15 Promoting Vices

Designing the Web for Manipulation

Lukas Schwengerer

1 Introduction

It is Friday evening and you are exhausted from another day of overtime.[1] You want to relax by watching a TV series. Say you want to watch *The Wire*. You wonder where you can access episodes of *The Wire*. You do not have the box set, nor is it running on any channel right now. So you decide to find out which streaming service offers episodes. You open Google Search, and the field for your search input jumps right to your attention. You start to type "The Wire Stream" into the box. As soon as you reach the "W" Google Search suggests to autocomplete to "The Wire". As soon as you reach the "S" Google suggests your intended query "The Wire Stream", so you hit enter. Immediately, Google not only presents you websites that likely tell you which streaming service provides the opportunity to watch *The Wire*, Google Search itself presents all available choices directly at the top. Clearly visible. Impossible to miss. Even when you are exhausted you can find the right streaming service in a matter of seconds. It is that easy.

The Google Search website is a paradigmatic example of so called *user-friendly design* – design that makes it particularly quick, easy and efficient to use a website for the task the user aims to complete. User-friendly design is intended to increase processing fluency in users and successful user-friendly design makes cognitive processing of a website faster and easier.[2] I will focus on tasks aimed at gathering information, but the same ideas can also apply to other forms. The website is designed in a way that makes all required information as obvious as possible and avoids features that are distracting or difficult to access. You likely have also experienced websites that did not include speed and ease of use as a goal during their design process: flickering colours and moving backgrounds fighting for your attention, low contrast in colour that renders text illegible, dead links that prompt frustration and the dreaded Papyrus as the font of choice. And most likely you avoid any such website at all costs. It seems natural to claim that the user-friendly site has epistemic and practical advantages over the user-unfriendly one. Just take the earlier example: it is a lot easier to find out where I can find a streaming service including "The Wire" with a website featuring user-friendly design.

DOI: 10.4324/9781003205425-18

That's the good news, and I will not deny these benefits. But the bad news is that the same design principles also come with a particular epistemic problem: user-friendly design tends to promote an intellectually vicious attitude towards a website. It tends to promote an *overly trusting attitude*.

The guiding idea in the background is that trust is an unquestioning attitude I can have towards a person or object (Nguyen forthcoming). When I trust a friend I will usually not question their information, nor their intentions. Often, my trust is related to a person – such as the friend I trust. But unquestioning attitudes are possible towards objects also. A climber trusts a rope, and I trust the bridge I am walking on. In these cases objects are unquestioned in performing their intended function. Trusting a website can be read in both ways. I can trust an author of a website, or I can trust the website itself. I will work with the latter reading, but an argument with the same structure is available with the former reading. The object reading is easier to combine with a framework of cognitive integration, and it has the benefit of being applicable even when the website would be largely created by algorithms without much deliberate human input.

I can compare the degree of trust we have towards an object or person with the degree of trust the object or person deserves: their trustworthiness. Sometimes, I do not question a person as a source of information, even though I should. They do not know what they claim to know. Similarly, sometimes I trust a rope that I should not trust. I do not question its stability, but it is already damaged and cannot hold my weight reliably. If the trust towards something exceeds its trustworthiness I will speak of an overly trusting attitude. My aim is to show that such a mismatch between trust and trustworthiness can arise for websites because of user-friendly design. The ease and speed with which I parse information from a website changes the trust I have towards a website in a way that is unrelated to the trust the website deserves. For instance, I take the information that Google Search provides me at the top of the results to be obviously correct even though it does not deserve such an unquestioning attitude. I develop this attitude in part because Google Search is especially easy and quick to use, and my psychology functions such that ease and speed of processing induces an increase in trust. Hence, I end up trusting the Google Search result more than I ought to. I have an overly trusting attitude. An attitude that is not justified by the trustworthiness of the website itself that makes me vulnerable to manipulation.

The road to manipulation is straightforward. If a manipulator can induce an unquestioning attitude of trust, then they will be able to manipulate the beliefs of the trusting person without much effort. Hence, if someone wants to manipulate via a website, they can use psychological effects of user-friendly design for the website to generate a gap between the trust users assign to the website and the trust the website deserves. And in doing so a manipulator makes users more intellectually careless and their beliefs easier to manipulate. It is this danger of users being exploited via psychological features affecting trust judgements and the mechanism involved that will be the target of my discussion. User-friendly design is also manipulation-friendly design in this specific way – or so I will argue.

My plan is the following: I will start with a sketch of the argument against user-friendly design based on cognitive integration. Then, in Sections 3 and 4, I develop the groundwork for a refined version of the argument. In Section 3, I give an overview of virtue and vice epistemology, showing how an overly trusting attitude is detrimental for intellectual virtues. In Section 4, I present my preferred version of the extended mind thesis and cognitive integration. Section 5 features the expanded argument in detail with a focus on the empirical support that brings us from user-friendly websites and cognitive integration to the overly trusting attitude. I conclude in Section 6 with directions of how to limit the epistemic badness of user-friendly design without giving up on the benefits.

2 The Argument From Cognitive Integration Against User-Friendly Design

My aim is to show how a website[3] promotes an overly trusting attitude based on its user-friendly design. I do this by treating the website as an artefact that can be cognitively integrated – that can be part of an extended mind. I thereby treat a website akin to a tool that can be used to enhance the abilities of an agent. In particular, I am interested in the epistemic abilities and the epistemic actions of agents. Here an epistemic action is understood as an action with a function to improve an agent's cognition such that some cognitive tasks become easier or in some cases become possible in the first place (Kirsh and Maglio 1994). For instance, we can use pen and paper to enhance our ability to perform arithmetic and then perform epistemic actions in writing and reading numbers and symbols on the paper. Pen and paper become part of the information-processing system and therefore deserve part of the credit for successful performance of a task (Clark and Chalmers 1998). This sort of cognitive integration is independently plausible with significant explanatory power (Hutchins 1995; Clark and Chalmers 1998; Sutton 2006; Clark 2008, 2010; Menary 2010; Heersmink 2015) and when applied to a website constitutes the basis for my argument against user-friendly design. Using the framework of cognitive integration also provides us with a good way to explore the fine-grained mechanisms that lead to an overly trusting attitude.

The core of the argument was first pointed to by Smart (2018, 297) and starts with the assumption that a website is a particular kind of artefact which can be integrated in a cognitive system to varying degree. The explicit reference to degrees of integration already hints at my preferred theory of cognitive integration: a second wave theory of the extended mind (Sutton 2010; Heersmink 2015). In Section 4, I provide further details on this account of cognitive integration. For now, all I need is the idea that cognitive integration of artefacts comes in degrees. Artefacts that are relied on frequently and without much conscious effort (e.g., a white cane or a smartphone) are integrated to a higher degree than artefacts with one-off uses that require a significant conscious effort in interaction (e.g., a ticket terminal at an airport). This idea also provides the basis for the second premise in the argument. User-friendly design of a website leads to higher cognitive integration because it lowers the effort necessary to engage with the website. Of course, user-friendliness is not the only factor. However, given that as a design principle user-friendliness aims at making the user experience as effortless and quick as possible, and that effort and speed of the engagement with an artefact partially constitute how cognitively integrated an artefact is, it seems straightforward to conclude that user-friendly design also promotes cognitive integration. To make my argument work I now need to bring a dimension of trust into the picture. Hence, the third premise is an empirical claim that higher cognitive integration generally comes with higher trust towards the artefact for non-epistemic reasons. Crucially, that trust is not fully warranted as it is not built on a proper epistemic basis.[4] Hence, I conclude that user-friendly design promotes an overly trusting attitude.

In Section 5, I will modify the argument from cognitive integration slightly based on the particular account of cognitive integration I work with. But for now this general structure is sufficient:

1. Websites are artefacts that can be cognitively integrated.
2. User-friendly design promotes cognitive integration.
3. Generally, cognitive integration promotes trust in an artefact to a degree that is not fully epistemically warranted.
4. C: Generally, user-friendly design promotes an overly trusting attitude towards a website and its content.

The work in this argument is done primarily by premise 3, which can be established by paying attention to empirical research on judgements of trust and confidence in relation to the speed and ease of processing information. As I will show later in detail, there is considerable evidence that points towards an increased feeling of trust and the assignment of higher credence purely because information is processed more easily (Alter and Oppenheimer 2009). As Smart (2018, 297) suggests, this empirical research on judgements of trust in relation to the fluency of processing information shows us that properties that constitute higher cognitive integration also come with higher trust in an artefact and its outputs. User-friendly design aims at speed and ease of user interactions, increases cognitive integration and leads to an overly trusting attitude towards a website. This is exactly the conclusion I aim at.

I have now provided a general argument from cognitive integration showing that user-friendly design leads to an overly trusting attitude towards a website.[5] For the rest of the chapter I want to spell out and support the argument in detail. Moreover, I provide an analysis of the mechanisms involved. To do so, I need to build on the not-yet-fully explained notions of trust and cognitive integration before looking at the empirical evidence supporting premise 3.

3 Trust, Intellectual Virtues, and Intellectual Vices

Before arguing for the plausibility of all premises in my argument I need to provide some background on the problems with an overly trusting attitude. After all, I want to show that user-friendly design should worry us epistemically because it leads to an overly trusting attitude. But what is so bad about trusting a website too much?

First of all, it is not all that clear how trust applies to websites. Usually, trust is taken to be an interpersonal affair (Baier 1986). How can I trust a website if I do not treat it as a form of testimony? Trust is often distinguished from mere reliance (e.g., Baier 1986; Hawley 2019). For instance, trusting a chair not to break seems to be mere reliance. Trust proper seems to be normatively laden in ways that trusting the chair is not. I will not blame a chair for breaking – or at least only in jest. And neither are chairs praised for being trustworthy when they do what they are supposed to do. On the other hand, if I trust your word I do not merely rely on your testimony. I blame you as a person if you betray my trust with a lie. When discussing cognitive integration trust has to be understood as a more general term that also applies to artefacts. One option would be to simply stipulate that the term trust here refers to both reliance and trust proper. However, I think a more motivated solution is to analyse trust with Nguyen (Forthcoming) as an unquestioning attitude: trust is a suspension of deliberation. When we trust someone or something we leave aside all questions of whether the person or artefact will be reliable. Trust in this sense is not necessarily targeted towards agents. Nguyen's notion fits well with the notion of trust that is in play in the debates on cognitive integration (Clark and Chalmers 1998; Heersmink 2015). Importantly, this does not commit me to a binary notion of trust. As Nguyen explains, "[o]ne can trust with varying degrees of unreservedness, since one can hold the dispositions with varying degrees of force" (Nguyen forthcoming). This is important for my argument, because theories that allow for degrees of cognitive integration also demand a gradual notion of trust.

I now have an adequate notion of trust in place and can look at a theoretical foundation of the epistemic badness of trusting too much. My suggestion here is the following: an agent's overly trusting attitude leads to behaviour that is epistemically improper. It promotes vicious epistemic behaviour over virtuous epistemic behaviour. To spell out this idea I rely on some general ideas of virtue epistemology as a theory of knowledge that puts the agent and its role in acquiring knowledge at the centre of attention. To understand knowledge – so the virtue epistemologist – I ought to look at what makes potential knowers good or bad thinkers (Battaly 2008). I am limiting myself to epistemic responsibilism,[6] which pays attention to character traits that constitute intellectual virtues and vices (e.g., Zagzebski 1996; Baehr 2015). Intellectual virtues help in acquiring knowledge whereas vices are obstacles to knowledge.

Intellectual virtues in the responsibilist sense are directly impacted by an overly trusting attitude. Take intellectual carefulness. An agent is intellectually careful when they avoid intellectual errors, including the formation of false beliefs (Baehr 2015). To be intellectually careful one has to be aware of the risks of any particular belief-forming process. I need to know in which ways I might go wrong in forming beliefs in order to avoid mistakes. I need to know how easily I could fail in acquiring knowledge through a particular source. Only then I can properly judge how careful I have to be and only the appropriate amount of care is virtuous. Suppose I am reading a newspaper. Being overly careful in forming beliefs based on the newspaper's content is not virtuous because it leads to missing out on knowledge. The newspaper might be a good source of information for the results of the latest football matches, but I am reluctant to base my beliefs about football results on the newspaper. I miss out on knowledge. But being overly careless is not virtuous either, because it leads to false beliefs. Suppose the newspaper has an insufficiently funded science section and frequently misrepresents scientific studies. If I am careless and base my beliefs on the newspaper's science content I end up with false beliefs. I need to be careful to the proper degree – the degree that this particular source of belief deserves. But my judgement on how careful I ought to be can be influenced by the amount of trust I put into a source of beliefs. When I trust the newspaper I will be rather careless because I will not question it as a source of knowledge. This is fine if the newspaper is worthy of my trust, if it is indeed a good source of information. Then my unquestioning attitude usually leads to knowledge. However, if I overly trust a source – if I trust it more than the source deserves – I will not be careful enough. An overly trusting attitude destroys the virtue of intellectual carefulness.

There are similar worries for trust in relation to other virtues. A high amount of trust will lead us to give up on intellectual autonomy to an extent that we ought not to. It will lead to us being less thorough than we ought to and less open-minded. If we highly trust a source, we stop enquiries early and are not willing to take other sources into consideration. All these problematic influences of trust on intellectual virtues stem from the same source. Intellectual virtues all aim at manifesting a character trait to a particular degree in a particular situation. The ideal intellectually virtuous agent is as careful as the situation requires, as autonomous as the situation requires, as open-minded as the situation requires. Rarely anyone fits the ideal, but that is at least what agents should aim for, and what they can get reasonably close to. The ideal is set by the situation the agent is in, and the further we diverge from the ideal the worse epistemic agents we are.

If we consider the effects of trust on intellectual virtues, we can capture the relevant properties of the situation in terms of the trustworthiness of artefacts[7] involved: an intellectually virtuous agent will act in ways that are partially determined by the trustworthiness of relevant artefacts. The amount of intellectual carefulness required is set by the trustworthiness of the artefact. An agent will act intellectually careful if they put trust in the artefact roughly equal to the trust the artefact deserves. Trust and trustworthiness have to match. Whenever they are too far apart, the agent will end up acting in an intellectually vicious way. Even if they might be generally intellectually virtuous, the virtues will be unable to manifest in the concrete situation because of the mismatch between trust and trustworthiness. This in turn leads to epistemically bad consequences: the formation of false beliefs or missing out on knowledge. An overly trusting attitude therefore qualifies as an epistemic vice – it gets in the way of knowledge (Cassam 2019).

I have now shown why an overly trusting attitude should worry us and therefore why user-friendly design should worry us. Putting more trust into an artefact than it deserves leads to intellectually vicious behaviour. It stops us from being appropriately intellectually careful by misguiding us in our judgements. And being intellectually careless makes us a target for manipulation. A website's author can influence beliefs and resulting actions more easily if they can prompt the user to be careless in their belief formation. Careless users form their beliefs in ways that they would not deliberately endorse. This sort of careless belief formation fits with a general idea of classifying "an effort to influence people's choices . . . as manipulative to the extent that it does not sufficiently engage or appeal to their capacity for reflection and deliberation" (Sunstein 2016). By pushing users towards carelessness these users cannot sufficiently manifest their capacities for reflection and deliberation anymore. Hence, they stop forming beliefs virtuously. With these general results in place, I can now come back to developing the argument from cognitive integration in detail. To start, I will expand on my preferred theory of cognitive integration.

4 Cognitive Integration

Humans are proficient in using and shaping the environment to make their lives easier. We do our calculations on paper. We use post-it notes, notebooks or smartphones to remember important tasks. Humans excel in outsourcing cognitive work to the environment. Clark and Chalmers (1998) were the first to use this observation to argue that in all these cases the environment is part of the cognitive process, labelling their view *the extended mind thesis*. Cognition and mental states are not limited to the brain and skull. They leak into the environment.

This thesis is not uncontroversial. Opponents of cognitive integration models suggest that cases used to motivate the extended mind thesis are better explained otherwise because they are too different from our internal cognitive processes (cf. Rupert 2004; Sterelny 2004) or lack features that our internal mental states have (Gertler 2007). Perhaps there are even differences in the nature of content (Adams and Aizawa 2010). I will not discuss these objections here. If you find objections to cognitive integration compelling I can still retreat to the argument from testimony hinted at in Note 5. In this case you can skip directly to Section 5 and the empirical evidence that supports both the argument from testimony and the argument from cognitive integration.

Clark and Chalmers focus on a parity between the functional role the environment plays and the role that something could play inside our brain as the deciding factor for extended minds. For instance, an extended belief has to be functionally on par with a biological belief. In contrast, a second wave of theories of the extended mind (e.g., Sutton 2006, 2010; Menary 2010; Heersmink 2015) focuses on artefacts that expand the cognitive realm and allows humans to succeed in cognitive tasks that often were not possible at all without these artefacts.[8] Besides focusing on the complementary nature of extended cognitive processes, the second wave theorists also leave the largely binary nature of Clark and Chalmers's (1998) model behind. They argue that we can describe our relations to artefacts more appropriately if we think of cognitive integration as covering different dimensions, not on all of which an artefact has to be equally integrated. Take for instance, Heersmink's (2015) suggested framework of dimensions of cognitive integration. In this framework we can evaluate how integrated an artefact is among eight different (although related) dimensions. I take the shorter descriptions of these dimensions from Schwengerer (2021); for the extended presentation, see Heersmink (2015, 582–92):

Information Flow – the directions that information is passed on between an agent and an artefact.

Reliability – the frequency an artefact is used to impact the agent's cognitive processes.

Durability – the permanence of one's relation to an artefact.

Trust – the degree to which one takes the information provided by an artefact to be correct.

Procedural Transparency – the degree of fluency and effortlessness in interacting with an artefact.

Informational Transparency – the degree of fluency in receiving, interpreting, and understanding information from the artefact.

Individualisation – the degree to which an artefact is personalized or can be used by anyone.

Transformation – the degree to which the cognitive capacities of an agent change in virtue of using an artefact.

These dimensions allow a more fine-grade analysis of the human–artifact relationship. For instance, think of a notebook I take with me whenever

I leave my home. My notebook might have a two-way information flow. I write in the notebook and read information from it. I use my notebook only every once in a while, so the notebook is not highly integrated on the reliability dimension. It ranks higher on durability, because I keep the same notebook with me for a long time. It also ranks highly on trust. I rarely doubt what is written in my notebook. If I read that I have to finish this chapter on Friday, I believe that to be the case. Both transparency dimensions are also satisfied to a high degree. There is barely any effort required to open and read my notebook. Given that I usually have only a few recent entries that matter, I can also find the relevant entries quickly and easily. Moreover, because it is written in my own language and in my own style of talking and thinking it does not take much effort to interpret and understand the content either. How the notebook ranks on individualisation is unclear. On the one hand, it is not especially individualised, because anyone can read the contents of, or write into, my notebook. But on the other hand, everything in the notebook is written by me and for me. Finally, the notebook ranks relatively low on the transformation category. All – or at least most – of what I use the notebook for could be achieved by me without the notebook as well. Just with a little less convenience.

For the rest of the chapter I will work with this picture of cognitive integration suggested by Heersmink. The additional flexibility allows this theory of cognitive integration to deal with objections more easily. For instance, a general worry for theories of the extended mind is that too much becomes part of one's mind. The dimensions of integration framework can make this problem more palatable by suggesting that most things around us are integrated to only a very small degree on particular dimensions. They are not fully part of one's mind. More importantly, for my purpose, Heersmink's theory fares a lot better if I want to combine it with virtue epistemology. Whereas for Clark and Chalmers, only highly trusted artefacts can be integrated, Heersmink allows for integration of artefacts even while I do not trust the artefact fully. In his framework, I can distinguish between epistemic dimensions – which consist of only the trust dimension – and the other, non-epistemic dimensions.[9] For instance, a website can be highly integrated on reliability, durability, procedural transparency and informational transparency but still shows only a low integration on the trust dimension. Hence, I can cognitively integrate a website that is not very trustworthy and still remain intellectually careful – an option not available in the Clark and Chalmers account. The integration just has to be limited to the non-epistemic dimensions. And this is exactly what I aim for: cognitive integration that allows one to frequently, quickly and easily perform an epistemic action, without sacrificing on epistemic virtues and standards.

Unfortunately, this is possible only in theory. In practice, humans are a lot worse in isolating the trust dimension from other dimensions of integration. Cognitive integration spills over from the non-epistemic dimensions to the sole epistemic dimension of trust. This empirical claim is at the core of the argument from cognitive integration. And I am now in a position to show why this is the case, before looking for ways that help us isolate different dimensions of cognitive integration.

5 How User-Friendly Design Promotes Vices – The Expanded Argument From Cognitive Integration

Let me start by restating the initial argument from cognitive integration:

1. Websites are artefacts that can be cognitively integrated.
2. User-friendly design promotes cognitive integration.
3. Generally, cognitive integration promotes trust in an artefact to a degree that is not fully epistemically warranted.
4. C: Generally, user-friendly design promotes an overly trusting attitude towards a website and its content.

I am now equipped to modify the initial premises in light of Heersmink's theory of cognitive integration. The first premise can stay as is, but premises 2 and 3 have to be modified. Both premises 2 and 3 are too general with regard to cognitive integration. The argument needs to allow for the conceptual possibility of cognitive integration without an overly trusting attitude. And the dimensions of integration framework make this possible by distinguishing between epistemic dimensions and non-epistemic dimensions. Only in virtue of formulating premise 2 solely with non-epistemic dimensions in mind the full force of the argument will be present. If premise 2 already included high integration on the trust dimension without showing specifically that they result from non-epistemic factors the argument would be question-begging at best. Similarly, what premise 3 aims at is that non-epistemic factors in cognitive integration usually impact the extent to which one trusts an artefact. Only if this connection is established I can conclude that whatever trust is generated by cognitive integration is not fully epistemically warranted. Hence, premises 2 and 3 have to be reformulated as follows:

2. User-friendly design promotes cognitive integration on non-epistemic dimensions (dimensions other than trust).
3. Generally, cognitive integration on non-epistemic dimensions promotes an increase in cognitive integration on the trust dimension in a way that is not fully epistemically warranted.

Perhaps, the additional clause "in a way that is not fully epistemically warranted" is not required, given that the non-epistemic dimensions are responsible for the difference in the trust dimension. However, the clause is still a safeguard against the idea that some of the non-epistemic dimensions could be potentially used as an indicator of the care put into a website – and hence also as an indicator for truth conduciveness.[10] We can now state the extended argument from cognitive integration:

1. Websites are artefacts that can be cognitively integrated.
2. User-friendly design promotes cognitive integration on non-epistemic dimensions (dimensions other than trust).
3. Generally, cognitive integration on non-epistemic dimensions promotes an increase in cognitive integration on the trust dimension in a way that is not fully epistemically warranted.
4. C: Generally, user-friendly design promotes an overly trusting attitude towards a website and its content.

And given the discussion on epistemic virtues and vices in which I showed that intellectual virtues are incompatible with an overly trusting attitude I can now reach a further conclusion:

> C2: Generally, user-friendly design promotes an intellectually vicious engagement with a website and its content.

This is the worry that I have been following throughout the chapter. If the argument is sound we should be apprehensive about websites with user-friendly design because they foster a form of user interaction that makes users intellectually vicious – intellectually careless. What is still missing is the evidence for premise 3. Why should one believe that the trust dimension cannot be isolated from other dimensions of cognitive integration? Why should high integration on non-epistemic dimensions spill over to the epistemic dimension of trust?

My answer here is an empirical claim. Human beings have a psychological make-up that makes it difficult to prevent non-epistemic dimensions from contaminating the epistemic one. Our psychology cannot, or at least not easily, keep the ease and speed of cognitive processes apart from a judgement of trust. When some process comes quickly and easily to a person, they tend to trust the result of that process more, purely for the epistemically irrelevant aspects of speed and ease of processing. Aspects that by and large[11] have no relation to the truth of the output given by that process.

The main source of evidence for this claim are studies about the influence of processing fluency on judgements of trust and credence. The effects of processing fluency are well researched and support a general conclusion that the easier it is to process information, the more likely we are to believe that information (Alter and Oppenheimer 2009). Let me look at a small selection of these studies to illustrate the point, before applying the observations to user-friendly design.

Reber and Schwarz (1999) provide evidence that statements that are easier to read are taken to be more likely to be true. They presented subjects with statements in colours that made them easier or more difficult to read.

For instance, they showed statements in the form of “Town A is in country B” (e.g., “Lima is in Peru”) and varied the visibilities of the colours used. Blue and red were highly visible on a white background but yellow or light blue less so. The experimenters ensured that statements for all visibility ranges were balanced – statements in red were not more obviously true than statements in yellow. After presentation of a statement the subjects had to decide whether the statement was true or false. Subjects were told the colours were meant to measure reaction times with different colours to disguise the actual goal of the study and prevent manipulation. The results show that statements written in colours that could be read more easily were endorsed significantly more frequently than statements written in colours that were less visible. In other words: subjects judged statements to be more likely to be true, merely because they had an easily readable colour. The most plausible explanation is that the information processing was more fluent – it was easier and faster to read the visible colours.

McGlone and Tofighbakhsh (2000) observe a similar effect of processing fluency in the effects of rhyming. Subjects were confronted with aphorism that they were not familiar with that they had to judge on their accuracy on a scale of 1 (not at all accurate) to 9 (very accurate). The complete list of aphorisms featured pairs of rhyming and non-rhyming versions such that for each pair the experimenters could compare the accuracy judgement for the rhyming and the non-rhyming versions. For instance, the list included “Woes unite foes” and “Woes unite enemies”. As a control measure they also included pairs in which neither version was rhyming. For instance, “Good intentions excuse ill deeds” and “Good intentions excuse ill acts”. It turned out that if the subjects were not warned of potential effects of rhyming, they assigned higher accuracy to aphorisms that did in fact rhyme. They propose that “this effect is a product of the enhanced processing fluency that rhyme affords an aphorism such as ‘What sobriety conceals, alcohol reveals’ relative to a semantically equivalent nonrhyming version” (McGlone and Tofighbakhsh 2000, 427). Again, speed and ease of processing comes with an increase in perceived accuracy.

Finally, Oppenheimer (2006) provides evidence that easier to process texts are deemed to be written by more intelligent authors. In particular, he shows that using overly complex words comes with being judged of lower intelligence. This relationship held regardless of the quality of the text in question. This result might not be completely surprising, given that every writing guide suggests simple prose, but it is again further evidence that processing speed and ease impacts judgements about the epistemic merits of some perceived informational content. Oppenheimer explicitly states the results of these judgements are best explained by considering processing fluency (Oppenheimer 2006, 151).

These examples are a mere glimpse at the evidence available. Alter and Oppenheimer’s (2009) meta-analysis includes an abundance of similar studies that all point in the same direction: human psychology infers from processing fluency broadly epistemic features, even when such an inference is not justified. Most importantly, processing fluency leads to more trust and giving higher credence to information processed fluently. I can import these results directly into Heersmink's cognitive integration framework. Processing fluency – the speed and ease of processing – is captured by procedural and informational transparency. I thereby have identified evidence for at least two non-epistemic dimensions of cognitive integration that spill over to the trust dimension. An increase in cognitive integration on transparency dimensions also leads to an increase on the trust dimension, as supported by the empirical evidence. And it seems clear that these non-epistemic dimensions have no relation at all to truth. Take the mentioned colour effect in Reber and Schwarz (1999). It seems obvious that the colour a statement is written in has no connection to the truth of that statement. These effects are exactly what I am looking for to establish premise 3: cognitive integration on non-epistemic dimensions leads to an increase in cognitive integration on the trust dimension in a way that is not fully epistemically warranted. The ease of reading a text increases the integration on the non-epistemic dimension of procedural transparency in a way that also increases the integration on the trust dimension.

Taking a step back the same idea can be applied more generally to user-friendly design – design that makes it particularly quick, easy and efficient to use a website for the task the user aims to complete. Making a cognitive process particularly quick, easy and efficient is nothing else than increasing processing fluency. And given that processing fluency increases perceived trust, premise 3 is established, and I can conclude that user-friendly design leads to an overly trusting attitude towards a website and its content.

One might wonder whether there is an alternative reading available. Perhaps, fluency does not lead to an overly trusting attitude, but lower fluency leads to a lack of trust. This does not seem to be the right interpretation, because studies of repeated presentation of the same content point towards processing fluency influencing judgements of trust beyond the trustworthiness of a source (Hasher, L., Goldstein, D. and Toppino 1977; Begg, I. M., Anas, A. and Farinacci 1992). Hence, there is clear evidence of trust due to processing fluency exceeding trustworthiness of a source.[12]

Of course, even though this is bad news, it need not be terrible news yet. All that I have established is that user-friendly design promotes an overly trusting attitude and therefore also promotes an intellectually vicious engagement with websites. But nothing has been said to the extent of excess in trust and intellectual viciousness. I have not established that user-friendly design always leads to high agential gullibility, the kind of gullibility in which we too eagerly accept an artefact and its processes as trustworthy (Nguyen forthcoming). Trusting a website a little more than the site deserves is perhaps not that big a problem. But the worry looms that developments to make the user experience even faster, even easier and more comfortable brings us to a larger and larger gap between our trusting attitudes and the trust a website deserves. A gap that could be exploited by people aiming to manipulate us for their gain by eliciting false beliefs that prompt actions that we would not have otherwise performed. How can we stop that? This is what I will address in the final part.

6 Fixing the Web

I have established that user-friendly design leads to an overly trusting attitude towards a website. This should worry us, even if I have not shown that the excess of trust is already at a particularly dangerous level. There are at least three different responses available that I will sketch in turn.

First, we could abandon user-friendly design principles. Stop making websites accessible, use illegible fonts and colours. Remove all forms of personalisation that increase the ease of using a website. But obviously this cannot be the way to go. It is a clear case of throwing the baby out with the bathwater. We should not sacrifice all epistemic benefits we get from websites just because of a worry of an overly trusting attitude. Moreover, economic pressures make this option practically impossible. The market forces will always promote user-friendly websites over completely unusable ones.

Second, we could limit user-friendliness. The aim here would not be to stop us from being overly trusting completely but to limit the extent to which our trust exceeds the trust the website deserves. As long as the gap between an agent's trust in a website and the website's trustworthiness is not too big, the potential damage is also limited. An agent might end up with some false beliefs and miss out on some knowledge, but by and large the agent's belief formation will be truth conducive because the agent's behaviour is not too far off from that of an ideal, intellectually virtuous agent. The agent can still be close enough to the required intellectual carefulness. Maybe that is good enough for all our purposes.

How these limits on user-friendliness look in practice is a difficult question. To give you one example of such a limit, consider a law that restricts a website's use of personalisation via tracking cookies. If the website cannot personalise efficiently, then the website loses a tool in increasing user-friendliness. It can no longer predict efficiently what a user wants to do. Hence, the user will likely be required to take an extra step and reduce their processing fluency.

Finally, third, we could look for strategies that stop or compensate the spill from non-epistemic dimensions of integration to the epistemic dimension of trust. This is the ideal solution. It allows to increase user-friendliness with all its benefits while it prevents the design to influence trust in a website. Strategies here might be available on a structural level and on the level of the individual user. On a structural level one approach is to provide means that artificially lower the integration on the trust dimension.[13] The aim here is to counteract the spill from non-epistemic to epistemic dimensions. This can be achieved by providing some sort of psychological defeater to the agent when they visit a website: a consciously available reason that decreases justification with regard to the contents of the website. In fact, in the European Union there is already a version of this approach established – although likely not with this goal in mind. The General Data Protection Regulation forces website providers to make their personalisation via tracking cookies obvious and explicit. Websites have to inform users in the European Union of their tracking mechanisms and users can choose to continue to the website while declining those tracking cookies that are not necessary for the core functioning of the website. By being presented with a pop-up pointing to the tracking cookies, the mechanisms behind the website become more salient to users with enough background information. The necessity of accepting tracking cookies functions as a warning that can decrease trust in a website and thereby compensates some of the effects of user-friendly design. As is, there are still some hurdles for the effectiveness of these warning signs. As long as the owner of a website is in full control of how to include these pop-ups the intended effects could be mitigated. The design of these pop-ups itself might influence their impact on the trust assigned to a website. Companies such as Facebook or Google have the resources to design pop-ups in a way that clicking on them is quick and effortless, compared to other sites. In the worst case, this could lead to sites that warrant higher trust to have badly designed pop-ups that lower trust significantly, but sites that warrant only lower trust to have perfectly engineered pop-ups without much of an effect on assigned trust. To counteract this issue the implementation of such pop-ups ought to be standardised – which perhaps moves the solution back towards the second option discussed.

Moreover, for these pop-ups to have the desired effect they require substantial background knowledge on what they actually indicate. Making personalisation salient does not do the trick if one has no idea about the effects of personalisation. However, this might be supplemented by a strategy on an individual level. The goal thereby is to improve the relevant cognitive abilities of users so that they are able to competently respond to available defeaters by lowering trust put into a website.[14] Heersmink (2018) suggests a version of this strategy with an emphasis on educating for online intellectual virtues, that is, an emphasis on teaching how to apply instances of general intellectual virtues in an online environment based on relevant background knowledge. Part of this educational goal is internet literacy skills, which then in turn allow an agent to apply their general intellectual virtues properly in the online environment. It might be a long shot to train us to not be victims to the psychological effects of processing fluency, but it is less of a long shot to teach us all we need to use institutionally mandated prompts as a way of making defeaters salient. Perhaps, it is even possible to acquire online intellectual virtues that by themselves decrease the default trust for websites, such that not even a salient defeater is necessary to compensate for fluency effects via user-friendly design. The challenge is to find concrete ways of teaching these intellectual virtues. Kotsonis (2020) argues that teaching for intellectual virtues in a social media environment is possible. Similarly, Heersmink (2018) remains hopeful that we can teach online intellectual virtues properly. However, the details of how such an education towards online intellectual virtues exactly looks like are still up in the air, which leaves plenty of work for future research.

Notes

1. An earlier version of this chapter was discussed in the "Manipulation Online" workshop series organised by Fleur Jongepier and Michael Klenk, a research meeting organised by Andreas Müller and a seminar at the University of Duisburg-Essen. Thank you to all participants. Further thanks to the editors of this volume for helpful suggestions.
2. This is the rough definition of "user friendly design" that I work with. User-friendly design has to be kept distinct from *persuasive design*. Persuasive design uses psychological and social means to change user behaviour (cf. Fogg 2009a, 2009b). In contrast, user-friendly design is solely focused on making it as easy as possible for the user to perform a task. It is not aimed at changing the task the user wants to perform. Some design choices that aim at speed and ease of use can also influence the tasks intended. Autocomplete features might fall into this category in a dangerous way (Noble 2018). I will bracket this issue.
3. Although the argument applies to some online systems other than websites I will limit myself to websites.
4. Perhaps not all of the trust is unwarranted, because sometimes aspects that play a role in user-friendly design and cognitive integration are also indicators for the care put into a website and hence plausibly play a role in justifying beliefs formed in relation to a website. For instance, correct spelling is no direct warrant for a claim but might be an indicator for care put into a website and provide higher-order warrant (Tollefsen 2009). However, I argue that at least some amount of trust lacks an epistemic ground because it is based on the effort required to engage with the website and not on any feature indicating truth-conduciveness.
5. The argument from cognitive integration is not the only one available. A similar argument can be provided if we take websites to be instances of testimony. Bracketing issues of who the trust would be directed at the argument would have roughly the following steps:

    1. Information written on and read of a website constitutes a form of testimony.
    2. Generally, user-friendly design of a website increases trust in the website.
    3. Trust based on user-friendly design is not fully epistemically warranted.

    C: Generally, user-friendly design promotes an overly trusting attitude towards a website as a source of testimony (From 1, 2 and 3).

    Thank you to Eva Schmidt for suggesting this version.
6. The alternative is virtue reliabilism. See Sosa (2007), Greco (2009), and Pritchard (2012).
7. And people, but for simplicity I focus on artefacts here.
8. This is not a complete contrast to Clark's work but rather a contrast to the early formulations of the extended mind thesis. See Wilson and Clark (2009).
9. Heersmink himself is not committed to this distinction.

10. For a discussion of similar non-obvious indicators for truth-conduciveness in websites, see Tollefsen (2009).
11. Again, there might be relations in some cases. However, in the following empirical examples it will be clear that we often trust because of speed and ease of processing information that has absolutely no relation to truth.
12. This does not rule out that lack of fluency also leads to a lack of trust. Perhaps, there is a particular point of fluency that helps us to neither overly or underly trust a source. I bracket this issue. All I require is that high fluency leads to an overly trusting attitude.
13. The approach can be broadly qualified as a version of what Lewandowsky, Ecker, and Cook (2017) label "technocognition". See also Kozyreva, Lewandowsky and Hertwig (2020).
14. This approach qualifies as a cognitive "boosting" strategy (cf. Hertwig and Grüne-Yanoff 2017; Kozyreva, Lewandowsky and Hertwig 2020).

7 References

Adams, Fred, and Ken Aizawa. 2010. "Defending the Bounds of Cognition." In Menary 2010, 67–80.

Alter, Adam L., and Daniel M. Oppenheimer. 2009. "Uniting the Tribes of Fluency to Form a Metacognitive Nation." *Personality and Social Psychology Review* 13 (3): 219–35. doi:10.1177/1088868309341564.

Baehr, J. 2015. "Cultivating Good Minds. Retrieved from: A Philosophical and Practical Guide to Educating for Intellectual Virtues." https://intellectualvirtues.org/why-should-we-educate-for-intellectual-virtues-2–2/.

Baier, Annette. 1986. "Trust and Antitrust." *Ethics* 96: 231–60.

Battaly, H. 2008. "Virtue Epistemology." *Philosophy Compass* 3 (4): 639–63.

Begg, I. M., Anas, A., and S. Farinacci. 1992. "Dissociation of Processes in Belief: Source Recollection, Statement Familiarity, and the Illusion of Truth." *Journal of Experimental Psychology* 121 (4): 446–58.

Cassam, Quassim. 2019. *Vices of the Mind: From the Intellectual to the Political*. Oxford: Oxford University Press.

Chatterjee, Samir, and Parvati Dev, eds. 2009. *Proceedings of the 4th International Conference on Persuasive Technology – Persuasive'09*. New York, NY: ACM Press.

Clark, Andy. 2008. *Supersizing the Mind: Embodiment, Action, and Cognitive Extension*. New York, NY: Oxford University Press.

Clark, Andy. 2010. "Memento's Revenge: The Extended Mind." In Menary 2010, 43–66.

Clark, Andy, and David Chalmers. 1998. "The Extended Mind." *Analysis* 58 (1): 7–19.

Fogg, B. J. 2009a. "A Behavior Model for Persuasive Design." In Chatterjee and Dev 2009.

Fogg, B. J. 2009b. "Creating Persuasive Technologies." In Chatterjee and Dev 2009.

Gertler, Brie. 2007. "Overextending the mind?" In Gertler, & Shapiro 2001, 192–206.

Greco, John. 2009. "Knowledge and Success from Ability." *Philosophical Studies* 142 (1): 17–26.

Hasher, L., Goldstein, D., and T. Toppino. 1977. “Frequency and the Conference of Referential Validity.” *Journal of Verbal Learning and Verbal Behavior* 16 (1): 107–12.

Hawley, Katherine. 2019. *How to be Trustworthy*. Oxford: Oxford University Press.

Heersmink, Richard. 2015. “Dimensions of Integration in Embedded and Extended Cognitive Systems.” *Phenomenology and the Cognitive Sciences* 13 (3): 577–98.

Heersmink, Richard. 2018. “A Virtue Epistemology of the Internet: Search Engines, Intellectual Virtues and Education.” *Social Epistemology* 32 (1): 1–12. doi:10.1080/02691728.2017.1383530.

Hertwig, Ralph, and Till Grüne-Yanoff. 2017. “Nudging and Boosting: Steering or Empowering Good Decisions.” *Perspectives on Psychological Science* 12 (6): 973–86. doi:10.1177/1745691617702496.

Hutchins, Edwin. 1995. *Cognition in the Wild*. Cambridge, MA: MIT Press.

Kirsh, David, and Paul Maglio. 1994. “On Distinguishing Epistemic from Pragmatic Action.” *Cognitive Science* 18 (4): 513–50.

Kotsonis, Alkis. 2020. “Social Media as Inadvertent Educators.” *Journal of Moral Education*, 1–14. doi:10.1080/03057240.2020.1838267.

Kozyreva, Anastasia, Stephan Lewandowsky, and Ralph Hertwig. 2020. “Citizens Versus the Internet: Confronting Digital Challenges with Cognitive Tools.” *Psychological Science in the Public Interest: A Journal of the American Psychological Society* 21 (3): 103–56. doi:10.1177/1529100620946707.

Lewandowsky, S., U. K. Ecker, and J. Cook. 2017. “Beyond Misinformation: Understanding and Coping with the ‘Post-Truth’ Era.” *Journal of Applied Research in Memory and Cognition* 6 (4): 353–69.

McGlone, M. S., and J. Tofighbakhsh. 2000. “Birds of a Feather Flock Conjointly (?): Rhyme as Reason in Aphorisms.” *Psychological Science* 11 (5): 424–28. doi:10.1111/1467-9280.00282.

Menary, Richard, ed. 2010. *The Extended Mind*. Cambridge, MA: MIT Press.

Nguyen, C. T. forthcoming. “How Twitter Gamifies Communication.” In *Applied Epistemology*, edited by Jennifer Lackey. Oxford: Oxford University Press.

Noble, S. U. 2018. *Algorithms of Oppression*. New York, NY: New York University Press.

Oppenheimer, Daniel M. 2006. “Consequences of Erudite Vernacular Utilized Irrespective of Necessity: Problems with Using Long Words Needlessly.” *Applied Cognitive Psychology* 20 (2): 139–56.

Pritchard, Duncan. 2012. “Anti-Luck Virtue Epistemology.” *Journal of Philosophy* 109 (3): 247–79.

Reber, R., and N. Schwarz. 1999. “Effects of Perceptual Fluency on Judgments of Truth.” *Consciousness and Cognition* 8 (3): 338–42. doi:10.1006/ccog.1999.0386.

Rupert, Robert D. 2004. “Challenges to the Hypothesis of Extended Cognition.” *Journal of Philosophy* 101 (8): 389–428.

Schwengerer, Lukas. 2021. “Online Intellectual Virtues and the Extended Mind.” *Social Epistemology* 35 (3): 312–22.

Smart, P. 2018. “Emerging Digital Technologies: Implications for Extended Conceptions of Cognition and Knowledge.” In *Extended Epistemology*, edited by J. A. Carter, Andy Clark, Jesper Kallestrup, and Duncan Pritchard, 266–304. Oxford: Oxford University Press.

Sosa, Ernest. 2007. *A Virtue Epistemology: Apt Belief and Reflective Knowledge*. Oxford: Oxford University Press.

Sterelny, Kim. 2004. "Externalism, Epistemic Artefacts and the Extended Mind." In *The Externalist Challenge*, edited by Richard Schantz, 239–54. Berlin: De Gruyter.

Sunstein, Cass R. 2016. "Fifty Shades of Manipulation." *Journal of Marketing Behavior* 1 (3–4): 214–44. doi:10.1561/107.00000014.

Sutton, J. 2006. "Distributed Cognitions: Domains and Dimensions." *Pragmatics and Cognition* 14 (2): 235–47.

Sutton, J. 2010. "Exograms and Interdisciplinarity: History, the Extended Mind and the Civilizing Process." In Menary 2010, 189–225.

Tollefsen, Deborah P. 2009. "'Wikipedia' and the Epistemology of Testimony." *Episteme* 6 (1): 8–24.

Wilson, R. A., & Clark, A. 2009. "How to situate cognition: Letting nature take its course." In M. Aydede, & P. Robbins (Eds.), *The Cambridge Handbook of Situated Cognition* (pp. 55–77). New York: Cambridge University Press.

Zagzebski, Linda T. 1996. *Virtues of the Mind: An Inquiry into the Nature of Virtue and the Ethical Foundations of Knowledge*. Cambridge: Cambridge University Press.