The value of value capture

Discussion Note

The Value of Value Capture: Using Gamification to Address Addiction and Sequential Choice Problems

Michael Ridge[*]

Gamification, roughly the use of game-like elements to motivate us to achieve practical ends "in the real world," makes large promises. According to Jane McGonigal, whose book is a New York Times bestseller, gamification can save the world by channeling the amazing motivational power of gaming into pro-social causes ranging from alienation from our work to global resource scarcity and feeding the hungry (McGonigal 2011). Even much more modest aims like improving personal fitness or promoting a more equitable division of household labour provide some license for optimism about the ability of gamification to improve our lives in more humble but still worthwhile ways. A variety of concerns have been raised about gamification, including that it can lead to the exploitation of workers and is motivated by corporate greed (see, for example, Bogost 2014 and Rey 2014). Some of these concerns apply more forcefully to the use of gamification in the workplace, and I share those concerns too. In my view, gamification is much more promising outside the workplace, where people embrace it autonomously as a means to motivating themselves.

In this paper I will focus on a concern pressed forcefully by Thi Nguyen, which applies equally even to this more promising form of gamification. Nguyen argues that there is a dark side to gamification: "value capture." Roughly, gamification works in large part because it offers a simplified value structure—this is an essential part of its appeal and motivational power. According to Nguyen, simplified values are a large part of the appeal of games more generally. The simple values they provide offers a kind of clarity and a relief from the more complex moral and social values we must navigate in the real world. However, especially in the context of gamification which exports these

---

value schemes into our real-world lives, there is a risk that these overly simplistic models will displace our richer, subtle values and that this will make our lives worse: this is value capture. Value capture can then be defined as the replacement of rich and subtle values in the real world with simplified (often quantified) values, where those new values replace the original, richer ones in our motivational economy.

Nguyen's point is well-taken. The way in which number of steps taken per day can, for an avid user of "FitBit," displace more accurate measurements of how one's activities contribute to one's fitness is a compelling example. If I become so obsessed with "getting my 10,000 steps" that I stop making time to go to the gym, jog or do my yoga/Pilates then that is not a net gain. However, there is an important range of cases that Nguyen's discussion ignores but which provide an important exception to his critique: value capture relative to behaviours that are addictive and destructive. Here I have in mind things like alcoholism, drug addiction, and gambling addiction. With these kinds of activities, value capture can not only be good but essential to a person's well-being because of (and not despite) its displacement of the person's more rich, subtle values. Interestingly, the point is not limited to cases of addictive behaviour, though they put the point in its most sharp relief. Any situation in which making rational decisions one by one can leave one worse off than "blindly" following a policy which is itself rational to adopt also turns out to illustrate the point, thus further expanding the role for value capture as itself a force for good. The more general point is that certain kinds of sequential choice problems carve out an important and theoretically interesting exception to Nguyen's worries about value capture. In these kinds of choice contexts, value capture not only does not make our lives go worse, it may even be essential to making our lives go better.

In fairness, Nguyen himself does recognize that there are forms of gamification which avoid his critique. Nguyen clearly intends only to issue an important cautionary note: "Gamification: Use Judiciously and With Extreme Caution!" rather than "Never Gamify!" However, the exceptions he explicitly recognizes all seem to be ones in which the correct values are very simple to begin with. The point is that in these cases the risk of gamification replacing subtle and more complex values with more simple and less accurate ones is minimal. Here is a representative passage:

> Clearly, gamification can sometimes be a tool for getting around weakness of the will, and so enhancing our autonomy. A system like Duolingo is a particularly good case of autonomy-enhancing gamification, because the target of the system – learning a language – is a relatively clear goal. Learning a language is, by its nature, a clear and simple target, and is easily translated into a set of game-like goals with little loss of content. (Nguyen 2020, 210)

Although he never explicitly says that this is the only sort of exception worth considering, all the cases he discusses that avoid his worry about value capture are cases in which the correct values are simple to begin with. This suggests that Nguyen's tacit view is that unless the target goals are relatively clear-cut and simple then we should

probably avoid gamification because of the risk of value capture. Or, perhaps more accurately, we should in that case at least be very cautious about using gamification unless we can be sure that doing so will not leave us with overly simplistic and less accurate values. Perhaps there are forms of designing gamification so that it does not simplify the rich and complex values with which we start, but otherwise we should reserve gamification for activities in which (like learning a language) the original "target" is relatively clear cut. Use gamification only when either (a) the correct values are simple enough to be well-modelled by the values of the gamification in question or (b) the gamification is subtle enough to capture the correct values in spite of their complexity and subtlety, and somehow does so without losing the motivational advantages of gamification which typically arise out of simplicity and value clarity to some extent[1].

This is a natural view, but it misses something important. One way of expressing the value capture worry is that value capture leaves us with "the wrong values." However, there is here a potential equivocation on 'wrong values'. This might simply mean 'inaccurate values'— values which are simply incorrect—taking something to be good, or better than something else, when it is not. If, for example, we assume an objective view of human well-being then there will ways of understanding well-being which are simply false. If a complex "objective list" theory of well-being is correct then someone who endorses a simple hedonistic view of well-being is simply getting it wrong— they will have false evaluative beliefs about well-being. Call this the alethic sense of 'wrong values'. On the other hand, by 'wrong values' we might mean values which are such that endorsing them will make your life go worse relative to the relevant objective values. These values might be prudential or moral, depending on the context. Call this the practical sense of 'wrong values'.

The possibility of this equivocation is easily missed because these two tend to go hand in hand. If we endorse the wrong values in the alethic sense then we will tend to pursue those wrong values and this, in turn, will typically mean we do worse than if we had pursued the right values, thus making those the wrong values in the practical sense as well. The case of someone who quits going to the gym and lifting weights in order to just "get more steps" on the FitBit because he comes to believe (implicitly) that is the best way to pursue his health is a clear example. Still, the two senses of 'wrong values' are conceptually distinct and sometimes they can come apart. Indeed, it is precisely when they do come apart that I want to argue that gamification can be a useful tool, precisely because it replaces our rich, subtle values with more simplistic and less accurate ones. Indeed, it might be that without gamification we would struggle to adopt the val-

---

[1] Another caveat I lack the space to explore here but worth mentioning is that forms of gamification that are designed to encourage playfulness may mitigate Nguyen's concerns about value capture. I discuss the need to incorporate what I call the "playful stance" in Nguyen's work in Ridge 2021.

ues we need to adopt because it will be too clear to us that they are mistaken. We may need gamification to help us adopt the needed values in much the same way that someone who is convinced by Pascal's Wager needs church to adopt the beliefs they think it is in their interest to have but which they otherwise will struggle to adopt due to the apparently decisive counter-evidence. The very thing that makes value capture insidious and problematic in other cases is what makes it so useful in these cases.

Consider alcohol. For a given person, it might be ideal for that person to drink moderately. They enjoy wine, suppose, and they enjoy socializing with others in contexts in which alcohol acts as a "social lubricant" for all concerned. They also enjoy unwinding with a nice glass of wine with dinner, especially one that complements their meal. On the other hand, by far the worst outcome for this person would be becoming an alcoholic, which would destroy their career, their family and their reputation, leaving them miserable with low self-esteem and compromised autonomy. Let us further suppose that this person is (due in part to their genetic endowment) considerably more at risk of alcoholism than most people and they have gone through problematic periods of binge drinking in the past. This outcome is so bad that even a small risk of it would mean it would be wiser for the person to abstain from alcohol altogether than to drink moderately. However, the risk of this person becoming an alcoholic on the basis of any single drink is, we may stipulate, extremely low. Indeed, the risk is low enough for the benefits of drinking to outweigh the risks in terms of expected utility. The person may therefore rationally decide to have a drink on each of a number of occasions, when after enough such occasions the overall risk of her becoming an alcoholic is unacceptably high. This is a problem, and indeed a recognized problem in the literature on rational choice; it is a form of sequential choice problem (see, e.g. Bermudez 2010). In effect, each of the "person stages" of the person who faces a choice to drink is in a many person prisoner's dilemma with all the other person stages who face that choice. While each person stage is better off having the drink than not having it, no matter what the other person stages do, it is also the case that each person stage does better off if they all refrain than if they all drink – and indeed, each person stage is much worse off if they all drink. There is a dominance argument for each person-stage to drink, but if they all follow this policy then the person made up by these stages will predictably and tragically end up being much worse off.

This is a practical problem. One recognized solution to this sequential choice problem is for the person to "convince herself" that the disvalue of even a single drink is so great as to make it not worth the risk. Here it can help to remind herself vividly of the costs, perhaps reviewing internet discussions of the harms of alcoholism, watching moving testimonials by recovering alcoholics, look at the damage done to livers by cirrhosis, and so on. David Hume's discussion of how the drunkard should vividly keep before his mind how his drunkard friend has died from his drinking problem. The

drunkard "has seen his companion die of his debauch, and dreads a like accident for himself. But as the memory of it decays away by degrees, his former security returns and the danger seems less central and real." (Hume 1978, book 1, part 3, section 8, p. 144) Richard Joyce makes use of this example in his discussion of moral fictionalism, as providing a useful model (Joyce 2001, 216) He suggests that the drunkard should not only keep the image of his unfortunate friend vivid before his mind, to help keep him on the straight and narrow. He should also ideally think "If I drink, that's what will happen to me," even though this is false or anyway potentially false and epistemologically unjustified. What is true and well justified is the more nuanced judgment, "If I drink then there is a 10% chance [say] of this happening to me." But, Joyce suggests, that thought is dangerous. If the drunkard thinks that then he will effectively take too great a risk of lapsing into disastrous alcoholism and binge drinking. This kind of case nicely illustrates the equivocation on 'wrong values'. In the purely alethic sense, the judgment "I ought never drink, since if I ever drink again then I will definitely end up like my poor friend" is a wrong value as this judgment is, let us suppose, false. However, it may be that this judgment is (or anyway reflects) the right values in the practical sense. If the agent endorses this thought then his life will go better—he will avoid undue risk of alcoholism and all its costs.

Although my example relied on the risk of addiction, this is not essential to the relevant choice problems with the structure of a many person stage's prisoner's dilemma. Smoking is, in fact, highly addictive, but even if it were not addictive (but still dangerous in the same way over time), it would provide a good example. The risk of getting cancer (or emphysema, asthma, etc.) from a single cigarette is so low as to be miniscule. So if someone gets a lot of pleasure from smoking, on each occasion of choosing whether to smoke a single cigarette it will be rational to do so. Yet consistently following this policy for long enough will pose a risk of cancer (etc.) great enough to make that policy disastrous. This problem does not rely on the fact that smoking is addictive. You just need a choice context with this diachronic structure. Richard Holton poses the problem well:

> I reason that since smoking forty cigarettes a day for the rest of my life will make a considerable difference to my chance of getting lung cancer, I should give it up. But should I deny myself the cigarette I was about to have? Smoking one cigarette will make very little difference to my chances of getting lung cancer, and anyway, that is all in the future. In contrast, one cigarette can give me considerable pleasure now. So why deny myself? Unfortunately the same argument will work just as well forty times a day for the rest of my life. What we need here is not just an intention to perform—or in this case, to refrain from performing—a specific action. We need a general intention concerning a certain type of action: what Bratman calls a policy. (Holton 2009, 10)

It is not hard to see why adopting and sustaining such a Bratmanesque policy is challenging. The temptation to reconsider any such general policy will often be overwhelming, and indeed arguably is rational. We need to be resolute choosers who simply

do not reconsider our policies unless relevant new information comes to light, but as a practical matter, this can be very difficult. When we add addiction to the mix it can be virtually impossible.

How does this matter to gamification and value capture? These kinds of cases carve out an exception to Nguyen's presumptive case against gamification that involves value capture that is interestingly different in character from the sorts of exceptions he explicitly recognizes. Recall that the exceptions Nguyen allows for are ones in which the original values are fairly simple and clear-cut, as with learning a language and Duolingo. In the case of smoking and drinking the relevant original values—those that would capture the ideal course of action for the agent, may well be rich and subtle. When one really ought to have a drink, for example, is often a subtle matter of judgment relying on multiple complex factors. Value capture here might indeed entail adopting much simpler and less accurate values as when someone decides that the right evaluative judgment is that they should never drink. The point, then, is simple. This will be the wrong value in the alethic sense, but not in the practical sense. The agent's belief that her life would go best if she never drank is false. Ideally, she would drink occasionally and in moderation. At the same time, the agent's belief is good for her— she is better off adopting this highly oversimplified value scheme than she would be adopting the more accurate one. The risk of alcoholism with the more accurate values is simply too great. The same point obviously applies in many cases of smoking, or indeed any case with this sequential choice structure. Nguyen's critique of value capture moves seamlessly from "we adopt oversimplified and hence wrong values" to "our lives go worse." But this argument is sound only if 'wrong values' in the premise that wrong values make our life go worse is read as 'wrong values' in the practical sense. In cases like the ones discussed here, though, the claim that value capture generates the wrong values is true only in the alethic sense. The argument's premises are therefore all true only if there is an equivocation on 'wrong values' between these two premises, making the argument invalid. The critique therefore does not go through. Value capture in these cases does not make our lives go worse. In fact, value capture makes our lives go better.

Indeed, these are cases in which value capture is good for the agent precisely because it helps encode false evaluative beliefs. Those false, oversimplified beliefs help the agent avoid undue risk of alcoholism (or lung cancer in the smoking case, etc.). Moreover, adopting and sustaining such false beliefs is not easy. Belief plausibly aims at the truth, after all, and clinging to false beliefs in the face of countervailing evidence can be challenging. Again, recall the practical problem of the atheist who accepts Pascal's wager. He thinks it in his interest to believe against what he takes to be the balance of evidence, but since belief "aims at the truth" this is not something he can do at will. Instead, he must manipulate himself by attending church, hanging out with reli-

gious people, etc. Gamification for addicts can be like church for atheists convinced by Pascal.

If there are forms of gamification which make this kind of value capture more feasible then they would be making our lives go better precisely because they lead to value capture—they help agents sustain the needed belief in useful fictions. As it happens, there are a range of gamification apps that help people deal with addiction; see Nichols (2018) for some examples of gamification aimed at helping people struggling with alcoholism.

Plausibly, many of these forms of gamification will lead to value capture by inculcating a belief that one should never drink in much the same way that FitBit inculcates false beliefs about the magic of 10,000 steps per day. Here is a representative characterization of how one of these forms of gamification ("Nomo") works:

> Nomo accurately breaks down the time you are sober to the minute, which means that it can show you your sobriety time in terms of years, months, weeks, days, hours, and minutes. Every time you reach a recovery milestone, Nomo issues a chip to celebrate your achievement and keep you motivated. (Nichols 2018)

It is not hard to see how engaging with this form of gamification could help reinforce a belief that one is only "one drink away from being an alcoholic" and make the kind of needed resolute choice more feasible. In fact, this may be false, strictly speaking. It might well be that the chance of alcoholism from just a single drink is very small, but the cumulative risk of drinking regularly leading to alcoholism may be great. In that case, having the wrong values in the alethic sense can help the agent have the right values in the practical sense. Unsurprisingly, similar forms of gamification and useful value capture can be found in the context of smoking too, see Nichols (2017).

More generally, value capture in cases with the sequential choice structure laid out here provide a powerful tool for making the needed resolute choices. So although Nguyen's admonition to treat gamification with a healthy dose of scepticism and caution because of the risk of value capture is, in general, well taken, there is another important range of exceptions worth bearing in mind when taking a broader perspective. Sequential choice problems like the ones discussed here reveal that value capture in the sense of inculcating oversimplified and mistaken values can itself be essential to promoting the agent's long-term interests.

I close by noting that matters are complicated in any given real-world case and of course the empirical details will matter to whether the possibility canvassed here is actually realized. The thought that "I'm only one drink away from being a drunk" is promulgated by Alcoholics Anonymous, and this reliance on useful fictions and "wrong values" might seem to provide confirmation for the views defended here. However, some evidence suggests that promulgating this kind of absolutism can become a self-fulfilling policy, so that people who fall off the wagon at all just assume all is lost and

start binge drinking (see, e.g. Ogborne 1989). This point is also well-taken, but it may just mean that the empirical details matter and that precisely which belief, if any, provides the needed useful fiction may well vary from case to case. Perhaps in the case of alcoholism the useful fiction is the less absolutist "The risk of having a drink is never worth it." Perhaps in fact, the risk sometimes is worth it, but believing this is just too dangerous, over time, to the person at risk of alcoholism. This less absolutist judgment might minimize the risk of self-fulfilling prophecy too, since the belief does not entail that having one drink ensures that the person will become an alcoholic—only that there is a significant risk of this outcome. The broader upshot of this paper is therefore not that gamification is guaranteed to be a good strategy for dealing with addiction; it will depend on the empirical details in each case. The point is rather that there is a sense in which value capture itself can be desirable and that we need to bear that possibility in mind when assessing gamification aimed at dealing with addiction and other sequential choice problems of the sort canvassed here.

References

Bermudez, J. 2010. "Pitfalls for realistic decision theory: an illustration from sequential choice." Synthese: 23-40.

Bogost, I. 2014. "Gamification is Bullshit." In Walz and Deterding (ed.) *The Gameful World*. Cambridge, MA: MIT Press: 65-81.

Holton, R. 2009. Willing, Wanting, Waiting. Oxford: Oxford University Press.

Hume, D. 1978. Treatise of Human Nature [1739]. Ed. L.A. Selby-Bigge. Oxford: Oxford University Press.

Joyce, R. 2001. The Myth of Morality. Cambridge: Cambridge University Press.

McGonigal, J. 2011. Reality is Broken: Why Games Make Us Better and How They Can Change the World. New York: Penguin.

McClennan, E. 2012. Rationality and Dynamic Choice. Cambridge: Cambridge University Press.

Nichols, H. 2017. "Best Apps For Quitting Smoking." Medical News Today. https://www.medicalnewstoday.com/articles/317633#Quit-Smoking-Slowly.
_____. 2018.

Healthline Media, n.d. "Best Apps to Stop Drinking Alcohol." Medical News Today. Accessed August 5, 2021. https://www.medicalnewstoday.com/articles/318913#Coach.me

Nguyen, T. 2020. Games: Agency as Art. Oxford: Oxford University Press.

Ogborne, A. 1989. "Some Limitations of Alcoholics Anonymous." Recent Developments in Alcoholism. https://pubmed.ncbi.nlm.nih.gov/2648498/

Rey, P. "Gamification and Post-Fordist Capitalism." In Walz and Deterding (ed.) The Gameful World. Cambridge, MA: MIT Press: 277-296.

Ridge, M. 2021. "Fun and (Striving) Games: Playfulness and Agential Fluidity," 2021. Journal of the Philosophy of Sport.