8-2022

When Knowing Prevents Doing: An Exposition on Commitment Hazards

Alexander Porter
alexporter@unomaha.edu

Follow this and additional works at: https://digitalcommons.unomaha.edu/university_honors_program

Part of the Metaphysics Commons, Other Philosophy Commons, and the Psychology Commons

Recommended Citation
https://digitalcommons.unomaha.edu/university_honors_program/187
Abstract

Information hazards are risks posed by potentially harmful true information. Information hazards include the risks posed by instructions on how to make a bomb, facts about world events which could cause harmful political unrest, or even the password to an email account being revealed. I will examine and explain one specific type of information hazard which can be seen as the subject of discussion in three well-known but disparate philosophical texts: information which draws attention away from individual control. I will argue that this idea shows up in Plato as what is being avoided through noble falsehoods in *The Republic*, in Nietzsche’s concept of slave morality in *The Genealogy of Morals*, and in Lacan’s account of desire as a primary motivator of action, as well as something we can only have if we do not know what we want. In each of these examples, greater access to information leads people to believe they lack control of their lives. This is a problem because modern psychological research indicates that a belief that you don’t have control over your life makes people less happy and less capable of self-control.
Introduction

The aphorism ‘knowledge is power’ has been widely accepted into our popular culture, but it doesn’t seem to be true. As a matter of fact, there are often cases where knowledge makes an individual weaker, causes harm, or prevents people from being able to act. One underappreciated work of contemporary philosophy which covers exactly this is Nick Bolstom’s Information Hazards, a Typology of Potential Harms From Knowledge. In it, Bolstom catalogs different ways in which certain kinds of information could cause harm to individuals, depending on both the kind of harm caused and the kind of information causing the harm. He refers to these potential kinds of harms information can cause as “information hazards”.

One kind of information hazard that is particularly relevant both in contemporary life and throughout many classic texts of philosophy and psychology is the ‘commitment hazard’. A commitment hazard is a specific type of information hazard which prevents an individual from being capable of committing to a course of action which would be beneficial. We can imagine smaller scale instances of commitment hazards, such as an individual choosing to skip a workout because they learned that the risk posed by skipping a single workout is small. Smaller hazards like this should still be avoided with great care, because, to continue the example, even missing a single workout aids in harming your commitment to the larger lifestyle changes working out more usually entails.

In addition, there can be much larger scale commitment hazards. Take an existential nihilist as an example. An existential nihilist is someone who believes (possibly correctly) that life has no intrinsic meaning. If existential nihilists are correct, it
would still be the case that acquiring the knowledge that their view is correct would pose a hazard, as the person who holds that view might be likely to take actions that could make their life less happy and fulfilling.

I begin with an analysis of Bolstom’s typology of information hazards - particularly commitment hazards - and then illustrate instances in classic works of philosophy and psychoanalysis in which the dangers of commitment hazards are discussed (albeit with different terminology). It’s especially relevant to discuss examples of commitment hazards not only because they might be well known, but also to show that the problem of commitment hazards goes back very far, but does not have any good solutions. Next, I will explain a few contemporary psychological explanations for commitment hazards. Finally, I will defend the view that there is some information you should deliberately hide from yourself or others as a better way of living life.

**Information Hazards**

Information hazards are potential risks which come from certain pieces of information. These pieces of information can be something concrete, like a sheet of instructions for building a bomb, or they can be something more abstract, like information which encourages acting selfishly to the detriment of others. Bolstom lays out two different typologies for information hazards: one to describe the hazard’s mode of information transference and one to categorize a hazard’s effect. The specific taxonomy of information hazards he uses is less important to what I am focusing on in this paper than the idea of information hazards more generally.

It should be noted that an information hazard remains a hazard even if it ends up having no real world harm. An information hazard also remains a hazard if it has a
potential to harm others but not the person who sees the information. Instructions to build a bomb remain a hazard to peoples’ wellbeing even if the bomb described in those instructions is never built, and even if it doesn’t harm the person who could make it. For more concrete information hazards this is a lot easier to see than for more abstract information hazards. For the purposes of my discussion, it is important for me to make my own distinction between types of information hazards.

Concrete information hazards are pieces of information that pose a threat to (or enable someone to do something that poses a threat to) individuals’ physical wellbeing. The previous example of instructions to build a bomb would be an example, as would other instructions on how to build or do things that hurt people physically. Abstract information hazards are pieces of information that pose a threat to (or enable someone to do something that poses a threat to) an individual’s mental wellbeing. There are some things people could more generally agree to being abstract information hazards, like certain kinds of verbal or emotional abuse or learning about a physical ailment which you cannot do anything about. Other things might be more contested as abstract information hazards, even including philosophical concepts (such as determinism and, in my view, nominalism, and physicalism). These more contested types of abstract information hazards seem to be more contested not because the harm they may cause is less real, but because the harm they may cause is less immediate. Later in this paper I will be defending each of the given examples as being information hazards.

Before that discussion, it might be worth asking how context sensitive an abstract information hazard is. We could imagine a theoretical being who experiences a degree of harm from any piece of information they are exposed to. This is a reasonable
question to be asked about what an abstract information hazard is, but it’s outside of the scope of what I am interested in for this paper. Further discussion and clarification on what an abstract information hazard is might be important work. For my purposes, we can simply say an abstract information hazard is a piece of information which would cause some degree of emotional or mental harm to most normal people under most normal circumstances. In this way, abstract information hazards are culturally contextual to some degree. I wanted to delineate between abstract information hazards and concrete information hazards specifically because the issue of context sensitivity does not seem to be present for concrete information hazards.

Commitment hazards can be seen as a type of abstract information hazard under my taxonomy, and they are the type of information hazard the remainder of this paper will be focusing on. Bolstom’s paper on information hazards was the first time anyone talked about them with that terminology, but the general concept of information that harms people (especially by causing inaction) has been present throughout the entirety of the history of philosophy. To investigate this tradition further, let us begin looking at some examples.

**The Republic**

*The Republic* is the most significant of the Socratic dialogues and is arguably the most important work of philosophy. *The Republic* follows a conversation, primarily between Socrates and Glaucon, about what an ideal society should look like. In this exposition I am primarily concerned with Book 3 of *The Republic*. In Book 3 Socrates explains how the rulers of the ideal society would be raised. This includes a great deal of exposition about the kinds of stories that should be told to children (Plato, 415a), the
most important of which involves a type of social stratification. Socrates says we ought
to tell children that people have souls that are made of different types of materials.
Members of the ruling class have gold mixed into their souls, and the lower classes
have souls mixed with silver, iron, or bronze. It should be noted that the metal of a
person’s soul is not necessarily the same as their parents, but it is rare that someone
will have a soul of a different class than their parents (415b).

On this specific bit of fabricated mythology, Glaucon asks Plato, “So, do you
have any device that will make our citizens believe this story?” (415c). Plato responds,
“I can’t see any way to make them believe it themselves, but perhaps there is one in the
case of their sons and later generations and all the other people who come after them”
(415d). Let’s put aside for the moment whether the story would effectively convince
anyone and look only at what effect Plato thought it would have. Socrates says
specifically that he imagines it would “help to make them care more for the city and
each other” (415d), and the implication of the rest of the book is that he believes it would
have a positive impact on the individuals, as well as the groups, as people would
believe they each have a particular nature to work towards. It is hard to imagine that
people would be less satisfied with their work if they believed that the nature of the work
they were doing is derived from how their very soul is.

Socrates calls this type of story a “noble falsehood” (Plato 414b), specifically
because it’s a lie which would have a positive effect. If people were instead to believe
that people do not have classes that are inherent to their very natures, it would be much
more difficult to successfully run an authoritarian society. Putting aside whether or not
an authoritarian society could ever be justified, holding certain beliefs under certain
circumstances allows individuals to be more happy and comfortable more easily within certain societies.

If you were to live in Plato’s ideal society and believed that people did not have souls which assign them to particular classes, it might make it more difficult to function within that society. If people are assigned to social classes on an arbitrary basis rather than there being a very good natural reason for the assignment of social classes, you might want to take actions against the system, which typically doesn’t work well for the individuals who do it. Even if you were not to take some kind of revolutionary actions, having the knowledge that their class system is made up might make your day to day life slightly more difficult. You might grow resentful of the people who are in the upper class, or not want to do the work you have been assigned because you know it is arbitrary, or many other reasons holding this belief which conflicts with the zeitgeist might result in causing you individual harm.

This is a key example of a commitment hazard. An individual obtains the true information that souls are not imbued with metals that assign them a class, which in turn makes it more difficult to commit to the actions they would have otherwise chosen. If an individual simply believed the lie that souls are imbued with metals that determine class, they would be able to more easily commit to the actions which are most beneficial to them within the constraints of that particular society.

The Genealogy of Morals

Nietzsche’s *The Genealogy of Morals* provides our second example of commitment hazards. In the first treatise, Nietzsche outlines a history of morality in which two completely different ideas of morality have emerged. The first moral system is
one which dichotomizes good and bad, with good things being strong, having good qualities towards some end, or traditionally associated with nobility. Bad things, in this dichotomy, are things which are weak, low quality, or which could be associated with commoners. He refers to this moral system elsewhere as “Master-Morality” (Nietzsche, *Beyond Good And Evil*, 271). The alternative moral system is one which dichotomizes good and evil. This moral system relies on the same values as master morality, but everything master morality deems as good it deems as evil, and everything master morality deems as bad it deems as good. He refers to this moral system as “Slave-Morality” (Nietzsche, *Beyond Good And Evil*, 271).

Nietzsche seemed to believe that we ought to adopt a master morality rather than a slave morality, but whether or not he believed this is aside from the point. If an individual were to be exclusively concerned with their own well being, it’s almost certain they would have a master morality rather than a slave morality. A slave morality is beneficial for people who are of a radical underclass and are unable to take power for themselves for a few reasons (it’s useful for coping, it might make coalition building easier), but it does not seem to be beneficial for most people in most normal circumstances in the modern world. It seems at least plausible that the Machiavellian traits implied by master morality would make it easier for an individual to act on their desires, do the things they want to do, and self-actualize more easily.

**The Four Fundamental Concept of Psychoanalysis**

Jacques Lacan’s seminar, *The Four Fundamental Concepts of Psychoanalysis* gives us our third and clearest example of commitment hazards. One of the titular four fundamental concepts of psychoanalysis is Drive. Drive is one of the primary motivating
factors in humans. It is completely distinct from our biological motivations, and it exists entirely, according to Lacan, as a psychological type of motivation. Drive gives people motivation through satisfaction within dissatisfaction. Satisfaction is derived from getting something that psychologically represents some larger thing which is missing, and the dissatisfaction is from not having the (symbolic) thing that is still missing when the real object that represents it is obtained. Love would be a good example of how Lacan conceives of the drive. There is a satisfaction felt when an individual loves another individual, and oftentimes even more satisfaction when that individual’s love is reciprocated. According to Lacan, loving another person is really satisfying because there’s a dissatisfaction in having a specific instance of love, but not having the symbolic thing that love is in an abstract sense. For Lacan love is one of the most powerful types of Drive specifically because it is a nonsensical concept (Lacan 61)[1].

Within Lacan’s framework, love is still something an individual can have in a real sense[2]. Because of this, an individual might eventually attain this real love. When they do this, the opposite of Drive’s motivation would happen (i.e, dissatisfaction in satisfaction rather than satisfaction in dissatisfaction). Because there would be nothing left to uncover, no greater thing to be obtained that love is pointing to, it would become unmotivating and unsatisfying.

It is because, to Lacan, love is not a sensical concept that it is such a powerful motivating force. When we have motivating forces that are based on real and immediate physical desires, like hunger, we take a single action that will fix this hunger. Because love is not real in any meaningful sense, it is capable of motivating a great deal of actions, because no single action can satisfy a desire that doesn’t point to anything real.
I take Lacan with a grain of salt, but there is certainly some truth in this idea. If an individual knows what they want\textsuperscript{[3]}, then they will get what they want, move on, and be done with it. If an individual wants something that’s representative of a greater whole but does not know what that greater whole is, that’s a type of motivation that could last them forever. In this way, we could see Lacan’s view as being that knowing what you want is a commitment hazard. What you really want is a piece of information which, if obtained, would result in your loss of motivation.

All three of the given examples are controversial in one way or another, and I don’t mean to say that I agree with any of them on the particulars of their ideas. What they all provide is an example, at least within their view, of how acquiring some form of knowledge might make us less capable of commitment.

**Locus of Control**

A therapeutic model can be used to understand the nature of commitment hazards. Locus of control is a concept in psychology that is especially prominent in cognitive behavioral therapy. Locus of control is the degree to which people believe their lives are under their own control as opposed to being controlled by external forces. An individual with an internal locus of control is someone who believes they have a great deal of control over the events in their life. In contrast, an individual with an external locus of control is someone who believes they do not have a great deal of control over the events in their life.

While there are situations in which an external locus of control may be beneficial, this is rarely the case. In lifestyle choices, individuals with an internal locus of control tend to have better long term health related behaviors (Norman et al.). Individuals with
an internal locus of control end up in a higher socioeconomic status (Cohen et al.) (although it is possible the causal direction could go either way). Students with an internal locus of control perform better academically (Vansteenkiste) and when students with an external locus of control receive counseling designed to make their locus of control internal they begin performing better academically (Whyte).

We can generalize this outward. It seems as though we have a good basis to say that an internal locus of control generally allows individuals to live better lives than an external locus of control. As such, it might be reasonable to consider information that makes individuals believe they have little control over their lives is hazardous. This appears to be the primary driver of commitment hazards. When an individual learns of a commitment hazard, it causes them to believe that they have less control over their life, and that belief makes them less likely to act. This is only intuitive: when an individual learns something which makes them believe their actions are pointless, they are less likely to act. The basis on which a difference in locus of control can make individuals less likely to act can be our foundation for understanding how commitment hazards cause people to avoid acting.

Locus of control has a great deal of correlation with similar beliefs, such as self-efficacy and expectancy shifts (Tyler et al), that is, an individual with a central locus of control is more likely to be more self-efficacious and to have a regular expectancy shift (i.e, to believe that when one thing happens after another thing once that those things are likely to be correlated). These shifts in personality seem to come from a difference in gray matter volume in several areas of the brain, specifically the anterior cingulate cortex, striatum, and anterior insula (Hashimoto et al), but of course we can’t say if
those differences in the brain cause certain beliefs or if the beliefs cause differences in
the brain[4].

Shifting locus of control - or whatever personality trait locus of control, self
efficacy, and expectancy shifts may be tied to - might be the primary way in which
commitment hazards work. If nothing else, research on locus of control gives us a good
way of understanding commitment hazards through a concept which has already been
studied thoroughly.

A prominent idea in popular culture is how people might live their lives differently
if they knew the exact date, time, and method of their own death. It is reasonable to
believe that if we knew these things some people would live their lives differently, but
why? If we know when our death will take place and knowing doesn’t impact our ability
to change it happening, it shouldn’t have any concrete impacts on our actions. This is
the nature of most abstract commitment hazards: they change our characteristics in
ways that have a negative impact in the long run. If we knew exactly when we would
die, the three previously mentioned traits would be negatively impacted, but it would not
enable us to take better actions in the world. Many philosophical beliefs have the
potential to harm people in similar ways.

Commitment Hazards in Philosophy

Let us first look at determinism, both because it is one of the most discussed
philosophical issues and one of the philosophical views with the most research behind it
in terms of psychology and neuroscience. Determinism seems, from my view, to be a
potential[5] commitment hazard. It seems intuitive that believing your actions - or
everything - is predetermined might have negative impacts on your life. This has been
backed up in a multitude of studies, such as one which showed that belief in
determinism predicts life satisfaction and positive affect (Li et al), willingness to cheat on
tests (Schooler), and job performance (Stillman et al). In all cases determinism is
associated with what we would typically consider to be worse outcomes (lower life
satisfaction, less positive affect, higher willingness to cheat, and lower job
performance).

While it seems clear that belief in determinism plays some kind of a causal role in
these negative effects, it is not the lone culprit. Determinism is a view which is a
consequence of larger metaphysical views which entail determinism. One example of
such views would be physicalism, the view that everything which exists in the universe
is reducible to physical terms. Determinism typically follows from a physicalist world
view. If everything in the world is reducible to physical terms and all physical things are
causally related then determinism is true.

Another view that is a close associate with those two (although not directly
related) is nominalism. Nominalism is the view that metaphysical categories and
abstract objects are not real. Whereas a metaphysical realist might say that different
things that are green all have a shared property of green which exists abstractly, a
nominalist would say those objects all merely have similar wavelengths of light and we
have arbitrarily grouped them into being ‘green’[6]. Nominalism might not be as harmful
as other beliefs, like determinism, but it seems plausible that there are harmful aspects
to it. For the nominalist categories are arbitrary in some sense. If the distinction between
things in our universe do not stem from any real system of how things are divided, the
(perceived) arbitrary nature of everything could potentially wear on us. An individual
who believes that category distinctions are real might feel more connected to the physical world, and might be more capable of taking action in it as a result.

More generally, and less empirically defensible, it seems that there is a harm in holding beliefs that reduce the experience of life to some other terms. The way people interact with and understand the world comes largely down to a matter of heuristics, feelings about the way things are, and rough approximations of real phenomena. More complicated schema of how the world works, even if accurate, might take away from this sense we have about how the world is, and make it something we want to interact with less.

We can analogize it to physics. Newtonian mechanics accurately describes how most of the physics we do will work, but it simply turns out incorrect results in some situations that we deal with less (at extremely high speeds, for example). This does not mean that Newtonian mechanics are not useful: Newtonian mechanics put a man on the moon. On the other hand, if we were dealing with all of the complexity that more modern systems of physics entail, like quantum mechanics, we might not have been able to put a man on the moon. The complexity of quantum mechanics could have bogged us down in the details so much that taking the actions that were sufficient to work could not have happened.

Our philosophical views seem to me to work like that. Being bogged down in questions like whether or not we have free will could create some kind of hesitancy or unwillingness to act that harms us. If determinism is true, or if any other belief is true for that matter, then it will likely be relevant on some level. In order to answer questions about the best way for people to live and the best way for societies to be structured,
ultimately we need to know the truth of the matter. For individuals trying to live the best versions of their lives however, a simpler and less accurate world view might have better results.

In general, the set of beliefs including determinism, physicalism, nominalism, and others is growing in popularity[7]. I believe this is a bad thing because these beliefs seem to cause a great deal of harm to the people who believe them, insofar as they might make individuals lives worse by disrupting their sense of purpose, meaning, or ability to act effectively in the world.

**In Defense of Knowing Less**

In this final section I will argue for the view that some individuals should not do philosophy. This is not to say that everyone should not do philosophy, and I think most people should, but there are some people who do not live in such a way that they will gain more from doing philosophy than they stand to lose from doing philosophy.

For most individuals, there is no reason to know the full line of argumentation for and against determinism. Furthermore, some risk is incurred that they could come to believe that determinism is true if they were to learn what the lines of argumentation look like. This risk of believing some things which could be harmful to the individual who believes them is incurred whenever individuals do any kind of philosophy, especially metaphysics and ethics. For this reason, it seems as though at least some individuals have an obligation to not do philosophy. For someone whose life is largely plotted out for them and will not involve a lot of big decisions, they don’t stand to gain a lot out of doing philosophy, insofar as it destabilizes the beliefs orienting their values and sense of purpose[8]. For such persons, doing philosophy might instead destabilize their view of
reality, and potentially cause them to become less happy and motivated, and more likely to do bad things like cheating.

This view is not a way of disrespecting the philosophical tradition, but rather it is a way of defending it. As demonstrated in the earlier sections of this paper, some of the greatest philosophers in history defended that, at least in some specific cases, there are potentially true things which we should not know because knowing those things will do more harm than good. Understanding the people and situations who should not be dealing with certain philosophical problems seems to be one of the more important questions in philosophy, and one which has gone critically underexplored.

Notes

1. Note well that while Lacan believes love is an incoherent concept, it is one of the things he wrote about the most in his body of work. This is not contradictory, he did psychoanalysis, which is chiefly concerned with the unconscious. It makes a lot of sense he would be so interested in people’s use of, in his view, a nonsensical term.

2. To clear up any confusion, Lacan’s view is that love is a type of narcissistic relationship one has with themself. To love someone is to imagine being loved by them, which is both a way of loving yourself through a symbolic object (the other person), and trying to attain wholeness (through combination of a subject and an object).
3. Note that when I use the word ‘want’ throughout this paper I mean it in a colloquial sense and not the psychoanalytic sense that Lacan and others use the word.

4. Given that things like the placebo effect can have real physical changes in the brain (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6013051/), it seems reasonable to believe that an individual’s beliefs with regards to things like locus of control can change their brain in physical ways (although I do not know of any research that has been done on this subject).

5. In Bolstom’s paper he states that an information hazard is specifically a “true” (Bolstom 1) piece of information that is potentially harmful, so I have opted to refer to these views as “potential information hazards”, as they are potentially true.

6. This is merely a simple example and I would not defend colors as being real abstract categories which exist.

7. Thomas Nagel refers to this commonly shared group of beliefs as the “Materialist Neo-Darwinian Conception of Nature” (Nagel), which is a name I like, although it’s a bit of a mouthful. I think this set of beliefs can be accurately described as falling under ‘scientism’ for the most part, and that the word should be used to refer to this group of beliefs.

8. We might look at the example of someone who is going to inherit their parents’ business. If the business is anything within the realm of things that are probably good or harmless, this person is almost certainly going to spend their life on their parents business, and if they are not then the decision not to do so was probably
made well before they did any philosophizing about whether or not they should run their parents business. It does, however, have the potential to give them doubts about what they are doing.
Works Cited


Schooler Jonathan W, Vohs Kathleen D, “The Value of Believing in Free Will: Encouraging a Belief