in these two variations of these cases, we can plausibly say that the CEO as well as Kent and Taylor deserve at least a certain amount of praise. At least they are clearly more praiseworthy in comparison to the original variations of these cases above. And yet they may not count as helping the environment or promoting ALS awareness in ways that are as clearly praiseworthy as their actions would be if they would be motivated to take action to help the environment or promote ALS awareness without any need to turn their activities into a game or some form of competition.

In short, the more gamification is needed to make people act in ways that promote good outcomes, the weaker their level of praiseworthiness for acting so as to produce these good outcomes is. Or so it might plausibly be thought. Does this matter, one might ask here, if the outcomes are good and important enough in the end?

In certain ways, it may not matter. In some cases – e.g., when it comes to saving the world from devastating climate change – it may not matter all that much whether people's activities or behaviors that lead to the saving of the climate are completely clearly praiseworthy activities that do not involve any forms of gamification to motivate them. The end in view might sometimes be so important that whatever can move people to act should be encouraged, whether or not this might involve some form of gamification that might lead the people to be a less clearly praiseworthy for their actions than if they could be moved to action without the need for any gamification.

However, we do often care about whether our own and other people's actions and behaviors are praiseworthy or whether there are factors – such as the use of gamification or whatever it might be – that makes it less clear whether we are praiseworthy or blameworthy for our actions. The types of ideas discussed in this paper clearly bear on this issue. The aim of this paper has not been to settle these questions once and for all. But hopefully the discussion above may stimulate further discussion of the intriguing topic of whether adding game-like elements to activities might in significant ways change what the different people involved might justifiably be praised or blamed for.

Declarations

Ethics Approval Does not apply. This is a purely theorical paper.

Conflict of interest None.

References

Allen, F. E. (2011). : Disneyland Uses 'Electronic Whip' on Employees, *Forbes*: https://www.forbes.com/sites/frederickallen/2011/10/21/disneyland-uses-electronic-whip-on-employees/?sh=76b0881251b3 (accessed on April 12, 2023).

Arora, C., & Razavian, M. (2021). Ethics of Gamification in Health and Fitness-Tracking. *International Journal of Environmental Research and Public Health*, *18*(21), 1–19.

Coeckelbergh, M. (2020). Artificial Intelligence, responsibility attribution, and a relational justification of Explainability. *Science and Engineering Ethics*, *26*, 2051–2068.

Danaher, J., Nyholm, S., & Earp, B. (2018). The quantified relationship. *American Journal of Bioethics*, *18*(2), 3–19.

De Jong, R. (2020). The retribution-gap and responsibility-loci related to Robots and Automated technologies: A reply to Nyholm. *Science and Engineering Ethics*, *26*(2), 727–735.

Edmonds, D. (2013). *Would you kill the Fat Man?* Princeton University Press.

Fogg, B. J. (2003). *Persuasive technology: Using computers to Change what we think and do*. Morgan Kaufmann.

Gorin, M. (2021). Gamification, Manipulation, and domination. In F. Jongepier, & M. Klenk (Eds.), *The Philosophy of Online Manipulation* (pp. 199–215). Routledge.

Hevelke, A., & Nida-Rümelin, J. (2015). Responsibility for crashes of Autonomous vehicles: An ethical analysis. *Science and Engineering Ethics*, *21*(3), 619–630.

Kamm, F. (2015). *The Trolley mysteries*. Oxford University Press.

Kim, T. W. (2015). : Gamification Ethics: Exploitation and Manipulation, *CHI 2015, Gamifying Research Workshop Papers*: 1–5.

Kim, T. W., & Werbach, K. (2016). More than just a game: Ethical issues in Gamification. *Ethics and Information Technology*, *18*(2), 157–173.

Knobe, J. (2003). Intentional Action and Side effects in ordinary Language. *Analysis*, *63*(279), 190–194.

Knobe, J. (2010). Person as scientist, person as moralist. *Behavioral and Brain Sciences*, *33*(4), 315–329.

Lanzing, M. (2019). : *The Transparent Self: A Normative Investigation of Changing Selves and Relationships in the Age of the Quantified Self*. PhD Thesis, Eindhoven University of Technology: https://research.tue.nl/en/publications/the-transparent-self-a-normative-investigation-of-changing-selves.

List, C. (2021). Group Agency and Artificial Intelligence. *Philosophy & Technology*, *34*, 1213–1242.

List, C., & Pettit, P. (2011). *Group Agency*. Oxford University Press.

Marczewski, A. (2017). The Ethics of Gamification. *Crossroads*, *21*(1), 56–59.

Maslen, H., Savulescu, J., & Hunt, C. (2020). Praiseworthiness and motivational enhancement: 'No Pain, no praise'? *Australasian Journal of Philosophy*, *98*(2), 304–318.

Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning Automata. *Ethics and Information Technology*, *6*(3), 175–183.

Nguyen, C. T. (2019). Games and the art of Agency. *Philosophical Review*, *128*(4), 423–462.

Nguyen, C. T. (2020a). How Twitter Gamifies Communication. In J. Lackey (Ed.), *Applied Epistemology* (pp. 410–436). Oxford University Press.

Nguyen, C. T. (2020b). *Games: Agency as Art*. Oxford University Press.

Nihlen-Fahlquist, J. (2017). Responsibility analysis. In: SO Hansson (Ed.), *The Ethics of Technology: Methods and approaches*. Rowman & Littlefield.

Nyholm, S. (2020). *Humans and Robots: Ethics, Agency, and Anthropomorphism*. Rowman & Littlefield International.

Nyholm, S. (2023). *This is Technology Ethics: An introduction*. Wiley-Blackwell.

O'Brien, L. (2015). Side effects and Asymmetry in Act-Type Attribution. *Philosophical Psychology*, *28*(7), 1012–1025.

Parmer, J. (2021). Manipulative design through Gamification. In F. Jongepier, & M. Klenk (Eds.), *The Philosophy of Online Manipulation* (pp. 216–234). Routledge.

Persson, I. (2013). *From morality to the end of reason*. Oxford University Press.

Persson, I., & Savulescu, J. (2012). *Unfit for the future*. Oxford University Press.

Pettit, P. (2015). *The Robust demands of the good*. Oxford University Press.

Porsdam Mann, S, Earp, BD, Nyholm, S, et al. (2023). Generative AI entails a blame-credit asymmetry. *Nature Machine Intelligence*.

Rubel, A., Castro, C., & Pham, A. (2019). Agency laundering and Information Technologies. *Ethical Theory and Moral Practice*, *22*(4), 1017–1041.

Scanlon, T. M. (2008). *Moral dimensions*. Harvard University Press.

Smids, J. (2018). : *Persuasive Technology, Allocation of Control, and Mobility: An Ethical Analysis*, PhD Thesis, Eindhoven University of Technology: https://research.tue.nl/en/publications/persuasive-technology-allocation-of-control-and-mobility-an-ethic.

Spahn, A. (2012). And lead us (not) into Persuasion… Persuasive Technology and the Ethics of Communication. *Science and Engineering Ethics*, *18*(4), 633–650.

Sparrow, R. (2007). Killer Robots. *Journal of Applied Philosophy*, *24*(1), 62–77.

Van De Poel, I., et al. (2015). *Moral responsibility and the Problem of many hands*. Routledge.

Véliz, C. (2020). *Privacy is Power*. Penguin.

Werbach, K. (2014). (re)defining gamification: A process approach. *Persuasive Technology: Lecture Notes in Computer Science*, *8462*, 266–272.

Williams, B (1973). A critique of utilitarianism. In Smart, J.J.C.& Williams, B. *Utilitarianism: For and Against*. Cambridge University Press. 77–150.