epistemе

ARTICLE

Information Hazards as Activity and Content: A Grounded Account of Dis/Misinformation

Omar El Mawas[1,2]

[1]Centre Cavaillès, Ecole Normale Supérieure, Paris, France and [2]Department of Philosophy, Durham University, Durham, UK
Email: omar.el-mawas@durham.ac.uk

Abstract

The study of dis/misinformation is currently in vogue, however with much ambiguity about what the problem precisely is, and much confusion about the key concepts that are brought to bear on this problem. My aim of this paper is twofold. First, I will attempt to precisify the (dis/mis)information problem, roughly construing it as anything that undermines the "epistemic aim of information." Second, I will use this precisification to provide a new grounded account of dis/misinformation. To achieve the latter, I will critically engage with three of the more popular accounts of dis/misinformation which are (a) harm-based, (b) misleading-based, and (c) ignorance-based accounts. Each engagement will lead to further refinement of these key concepts, ultimately paving the way for my own account. Finally, I offer my own information hazard-based account, which distinguishes between misinformation as content, misinformation as activity, and disinformation as activity. By introducing this distinction between content and activity, it will be shown that my account is erected on firmer conceptual/ontological grounds, overcoming many of the difficulties that have plagued previous accounts, especially the problem of the proper place of intentionality in understanding dis/misinformation. This promises to add clarity to dis/misinformation research and to prove more useful in practice.

1. Introduction

The study of dis/misinformation is currently in vogue, however with much ambiguity about what the problem precisely is, and much confusion about the key concepts that are brought to bear on this problem. I have two direct aims in this paper, as well as other less direct ones. Let us start with the direct ones. First, I will attempt to precisify the (dis/mis) information problem, roughly construing it as anything that undermines the "epistemic aim of information" – I will explain what that means in due course. Second, I will use this precisification to provide a new grounded account of dis/misinformation. The latter will take place through critically engaging with three of the more popular accounts of dis/ misinformation which are: (a) harm-based (b) misleading-based, and (c) ignorance-based accounts. Each engagement will lead to further refinement of these key concepts, ultimately paving the way for my own information hazard-based account.

As for the indirect aims, they are methodological but also therapeutic in nature. That is because the field of dis/misinformation studies has emerged largely due to practical concerns encapsulated by the urgent need to "combat" dis/misinformation. This urgency has led most scholars to jump right into trying to provide practical fixes which lack proper theoretical grounding. This paper attempts to provide one such grounding by connecting definitions with aims, as well as practical concerns.

Underlying the current classifications of dis/misinformation in the literature is the confusion regarding the proper place of intentionality in these classifications – call this the "placement problem[1]." There are likely moral as well as legal stakes in accusing an entity of intentionally spreading dis/misinformation; hence, diligence is required before describing such actions as intentional – a duty of diligence. But there is also an urgent need to counter dis/misinformation so as to prevent the erosion of truth from our information environment – a duty to act. These considerations pull in opposite directions.

My hazard-based account introduces the distinction between *activity* and *content*, which is absent in the current literature, leading to three key concepts which are *misinformation as content*, *misinformation as activity*, and *disinformation as activity*, where activities but not content involve intentions.[2]

This conceptual distinction not only solves the existing placement problem but also helps us to rethink the dis/misinformation problem itself – or simply the "information problem," which can now be broken down into two distinct but related problems, one pertaining to content – *the content problem*, and the other to activity – *the activity problem*. Briefly, the former concerns misinformation after it has been communicated, while the latter concerns its very act of communication.

While both information problems are deserving of our attention, I note that almost all scholars have been mainly concerned with the content problem, and accordingly I will do the same – this leaves addressing the activity problem for another occasion.

The account I offer helps to ease this tension by decoupling content from intention in our concept of misinformation in a way that allows us to address the content problem without having to worry about intentions. This does not entirely eliminate the concerns regarding intentions as, on my account, intentions remain constitutive parts of both disinformation and misinformation activities. Nonetheless, they become less of a problem in addressing the content problem of information as compared to other accounts on the market, and that itself constitutes progress.

In the final analysis, it will be shown that my account is erected on firmer conceptual/ontological grounds, overcoming many of the difficulties that have plagued previous accounts, especially the almost entire absence of theoretical grounding, the tension between diligence and action as well as the placement problem. This promises to add clarity to dis/misinformation research and to prove more useful in practice.

2. Precisifying the information problem

Enquiry can have diverse aims, including epistemic, psychological, and social ones. Its epistemic aims include knowledge, justified true belief, truth, etc. Communication can be a form of enquiry. Information as an activity is a form of communication and as such can have any of its said aims. Information as content is intrinsically epistemic, that

---

[1]The placement problem, as will be clear, is really a pseudo-problem, symptomatic of the mistaken mixing of content and activity in the current classifications.

[2]As will be clear, I also suggest a place for disinformation as content, understood as subspecies of misinformation as content. But it is not as central.

is having truth as a constitutive element – what information is and how we can makes sense of expressions such as "false information" will be discussed in due course. Accordingly, we can safely say that a minimal aim of information is truth. That is, we want our communication of information to be at least truthful or something close enough. The reason for the latter qualification is that in a great many cases truth is unavailable and there is an urgent need for epistemically based actions. In such cases, we settle for something close enough such as something with a good-enough degree of evidential support. Sometimes we don't even have that. As such, we aim for something more modest but not less important, namely avoiding and rejecting falsehood. It could then be said that

> *A modest epistemic aim of information as an activity is to seek and accept truth and/ or avoid and reject falsehood.*

This stage setting remark is important because when we speak of "information hazards" such as "information disorder" (Wardle & Derakhshan 2017; Wardle 2018) or "information pollution" (Meel & Vishwakarma 2020), or generically the now popular yet still ambiguous "fake news" (Lazer *et al.* 2018; Bernecker *et al.* 2021), or its seemingly more technical formulations under mis/dis/malinformation (Shu *et al.* 2020; Dame Adjin-Tettey 2022), we had better know what precisely the problem is. Why, precisely, is "fake news" or "disinformation" bad – a question whose answer, for most, perhaps seems too trivial to deserve serious consideration? But I think it is important to address it, especially given the conceptual confusion that currently exists. Having clarified what I take the epistemic aim of information to be, I am able to characterize "Information hazards."

> *An Information hazard is anything that through and including (semantic) content, can undermine the achievement of the epistemic aim of information.*

Here, "anything" makes the characterization quite broad. But it is also limited by the undermining taking place through content. The aim here to include all ways in which the epistemic aim of information can be undermined and not impose a specific stricture on what those can be – as will be further elucidated this allows to include activities as well as content, a distinction that will be crucial for clearing up certain confusions that exist in the current classifications.

I emphasize "through content" because the means by which the epistemic aim of information is undermined need not themselves be information/content-based even if they will have information results. For instance, drugging someone could undermine the epistemic aim of information but it would not count as information hazard as it does not take place through content. However, talking someone into a certain psychological state which affects their judgment-making abilities (e.g. warmongering) takes place through content and hence may in some cases count as an information hazard. This last point will be further explored later.

It is also important to note that I am addressing the problem from the vantage point of the receiver, that is, the individual. Of course, one can also address it from the vantage point of the source or the medium, and at some point we ought to include those in order to have a more comprehensive view. But in this paper, I focus on the receiver because I believe it is especially important given that the ultimate targets or end points of dis/ misinformation – depending on whether or not it includes intentionality, are specific *judgments* of particular individuals on a particular state of affairs. But even if the target

or end point is taken to be not judgments but behaviors/actions,[3] I would still maintain that so long as people are acting for reasons, dis/misinformation will involve judgments as a penultimate target (Alvarez 2016).

Attempts have been made to address the problem of information hazards from different standpoints, for example, ethical (Freiling *et al.* 2023), political (Levy *et al.* 2021), and psychological (Roozenbeek et al. 2022), each with their own definitions and classifications. But as noted by a recent review, there is a "remarkable amount of disagreement" over the classifications of key terms in the literature (Aïmeur *et al.* 2023). This has led to much conceptual confusion, leading scholars such as Tim Hayward (2023) to complain that while approaching the literature we find ourselves confronted with

> a multifarious set of different framings and assumptions, all ostensibly talking about the same general problem under different descriptions, without sufficient reflection on how the descriptions may have very different referents (2).

There are many problems with this multifariousness. First, it is unprincipled. Second, it is confusing, since key terms are employed differently between disciplines as well as within the same discipline. Third, and most importantly for practical purposes, conceptual confusion, if left unremedied, will carry over to inferences and likely affect policy recommendation. I would further point out that this is due to no small extent to the underspecification of what Hayward described as the "same general problem," which we specified above as the problem of "information hazards."

My main concern here is not to remedy all the confusion that exists in literature but to provide some groundwork to that end. In fact, I aim to do something very modest in intent and in philosophical theorizing, namely to criticize three of the more currently popular accounts of dis/misinformation (harm-based, misleading-based, and ignorance-generating-based) with the ultimate goal of providing what I take to be a better, grounded, alternative.

My criticism will be conceptual and ontological. By conceptual, I aim to show the current widely held definitions of dis/misinformation and their updates, are either inoperable, that is, too opaque to be useful, or that they fail to account for what we take to be clear cases of dis/misinformation. By ontological, I aim to show that these classification are not careful enough, subsuming what belongs to one ontological category under another, particularly when they take disinformation to be a species of misinformation.[4] After correcting that I will finally suggest my own account of dis/misinformation.

3. Current accounts

3.1. Harm-based account

Let us begin with one account that has gained much traction in recent years, not only in academic discourse but also gray literature, which is that of dis/mis/malinformation.

---

[3]Behavior usually belongs to the psychology literature, whereas action usually belongs to the philosophy literature. For my purpose, there is no difference between the two.

[4]While on the surface one of my conclusions will be that disinformation is a species of misinformation, this is very different from the mixing that exists in current literature. That is because I distinguish between content and activity, and my conclusion will be that disinformation *as content* is a species of misinformation *as content*.

Table 1. A classification of misinformation and related concepts along the axes of truth and intention to harm.

Truth/intention to harm (ItH)	True	False
ItH	(Mal)information	Disinformation
Not ItH		Misinformation

For instance, it is adopted by *the American's Cyber Defense Agency*,[5] *Institut national de la santé et de la recherche médicale (Inserm)*,[6] *the Canadian Centre for Cyber Security*,[7] and the *Council of Europe*.[8]

According to this widely adopted classification (Wardle & Derakhshan 2017):

- Misinformation is when false information is shared, but no harm is meant.
- Disinformation is when false information is knowingly shared to cause harm.
- Malinformation is when genuine information is shared to cause harm (5).

Some have questioned the coherence of "false information." They hold that information is true by definition and as such "false information" amounts to a contradiction. A notable proponent of that view is Luciano Floridi (2013) according to whom information is data that is well formed, meaningful, and truthful (31). As such, he construes misinformation as "semantic content that is false" where "semantic content" is a shorthand for well-formed, meaningful data (2010, 50). Others have contended that "information" does not require truth by pointing out that computer scientists and cognitive scientists use the term in a way that does not imply truth (Scarantino & Piccinini 2010).

It is not difficult to reconcile both by views by suggesting that "information" can have a loose and a strict sense, and only the strict sense requires truth and it is this strict sense that concern us here. For our purpose, then, one could simply replace "information" in the definitions above with "semantic content" – I sometimes will use "content" for short. This leads to the following:

- Misinformation is false semantic content that is shared but no harm is meant.
- Disinformation is false semantic content that is shared to cause harm.
- Malinformation is true information shared to cause harm.

It should be clear then that this classification takes place along the two axes of truth and intention-to-harm as shown in Table 1.

---

[5]Mis-, dis-, and malinformation – home page | cisa, accessed November 20, 2023, https://www.cisa.gov/sites/default/files/publications/mdm-incident-response-guide_508.pdf.

[6]"'Fake News' et Désinformation Autour Du Coronavirus SARS-CoV2." n.d. Salle de Presse de L'Inserm. Accessed February 15, 2024. https://presse.inserm.fr/cest-dans-lair/fake-news-et-desinformation-autour-du-coronavirus-sars-cov2/.

[7]Canadian Centre for Cyber Security, "How to Identify Misinformation, Disinformation, and Malinformation (ITSAP.00.300)," Canadian Centre for Cyber Security, February 23, 2022, https://www.cyber.gc.ca/en/guidance/how-identify-misinformation-disinformation-and-malinformation-itsap00300.

[8]"Information Disorder - Freedom of Expression - Www.Coe.Int," Freedom of Expression, accessed November 20, 2023, https://www.coe.int/en/web/freedom-expression/information-disorder#{%2235128725%22:[]}.

Here, we have three (four) key concepts, "truth"/"falsehood," "intention," and "harm." Everyone in the literature agrees that truth is an important ingredient in discussion of dis/misinformation. Although I largely agree with this diagnosis, I think that in many important cases this criterion can be problematic, but discussing that is for another paper. This leaves us with "intention" and "harm." I take issue with both.

Let us start with harm. What exactly do we mean by it? In their report, Wardle & Derakhshan mention the "harm" 21 times, but they do not attempt provide a definition or even a characterizations of it. They do say that harm can be caused to a person, a social group, an organization, or a country (2017, 20). They also provide some examples of where they think dis/misinformation has caused harm. For instance, they mention a disinformation campaign that was meant to damage Emmanuel Macron's chances during the 2017 French Presidential election, including a fraudulent article purporting that he was funded by Saudi Arabia (21).

Here we can ask what was that harm intended, and to whom exactly? One might think that the answer is rather trivial, namely that it is intended to harm the image of Macron. But can we be confident of that? What if the source of disinformation was of right-wing leanings, and whose aim was to support Marine Le Pen by all means necessary, meaning that they spread disinformation with the original intent *to* benefit Le Pen even if that *harms* Macron? In this case the intent to harm Macron, if at all existent, is only secondary to that of benefiting Le Pen. Or what if the source of disinformation thought Macron unfit to act a president and as such spread disinformation with the intent to benefit the French republic?

Take other examples, such as someone who is spreading that Covid is a hoax due to her distrust of the government, or an antivaxxer who knowingly or unknowingly is spreading "false information" because she genuinely thinks that vaccines are hazardous to one's health. Neither of those has any intention to harm, although they certainly have the capacity to harm. The problem with "harm" in the definitions of dis/misinformation, as I see it, is that it cannot be dissociated from the intent of the source – we are thinking of the *intent to* harm. Also, the source may not share our same values, so what we may take as harmful they may not. What makes things even worse is that, from an epistemic perspective, intentions are almost always opaque – call this the problem of intention opacity or "the problem of opacity" for short. We are seldom in a position to assert with confidence what someone's intentions are.

But even if we are confident of someone's intentions – say they explicitly state it, if we go back to the Macron's example. Let's say someone does spread information because she wants to benefit Le Pen, and she strongly believes that the survival of the French republic depends on Le Pen being elected. In this case, Macron is a just a place-holder for whoever is running against Le Pen. As such, the harm done to Macron is strictly accidental.

A plausible response here is to point out that the same actions may have more than one intent, some of which are *internal* to the action of disinformation itself, such as distorting Macron's image, and other *external* ones, such as saving the republic, and that it is the internal aim that is relevant when discussing harm. Although I find it to be a strong response, I still think that it is not entirely convincing because it fails to appreciate that when considering intentions, what matters more is not whether intentions are internal or external to particular actions, rather it is which intention has precedence in the mind of the actor. So even if we grant that the intent to distort Macron's image is internal to the action of disinformation, the intent to, say, benefit Le Pen is not an external one. It, in fact, takes precedence in the mind of the actor over the internal intention of the action and, in a way, conditions it. That is why I would still maintain that, the objection notwithstanding, distorting Macron's image may be made with no harmful intent even if *we know* it will most likely have caused harm.

One may try to improve the harm-based account by restricting the kind of harm involved in dis/misinformation to the epistemic kind, as opposed to, say, the physical, social or otherwise. This is certainly better than simple harm-based accounts, and it seems to be the (epistemic) direction in which later (misleading and ignorance generation) accounts are going. However, it still has the same drawbacks as other accounts that emphasize intent. To elucidate, consider the case of a "noble lie," which is falsehood deliberately propagated for a presumed greater good. A typical example is one propagated by the political elites to maintain social harmony. I find it difficult to accept that in such cases *epistemic harm is in fact intended*. Or take the more epistemic example of a noble lie, where an almost always reliable source of information, once having spread misinformation, decides to cover up with a lie to protect its credibility in the face of other very unreliable sources. In such, and similar cases, "disinformation" cannot even be said to be intended to cause epistemic harm, as those trusting this source are epistemically better off continuing to rely on it for information rather than others. Could the epistemic harm account be salvaged? I think it can be, if we distinguish between content and activity as well as introduce "capacity" into the definitions, all the while hyperspecifying the epistemic harm at stake in each situation, but then what we will end up with is very different from the current understanding of harm-based accounts, and will look more like my information-hazard-based account so I will not discuss it further.

From what has been said so far, including the problem of opacity, the disagreement over what counts as harm as well as the moral and legal stakes that hang on these, we are led to conclude that the "intent to harm" is inoperable and as such unsuitable in delineating dis/misinformation. The emphasis on operability cannot be overstated. That is because, as noted earlier, the recent explosion in the dis/misinformation literature is largely due to practical consideration. There is a general sense of urgency, and a need for action in the face of what some have even dubbed as an "information pandemic" or "infodemic" (Zarocostas 2020). As such, we not only want that our concepts be clear but also that they can be put to good use in addressing information problems.

3.2. Misleading-based account

Problems such as these have led many scholars to drop "to harm" entirely and replace it with something more specific, namely "to mislead" (e.g. Guess & Lyons 2020; Chadwick & Stanyer 2022). A notable example is Floridi (2013), who defines disinformation as "misinformation purposefully conveyed to mislead the receiver into believing that it is information" (260) – where information, for Floridi, is true. Another example is one of Fallis's (2014b) definitions of disinformation as

> information that is intentionally misleading. That is, it is information that – just as the source of the information intended – is likely to cause people to hold false beliefs (137).

It should be noted that cashing out dis/misinformation in terms of the intention to mislead already constitutes an improvement over the harm-based accounts since, unlike harm, it provides something more specified for us to evaluate, namely falsehood. It is important to keep in mind that "to mislead" in this context is generally taken to mean to cause false beliefs (see Harris 2023). Yet, misleading-based accounts of dis/misinformation are also unduly limiting since in many instances disinformation does not simply aim to cause false beliefs but instead aims to block true beliefs. A textbook example is how a small number of scientists backed by the tobacco industry were able,

despite overwhelming evidence, to sow the seeds of doubt about the link between the use of tobacco and cancer (Oreskes & Conway 2011). Here, and in similar cases, the aim is not necessarily to cause false beliefs but to prevent people from forming true beliefs despite the evidence. This consideration has later led Fallis to amend his definition of disinformation to include that it is likely to create false beliefs *or* prevent true beliefs, but even this, as will be clear, is also too narrow (2015, 420).

However, and notwithstanding its improvements, the misleading-based account also fails to address the problem of opacity. Also, and not unlike other accounts that lump content and activity, it suffers from the possibility of cross-classifying misleading semantic content. Is my grandfather spreading *mis*information or *dis*information when he shares a deliberately misleading Tiktok video on the side effects of the Covid-19 vaccine? The answer is not straightforward. This is because disinformation is sometimes taken to be distinct from misinformation and sometimes a subspecies of misinformation – depending on whether or not intentionality is included misinformation (see Fallis 2016).

All of this has led some to suggest that we drop intentionality altogether from our definitions and classifications of dis/misinformation (Corce & Piazza 2021). Others, such as Fallis (2014a), suggest that we can do away with intentionality by replacing it with "function," in that although such semantic content may not be *intended* to mislead, they nonetheless have a *function* of misleading. Here, Fallis is taking a page out of evolutionary biology, according to which function needs not involve teleology. But why is intentionality so important for most authors despite its problematic status?

The answer to me seems to be an ethical one. That is while false semantic content is spread readily, we want a way to distinguish between honest mistakes, and deliberate distortions, and intent plays a major role in that. I do not want to take away from this ethical consideration. But I take issue with the fact that it has been made integral to what I consider to be a largely epistemic but more importantly ontological/conceptual issue. As noted earlier, I do not intend for ontology to be understood in any deep sense, but I do want that our concepts and classifications be answerable to reality. Failure of current definitions and classifications to do so has led to the difficulties we have seen regarding the proper place of intentionality within our definitions of semantic contents.

Let's start with a question. Why is it that when speak of "information," that is true semantic content, we do not bother with intentionality? The answer to that is that semantic contents to the extent that they are contents, be they of propositions or representations – I am not interested in taking a position on this, are not the kinds of things that include intentionality. But when we speak of information, not as content but as a communicative *act* we then include intentionality. This distinction between *content* and *activity* is something I already alluded to earlier when I attempted to specify the current information problem as the undermining of the epistemic aim of information *as an activity*. Here intentionality determines the aim of activity. But I also said that what I suggested as a modest epistemic aim of information, which I construed as *to seek and accept truth and/or avoid and reject falsehood*, is actually a bare minimum condition for successful information as a communicative activity. In a nutshell, "information" can denote two *ontologically* different things, one is content which does not include intentionality and the other is activity which includes intentionality.

Bringing what we have learned so far to the problem of defining and classifying dis and misinformation, we can see where almost everyone has gone wrong. That is because they have mixed content with activity which has resulted in such "disfigured concepts." Let us look at "misinformation." It is currently used, I said, to denote false semantic content that is not intended to harm/mislead. But, the current definitions aside, what does our pre-theoretic understanding of "misinformation" tell us it means? The answer is that it depends on how it is used in a sentence.

When I say that "Sam is spreading misinformation," I mean something along the lines of that there is a thing that Sam is spreading, an object of Sam's activity, which is misinformation *as content*. To be precise the object is not really the content itself but the bearer of such content, whether that is taken to be propositions or representations – but let us not worry too much about that here. And content qua content, I said, is independent of intentions. But if I say that "Sam is engaging in misinformation," I do not take "misinformation" to be an object, for we do not engage in an object, but an activity. In this case, misinformation is the activity of spreading false semantic content with no intention to mislead – we are sticking with "mislead" for the moment. But what about disinformation? It seems that similar patterns of reasoning can be applied to it as well, in that as far as language is concerned, we can also have disinformation as content and disinformation as an activity, with the latter but not the former involving intentionality (to mislead).

Such observations simple though they may be, actually lead to interesting results, namely, that in so far as content is concerned both misinformation and disinformation mean the same thing: false semantic content, period – this will further be refined later. In this case, it is recommended that we drop one of these as it really is redundant. Now choosing which to drop and which to keep I think is a largely conventional issue, where a convention is understood as a solution to a coordination problem. So I will suggest that we keep "misinformation" and drop "disinformation" when we mean false semantic content. This is somewhat in line with some uses of misinformation where it is taken not to include intentionality. It follows from that that we are now able to say, as a first approximation, that someone can engage in disinformation (activity) by spreading misinformation (content). Here, we have malicious intent using false semantic content.

So what we end up with are three concepts, misinformation (as activity), disinformation (as activity), and misinformation (as content). Linking that to our previous discussion of the epistemic aim of information and information hazards, we are able to see that the undermining of epistemic aims of information can take place through activities and through contents. Granted, I said that for an activity to constitute an information hazard it will have to use contents. But it is important to tease out these two things when zeroing in on the problem because it allows us to address it with higher precision.

In fact, this distinction helps us to realize that the information problem can be broken down into two more specific but very much related problems, each perhaps requiring its own approach(es) and solution(s): the *content problem of information(hazards)* and the *activity problem of information (hazards)*.

The former, that is, the content problem, concerns misinformation "in the wild," that is regardless of who/how/why it is being communicated. It almost takes misinformation as existing as "facts" that need to be dealt with "out there." This generally deals with them, in a post hoc fashion, that is, after their dissemination (but see Lewandowsky &Van Der Linden 2021). The latter, that is, the activity problem, concerns the very communication activity which includes among other things, the relation between source and target, including epistemic and moral trust, expectations as well as other social psychological variables that go into communication.

The content problem, I would say, is the one that most scholars were/are concerned with, especially when they speak of "degradation," "fragmentation," or "pollution" of the "information environment" or similar spatial metaphors (e.g. Ahren et al.2016; Kraft *et al.* 2020; Toff & Kalogeropoulos 2020). This, however, they do unwittingly as they lack the conceptual distinction between content and activity, which ultimately created for them the placement problem.

I, on the other hand, recognize the conceptual distinction, and as such distinguish between the content and activity problem of information. And while, I acknowledge that solving the information problem will require attending to both of them. My "conceptual therapy" contributes to solving the content problem, if only, by exposing the placement problem that has plagued current accounts as a pseudo-problem. This, I believe, constitutes progress.

To summarize, our investigation has led us to conclude that the problem of information hazards can be broken down into a content problem and an activity problem, each of which, can lead to the undermining of the epistemic aim of information. We also ended up with three different concepts. The first is misinformation as content, which is intent-independent, the second is misinformation as activity and includes intent, and the third is disinformation as activity.[9] I also said that cashing out intent in terms of misleading, although an improvement on harm-based accounts, is nonetheless unduly limiting. In what follows I consider one of the most recent and more sophisticated accounts: Mona Simion's (2023), which replaces "misleading" with "generating ignorance."

3.3. Ignorance-based account

Simion, who is mainly concerned with disinformation, defines it as ignorance-generating content. More precisely, she defines it as follows:

> X is disinformation in a context C iff X is a content unit communicated at C that has a disposition to generate or increase ignorance at C in normal conditions (6).

Where to generation/increase ignorance for Simion is understood as:

> to fully/partially strip someone of their status as knower, or to block their access to knowledge, or to decrease their closeness to knowledge (6).

I have noted earlier how misleading-based account, although an improvement on harm-based accounts, are still incomplete, especially when to mislead is understood as to cause false beliefs. That's because I explained that one can disinform without causing false beliefs but, for example, by getting people to suspend judgment about a true claim when there's enough reasons not to. That's why Simion's account centers not on false beliefs but on the generation and/or increase of ignorance. This, to me, on the whole, is a welcome addition, but as I will explain later, its extent is not fully appreciated even by the author herself.

Now although Simion takes herself to be mainly interested in disinformation, she does discuss misinformation as she argues against what she calls "disinformation orthodoxy." The latter she cashes out in terms of three assumptions that she discusses and ultimately rejects (3).

The first assumption is that disinformation is a species of information. This, I have discussed earlier, and while I largely agree with Simion's conclusion, I said that it is not really a substantive disagreement since information can be used in a strict and loose sense and treating disinformation as a species of information can be understood as using the latter in a loose sense.

---

[9]As will be clear later, disinformation as content, can still have a place, but it ceases to be as central.

The second assumption is that disinformation is a species of misinformation. Here, Simion tries to make the argument that the mis prefix modifies as wrongly, badly, or not. While the dis prefix modifies as a depravation, exclusion and expulsion. She remarks that dis does not simply negate, but rather it "undoes." She provides the convenient example of "misplacing" as wrongly placing and contrasts it with "displacing" as taking out of place. However, Simion is well aware that her linguistic distinction is porous, even shaky, when she points out that there are cases where dis modifies as not such as in "disagreeable" or "dishonest." This leaves the question of whether disinformation is a species of misinformation unresolved – I will later argue that in certain cases disinformation is misinformation.

The problem with Simion, however, is that she, like everyone else in the field, fails to distinguish between content and activity. Consider how she combines both "content" and "communication" in her definition. This is also made clear at a more practical level, when she claims that disinformation, as opposed to misinformation, is not essentially false since she is able to disinform using true content and generating false implicatures (5). The key word her is "to disinform," which denotes a verb, and verbs refer to actions/ activities and not to contents.

The third assumption that she considers and rejects is that disinformation is essentially intentional/functional. Simion rejects this almost unanimously held assumption using first pragmatic considerations and second a thought experiment. I consider both in turn. The pragmatic consideration tells us that it is ill-advised to take disinformation as intentional, when, with the advent of black-box Artificial Intelligence (AI), a certain AI system may spread misleading content with neither intention nor function. In such cases, according to Simion, we have disinformation, but it includes neither intention nor function.

Information problems resulting from AI advancement notwithstanding, it seem to me that what worries Simion is also the content problem of information. Here seems to be an implicit assumption of hers, that the content problem of information hazard is largely a *dis*information one – in the conventional sense, or that the target of the problem deserving of our intervention is disinformation as opposed to misinformation. This, however, I think is a mistake, not least because I do not accept the current confused definitions, but more importantly if our worry is the erosion of truth from the information environment, then it makes little difference whether this is taken place through dis or misinformation. Next, let us consider her thought experiment.

Simion invites us to consider a trusted journalist that is interested in highlighting even the slightest of scientific disagreements, thereby making certain scientific issues seem far less settled than they really are. Examples she gives are climate change and vaccine safety. Simion maintains that this is a classic case of disinformation, which constitutes a litmus test for accounts of disinformation. Let us remember here that this example is meant to support, or more precisely illustrate, that disinformation is not, or should not be construed as intentional. After all, the said trusted journalist has no intention to mislead. Also, according to Simion, the journalist cannot be said to be spreading false content/ misinformation since he/she is communicating genuine, albeit fringe, scientific findings.

Notice that for Simion, as for others, the case is dichotomous. That is, it is either a case of misinformation or disinformation, and since it is shown that it is not the former – the journalist is not spreading false content, then it must be the latter, which makes it, according to Simion, a "classic case of disinformation spreading" (5). It should be clear, by now, however, that this, in fact, is a case of false dichotomy. As I have been arguing, the distinction between content and activity makes for a richer conceptual toolkit to construe the situation. So we need not concede that it is in fact a case of disinformation to begin with. In what follows I show how my account passes Simions's litmus test.

What activity is the journalist engaging in? On my account, the journalist is not engaging in disinformation, since the latter involves the intention to generate ignorance – here I am on board with Simion's replacement of misleading with ignorance generation as hers is more encompassing. What he/she is engaging in is the activity of misinformation. But how so, given that he/she is not using false content? My response is that the *activity* of misinformation can take place through false content *or* true content with false implicatures – this will be further refined. Indeed, what we care about with regard to content is the *capacity* of the content to generate ignorance. The reason why falsity is thought important in this context is that generally false content has a higher capacity to generate ignorance, if only because, taken at face value, false content itself generates ignorance.

Back to the journalist, what is he/she doing? Answer: misinforming the public by spreading true content with false implicatures. And this way we have answered Simion's challenge. But Simion's challenge does help us refine our account. After all, we started off by taking misinformation as content to strictly include false content, but now we realize that condition should be relaxed – I explain how that is done in due course.

After putting forward her account, Simion provides what she takes to be a comprehensive list of the means by which ignorance is generated/increased. This includes:

1- Disinformation via content that has the capacity to generate false beliefs.
2- Disinformation via content that strips knowledge via defeating justification.
3- Disinformation via content that has the capacity to induce epistemic anxiety.
4- Disinformation via content that defeat justification/doxastic confidence.
5- Disinformation via content that carry false implicatures (7).

I think that Simion's ignorance generating/increasing list is important even if I disagree with her classification. And although I think that she does better than perhaps all of the current accounts on offer, I think that her attempt at comprehensiveness fails partly for conceptual reasons that I have been discussing (i.e. failure to distinguish between activity and content), but also partly because she fails to consider, within the information hazard framework, means that are non-epistemic but which have epistemic results. What I am mainly concerned with here are psychological means, that is, cognitive and affective. Notice that Simion's discussion involves beliefs, justification, knowledge, defeaters, etc., all of these are epistemic in nature, and for good reasons. Ignorance itself is an epistemic concept. But people are not ideally epistemic agents any more than they are ideally rational agents.

Although Simion account is somewhat protected from such criticism when she includes "normal conditions" in her definition of disinformation – which could be said to include psychological aspects, I think this is problematic since normal conditions in communication or/and in content appraisal often involve causally relevant emotional/affective components and ignoring this point or brushing it aside undermines Simion's avowed attempt at comprehensiveness. Indeed, that psychological, including emotional, elements are usually involved or at least can be easily brought to bear on cognition is widely recognized, especially in fields that study judgment and decision-making (JDM), such as social psychology (see Keren & Wu 2015). Specific examples include the role of attention in JDM (Mrkva *et al.* 2020), or the role of emotions in JDM (Västfjäll & Slovic 2013). Such considerations play important roles in the means of ignorance generation. For example, inducing someone into a state of anger or fear may render them more likely to accept ignorance generating/increasing content that is mood-congruent (see Blanchette &

Richards 2010). All of that is to point out that despite its attempt at comprehensiveness Simion's list constitutes only a specific, precisely epistemic, subset of ignorance-generating means, and that a more comprehensive list would have to include psychological kinds of means.

4. Information hazard-based account

Let me link my discussion of Simion's to my initial remarks on the study of disinformation. I started out by pointing out the confusion in the literature that exists at the level of classifications of dis/misinformation. I defended that a key underlying reason for that is the lack of clarity of the problem that people are working to solve, and I attempted to remedy that by further clarifying what it take to be the problem and by making certain assumptions explicit. I made a couple of, what I take to be, innocuous assumptions in the study of mis/disinformation, beginning with information, as activity, itself.

I noted that, to me, a modest epistemic aim of information as an activity is to seek and accept truth and/or avoid and reject falsehood. I coined "information hazard" to denote anything that through and including content can undermine the achievement of the epistemic aim of information –what can be understood as the *means*. This is not so different from Simion's ignorance-based account, although, given her overemphasis on epistemology, she fails to appreciate ignorance-generating/increasing means that are not strictly epistemic. Also, given her knowledge-first epistemology, she cashes out her account in terms of knowledge, while mine concerns truth and falsehood. I find my suggestion more appealing as it does not require one to take a stand on the analyzability of knowledge. Lastly, despite the breadth of her ignorance-based account, and her attempt at comprehensiveness, the list Simion provides, I said, constitutes only an epistemic subset of means which, in her own words, are ignorance generation/increasing, and in my words constitute information hazards.

Now it becomes clearer why I construed the latter using the general "anything," because I wanted it to include activity and content, the epistemic and the non-epistemic including the psychological, which itself includes the cognitive and the affective. Having clarified all of that we are able to provide what I take to be a more satisfactory characterization of dis/misinformation.

Misinformation as content: Given a context C, target audience A, X is misinformation iff X is a content with a high capacity given C to undermine the epistemic aim of information for A. (X could have this high capacity either because it is itself false, or otherwise, e.g. likely given C to generate false implicatures for A – here Simion's aforementioned list is relevant).

Sometimes X will be C and A invariant, especially when the content of X is false. The latter can be dubbed "disinformation" (as content) and can be maintained as a subspecies of misinformation as content where X is false. Sometimes X will be A *invariant*, especially when there is a strong semantic link between the content and its implicatures. What this shows is that misinformation as content constitutes a gradient, with some cases taken to be clear cut cases of misinformation, whereas others are less clear cut and require appreciation of the context, and we will likely have cases where the situation is too blurry to tell.

Misinformation as activity: given C, A, X, source S, an activity M is misinformation iff M is an activity performed by S to communicate X to A at C with no intentions to undermine the epistemic aim of information for A.

Disinformation as activity: given C, A, X, source S, an activity D is misinformation iff D is an activity performed by S to communicate X to A at C with the intentions to undermine the epistemic aim of information for A.

Note that in both definitions of dis and misinformation X is used as defined in misinformation as content, which is roughly content with high capacity to undermine the epistemic aim of information. That is because on my account both activities can only take place through misinformation as content.

While it is true that the two forms of activities above include intentions, and I have criticized some of the account above for failing to solve the problem of opacity. I think that having clarified the problem as that of information hazard, and, further, breaking it down into the content problem and the activity problem, with the former problem being of greater concern for people currently working on dis/misinformation, I think that my classification largely, circumvents the problem of opacity by making misinformation as content intention-independent.

This, incidentally, but not without merit, answers Simions aforementioned black-box AI challenge – which is motivated by pragmatic concerns, where false content are spread by these bots with neither intention nor function to disinform. Previous accounts, which take disinformation to include intention, Simion rightly notes, fail to answer that challenge, while her account which removes intentions from disinformation succeed. My account says that there are ontological/conceptual considerations that come even prior to Simion's pragmatic concerns, and these are that dis/misinformation if treated as content should not have included intentions in the first place.

Alternatively, my answer is that the problem of these bots, or at least the aspect of which that Simion is worried about, belongs to the content problem of information hazard. Accordingly, what is being spread is misinformation, and if content is false then disinformation, and should be addressed accordingly. Hence, unlike Simion, whose pragmatically motivated account is fixated on disinformation – as construed by her, and as such is limited by the content problem, mine, which is conceptually/ontologically motivated, is broad enough to recognize that there are two kinds of information problems. This, I believe, helps to bring more clarity to the mis/disinformation research.

The last significant addition I offer over Simion's account, as I have pointed out earlier, is that she is overly concerned with the epistemological means that she failed to appreciate other, broadly psychological, ones. This I said is a problem for her claim of comprehensiveness, although to be fair, not to her account, because her construal of "ignorance" is in principle broad enough to include such considerations, even if, in practice, she fails to do so.

In the final analysis, unlike current accounts in the literature which lack theoretical basis, I have provided a theoretical framework through which to understand the current problem of information hazards. I believe that the requirement for the epistemic aim of information that I have put forward is modest enough to appeal to many scholars working in the field without compromising rigor – hence hopefully providing unity to what is currently a fragmented field. It is through this theoretical framework that we realize the sub-divisibility of the information hazards problem into two distinct but related problems. Also, the fact that it is conceptually grounded all the while having an eye for practical concerns helps to bridge the gap between social scientists, who are usually mostly concerned with practice, and philosophers, who are usually mostly concerned with theory. This is evidenced by the fact that it is able to accommodate actual cases of scientists and thought experiments of philosophers. This broad account, to be sure, is only a first step but remains nonetheless crucial for a more principled approach to the study of information hazards.

5. Conclusion

This paper aimed to contribute to removing the current conceptual confusion that exists in the dis/misinformation literature by providing a more ontological grounded account. I started by clarifying what I take the problem to be, which I construed as that of "information hazards." The latter I cashed out in terms of means that, involving content, undermine the epistemic aim of information as activity, where this aim is minimally taken to be to seek and accept truth and/or avoid and reject falsehood. Then I criticized three of the more currently popular accounts (harm-based, misleading-based, ignorance-generating-based) on ontological, conceptual, and practical grounds. I finished by offering my own information hazard-based account, which, being erected on solid conceptual/ontological grounds all the while having an eye for more practical concerns overcomes many of the difficulties that have plagued previous accounts, promises to add clarity to dis/misinformation research and prove more useful in practice.

References

Ahern L., Connolly-Ahern C. and Hoewe J. (2016). 'Worldviews, Issue Knowledge, and the Pollution of a Local Science Information Environment.' *Science Communication* **38**(2), 228–250.

Aïmeur E., Amri S. and Brassard G. (2023). 'Fake News, Disinformation and Misinformation in Social Media: A Review.' *Social Network Analysis and Mining* **13**(1), 30.

Alvarez M. (2016). 'Reasons for Action: Justification, Motivation, Explanation.' *The Stanford Encyclopedia of Philosophy*. Stanford, CA: Metaphysics Research Lab, Stanford University.

Bernecker S., Flowerree A. K. and Grundmann T., eds. (2021). *The Epistemology of Fake News*. Oxford: Oxford University Press.

Blanchette I. and Richards A. (2010). 'The Influence of Affect on Higher Level Cognition: A Review of Research on Interpretation, Judgement, Decision Making and Reasoning.' *Cognition & Emotion* **24**(4), 561–595.

Chadwick A. and Stanyer J. (2022). 'Deception as a Bridging Concept in the Study of Disinformation, Misinformation, and Misperceptions: Toward a Holistic Framework.' *Communication Theory* **32**(1), 1–24.

Croce M. and Piazza T. (2021). *Misinformation and Intentional Deception: A Novel Account of Fake News*. London: Routledge.

Dame Adjin-Tettey T. (2022). 'Combating Fake News, Disinformation, and Misinformation: Experimental Evidence for Media Literacy Education.' *Cogent Arts & Humanities* **9**(1), 2037229.

Fallis D. (2014a). 'A Functional Analysis of Disinformation.' *IConference 2014 Proceedings*. https://doi.org/10.9776/14278.

Fallis D. (2016). 'Mis-and Dis-Information.' *The Routledge Handbook of Philosophy of Information*, pp. 332–346. London: Routledge.

Fallis D. (2014b). 'The Varieties of Disinformation.' *The Philosophy of Information Quality*, pp. 135–161. Dordrecht: Springer Science and Business Media B.V.

Floridi L. (2013). *The Philosophy of Information*. Oxford: OUP Oxford.

Freiling I., Krause M. N. and Scheufele A. D. (2023). 'Science and Ethics of "Curing" Misinformation.' *AMA Journal of Ethics* **25**(3), 228–237.

Guess A. M. and Lyons A. B. (2020). 'Misinformation, Disinformation, and Online Propaganda.' *Social Media and Democracy: The State of the Field, Prospects for Reform* Cambridge: Cambridge University Press.

Harris K. R. (2023). 'Beyond Belief: On Disinformation and Manipulation.' *Erkenntnis*. https://doi.org/10.1007/s10670-023-00710-6.

Hayward T. (2023). 'The Problem of Disinformation.' *SSRN Electronic Journal*. http://dx.doi.org/10.2139/ssrn.4502104.

Keren G. and Wu G., eds. (2015). *The Wiley-Blackwell Handbook of Judgment and Decision Making*. Hoboken, New Jersey: Wiley-Blackwell.

Kraft P. W., Krupnikov Y., Milita K., Barry Ryan J. and Soroka S. (2020). 'Social Media and the Changing Information Environment: Sentiment Differences in Read Versus Recirculated News Content.' *Public Opinion Quarterly* **84**(S1), 195–215.

Lazer D. M. J., Baum A. M., Yochai Benkler A. J. Berinsky K. M. Greenhill F. M., Metzger J. M. et al. (2018). 'The Science of Fake News.' *Science* **359**(6380), 1094–1096.

Levy J., Bayes R., Bolsen T. and Druckman N. J. (2021). 'Science and the Politics of Misinformation.' *The Routledge Companion to Media Disinformation and Populism*, pp. 231–241. Milton Park, Abingdon: Routledge.

Lewandowsky S. and Van Der Linden S. (2021). 'Countering Misinformation and Fake News Through Inoculation and Prebunking.' *European Review of Social Psychology* **32**(2), 348–384.

Meel P. and Kumar Vishwakarma D. (2020). 'Fake News, Rumor, Information Pollution in Social Media and Web: A Contemporary Survey of State-of-the-Arts, Challenges and Opportunities.' *Expert Systems with Applications* **153**, 112986.

Mrkva K., Ramos J. and Van Boven L. (2020). 'Attention Influences Emotion, Judgment, and Decision Making to Explain Mental Simulation.' *Psychology of Consciousness: Theory, Research, and Practice* **7**(4), 404.

Oreskes N. and Conway M. E. (2011). *Merchants of Doubt: How a Handful of Scientists Obscured the Truth on Issues from Tobacco Smoke to Global Warming.* New York, NY: Bloomsbury Publishing USA.

Roozenbeek J., van der Linden S., Goldberg B., Rathje S. and Lewandowsky S. (2022). 'Psychological Inoculation Improves Resilience against Misinformation on Social Media.' *Science Advances* **8**(34), eabo6254.

Scarantino A. and Piccinini G. (2010). 'Information without Truth.' *Metaphilosophy* **41**(3), 313–330.

Shu K., Wang S., Lee D. and Liu H. (2020). 'Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements.' *Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities*, pp. 1–19. Cham: Springer.

Simion M. (2023). 'Knowledge and Disinformation.' *Episteme* **2023**, 1–12.

Toff B. and Kalogeropoulos A. (2020). 'All the News that's Fit to Ignore: How the Information Environment does and does not Shape News Avoidance.' *Public Opinion Quarterly* **84**(S1), 366–390.

Västfjäll D. and Slovic P. (2013). 'Cognition and Emotion in Judgment and Decision Making.' *Handbook of Cognition and Emotion* **252**, 271.

Wardle C. and Derakhshan H. (2017). *Information Disorder: Toward an Interdisciplinary Framework for Research and Policymaking.* Strasbourg: Council of Europe.

Wardle C. (2018). 'The Need for Smarter Definitions and Practical, Timely Empirical Research on Information Disorder.' *Digital Journalism* **6**(8), 951–963.

Zarocostas J. (2020). 'How to Fight an Infodemic.' *The Lancet* **395**(10225), 676.

Omar El Mawas is a Postdcotoral Researcher in Philosophy and Policy at the Ecole Normale Supérieure in Paris and an Honorary Fellow in the Philosophy department at Durham University. He is a former Postdoctoral Researcher in Philosophy of Science at the Institute for the History and Philosophy of Science and Technology at the Université Paris 1 Panthéon-Sorbonne. He is interested in the General Philosophy of Science (esp. The Scientific Realism debate) and its relation to Epistemology, Philosophy of Social Science, Evidence, and Objectivity.