

Bullshit activities

Kenny Easwaran

University of California, Irvine, Irvine,
California, USA

Correspondence

Kenny Easwaran, University of
California, Irvine, Irvine, CA, USA.
Email: k.easwaran@uci.edu

Abstract

Frankfurt gave an account of “bullshit” as a statement made without regard to truth or falsity. Austin argued that a large amount of language consists of speech acts aimed at goals other than truth or falsity. We don’t want our account of bullshit to include all performatives. I develop a modification of Frankfurt’s account that makes interesting and useful categorizations of various speech acts as bullshit or not and show that this account generalizes to many other kinds of act as well. I show that this illuminates some of Graeber’s classification of “bullshit jobs,” though it doesn’t fully agree with it.

Harry Frankfurt (1986) gave a famous account of bullshit, as a statement made without regards to truth or falsity, as opposed to ordinary truth-telling or lying. Austin (1975, 1979) argued that a large amount of language (perhaps even the majority) consists not of statements aimed at truth or falsity, but rather at performing other kinds of speech act than stating. Obviously, most such performative utterances are not bullshit, but I argue that there is a natural extension of Frankfurt’s concept of bullshit that applies to these other speech acts as well. Since many of the activities performed by speech acts can also be performed by other means, this extension of the concept of bullshit naturally goes beyond language and defines a broader range of bullshit activities. There has been some investigation of this concept in the context of sport (Howe, 2017), but not much other investigation of bullshit activity by philosophers. The anthropologist David Graeber (2013, 2018) has attracted public attention for his discussion of the phenomenon of “bullshit jobs.” While I do not endorse his particular diagnosis of which jobs are bullshit, I suggest that several aspects of his analysis of the phenomenon are in keeping with my extension of Frankfurt’s concept of verbal bullshit to bullshit activities in general, though some others are not.

One point of clarification before I begin. Some have argued that Frankfurt is incorrect to focus on the intentions of the bullshitter, and instead should focus on the features of the statement that

This is an open access article under the terms of the [Creative Commons Attribution-NonCommercial-NoDerivs](https://creativecommons.org/licenses/by-nc-nd/4.0/) License, which permits use and distribution in any medium, provided the original work is properly cited, the use is non-commercial and no modifications or adaptations are made.

© 2023 The Authors. *Analytic Philosophy* published by John Wiley & Sons Ltd.

constitutes the bullshit. At least sometimes, they think an audience can identify bullshit that the producer did not intend as such. While these authors may (or may not) be right about many of the ordinary language uses of the term “bullshit,” I think the concept that Frankfurt identifies is a pragmatically useful one to explicate and extend helps us identify an important class of activities. Along the way, I will show how audiences often apply something like this Frankfurian concept in calling something bullshit that was not produced in Frankfurt’s way, because they reasonably believe that it must have been.

1 | FRANKFURT ON BULLSHIT

Harry Frankfurt’s characterization of bullshit is as a statement that is made without concern for truth or falsity. Cohen (2002) and Wren (2013) argue that Frankfurt is wrong to focus on the mental state of the bullshitter, the “concern for truth,” and instead argue that we should identify which *statements* are bullshit. They propose that we only derivatively identify the bullshitter as the person who either intends to produce bullshit or who is disposed to regardless of having such an intention. Cohen (p. 332) ends up concluding that bullshit is any utterance that it is unclarifiable nonsense (and perhaps ideally, is unclarifiable nonsense that an innocent audience might think has some clarifiable meaning). Wren, similarly, concludes that bullshit is “noxious nonsense, with ‘noxious’ meaning being worthless or less than worthless, positively repellent or harmful, and ‘nonsense’ meaning not fitting in with or contributing to understanding or knowledge of an issue or topic, or not being relevant to the solution of a problem or the accomplishment of a task at hand.” (p. 113)

However, I agree with Frankfurt that the word “nonsense” (perhaps with an intensifier like “unclarifiable” or “noxious”) is already sufficient for these purposes, and it is useful to reserve the technical use of a distinct word like “bullshit” for a different category of speech act. Furthermore, as Frankfurt notes, the category of a lie “is identical neither with the falsity nor with any of the other properties of the statement the liar makes, but ... requires that the liar makes his statement in a certain state of mind – namely, with an intention to deceive.” There is value in making the term “bullshit” parallel to “lie,” so that it also depends on the state of mind of the producer. Once we have a characterization of what it is to lie or to bullshit, we can then use audience impressions of whether the speaker’s activity fits this characterization to figure out whether an audience is likely to judge a statement itself as a lie or as bullshit. (As I note later, defining it in terms of state of mind does not entail that the bullshitter is always aware of when they are bullshitting, since we are not always perfectly aware of our states of mind.)

As Grice (1957) noted, to mean something is to intend for one’s audience to come to believe it, and to intend this belief to come about by means of the audience recognizing one’s intention. This account is clearly not correct as an account of the semantic meaning of one’s utterance (as Grice noted in his 1975); this is better as an account of the pragmatic meaning of an utterance, which may diverge greatly from the semantic meaning, but it helps categorize honesty, lying, and bullshit (at least, as these categories apply to statements – in later sections I will consider parallel categories for other speech acts). An honest speaker means, in the Gricean sense, the content of what is said, and also believes it. A liar means it, but believes it to be false. I take Frankfurt to be proposing that a bullshitter does neither. As a rough characterization:

- A speaker is honest if their primary intention involves the hearer coming to believe what they say, by means of trusting their honesty, and they believe that what they say is true.

- A speaker is lying if their primary intention involves the hearer coming to believe what they say, by means of trusting their honesty, and they believe that what they say is false.
- Otherwise, a speaker is bullshitting.

There are of course challenges to these definitions. A collection of challenges to this account of lying is discussed by Mahon (2015) and citations therein, and challenges to the idea that bullshitting and lying are exclusive categories given by Stokke and Fallis (2017). My goal is not to give a perfectly extensionally adequate theory of the ordinary use of these terms, but to give a theoretically elegant account that generalizes to speech acts beyond statements and to acts beyond speech acts. Whether or not the concept I define is coextensive with any particular phrase we use in ordinary language, I hope that the way I define it shows that this concept is useful, and can do much of the work that we ordinarily do by using the word “bullshit.”

Frankfurt says at various points that the bullshitter is not motivated by a concern for the truth, the way that both the truth-teller and the liar are. However, at times he is ambiguous about whether he means that the bullshitter is unconcerned with the *actual* truth or falsity of their statement or that the bullshitter is unconcerned with whether the audience comes to *believe* in the truth of their statement. The account I have given allows for either. There is a characteristic intention that is required for a statement, and a characteristic belief that plays some role in making this intention sincere. When someone has the opposite of the belief, the statement is a lie. But when someone merely *lacks* one of the two components, then on my account that person is bullshitting. Note that on this account there are two ways a speaker can bullshit:

- A speaker can bullshit in the first way if their primary goal does not involve the hearer coming to believe what they say, by means of trusting their honesty.
- A speaker can bullshit in the second way if their primary goal involves the hearer coming to believe what they say, by means of trusting their honesty, but the speaker does not believe what they say to be true or believe it to be false.

I do not claim that this disjunctive account is what Frankfurt meant or that this is precisely how ordinary speakers use the word. But I do claim that it is a conceptually useful account somewhere in the vicinity of both. As I will show later, I think this account can productively be expanded to speech acts beyond assertions, and even to other acts.

Bullshit of the first kind is perhaps the most common. A few examples will help clarify what I mean by the primary goal not involving the hearer coming to believe what is said.¹

¹Technically, my account allows a statement to qualify as bullshit if it is made with the primary goal of the hearer coming to believe what is said, but *not* by means of trusting the speaker's honesty. As (Schiffer, 1972, p. 42) points out, Grice's own article is an example of this – Grice expects the reader to have the same intuitions as him about the cases, and to believe his claims for *that* reason, rather than out of trust.

However, in some earlier work (Easwaran, 2009), I argue that this is in fact the characteristic mode of mathematical writing, and much philosophical and other academic writing as well. Thus, I may be able to appeal to the account I develop in Section 2, and say that this work is an instance of a speech act *other* than an ordinary statement, made *with* its characteristic intention, rather than an *ordinary* statement made without the characteristic intention of a statement. Thus, I can say it does not qualify as bullshit.

If there are other cases where a speaker makes an apparent statement, with the intention that the hearer come to believe the content of the statement, but not by means of trusting the speaker, I suspect that they will be contrived enough that it is fine for my account to classify them either as bullshit or not. I thank Robbie Kubala for pressing me on this point.

Frankfurt's sole concrete example of bullshit in his essay is of this first kind – a speaker at a Fourth of July event goes on about “our great and blessed country, whose Founding-Fathers under divine guidance created a new beginning for mankind.” The point of such a speech is not to get the audience to believe anything about the divine guidance of the founders of the United States, nor about the importance of the United States for the history of mankind, but rather for the audience to believe that the speaker is an appropriately patriotic person. It may happen that some members of the audience do come to believe the content of these claims, and it may happen that some members of the audience roll their eyes at the banality of the utterance. Yet other members of the audience may not even fully process the content of the statement, and may just feel the emotional associations of various words that are used. But all of these members of the audience will perceive the speaker as the kind of person who makes public statements attributing preternatural significance to the establishment of the United States, and this perception is what the orator cares most about.

Frankfurt mentions the broader class of “pretentious bullshit.” As he says, “when bullshit is pretentious, this happens because pretentiousness is its motive rather than a constitutive element of its essence. ... It is often, to be sure, what accounts for his making that utterance. However, it must not be assumed that bullshit always and necessarily has pretentiousness as its motive.” I take it that a speaker is pretentious when they primarily intend to cause the audience to be impressed with the speaker. Pretentiousness can be bullshit when this intention does not involve the hearer believing what is said, but it can be a lie or even honesty if it does. A speaker who says, “I have a degree in mathematics from Stanford, so I understand all the statistical results in the latest CDC statement and think they are wrong” is pretentious, but is either lying or telling the truth, and not bullshitting. But a speaker who says, “The latest CDC report uses a Diffie-Hellman analysis of the quantized variance of case numbers from metropolitan and non-metropolitan counties, but a more useful analysis would use Bayesian p -values on municipal health system data” is bullshitting. In this case there is no meaning to be had in the statement, and the speaker just expects the audience to be intimidated or impressed by the jargon. Interestingly, on my account, even if the speaker said something perfectly true and correct, if they use the jargon primarily *intending* for the audience to be intimidated or impressed without being able to follow, they are still engaged in bullshit of the first kind. They are making an apparent assertion with the intention of impressing, rather than being believed.

Cohen (2002) suggests that certain academic movements may primarily be constituted by something like this sort of pretentious bullshit. However, instead of the obscurity being intended to make the audience think highly of the speaker, the obscurity is intended to make the audience think highly of the content. “Someone struggles for ages with some rebarbative text, manages to find some sense in it, and then reports that sense with enthusiasm, even though it is a banality that could have been expressed in a couple of sentences instead of across the course of the dozens of paragraphs to which they said someone has subjected herself.” While this may have something important in common with pretentious bullshit, it does seem to have an important difference in that the rhetorical technique is used to strengthen the belief in the content, rather than to believe something else about the author. Buekens and Boudry (2015) argue that it is thus better to characterize this sort of “obscurantism” as distinct from bullshit. My account agrees – so long as there really is an intention that the content come to be believed, the statement is not bullshit.

Furthermore, not every jargon-laden discussion is an instance of bullshit *or* obscurantism. As any academic knows, there are contexts in which the fastest and most precise way to communicate an idea to a colleague involves a few sentences dense with jargon. A non-specialist who sees such a statement may *think* it is just bullshit, but to the specialist, this is sometimes the best way

to communicate. I claim that the relevant difference is whether the speaker's intention is for the hearer to understand and come to believe what is said, or something else. As Åkerman (2009) suggests for complicated referential intentions, there are surely complicated uses of jargon where a speaker earnestly intends to be understood by the relevant audience, but where the relevant audience cannot find any way of successfully interpreting what was said. In this case, I claim that what was said was not in fact bullshit, but the audience would nevertheless be reasonable in having judged it to be bullshit, because that is the only interpretation they can find. I will not say more about which cases I take to be reasonable misinterpretations of this sort, which are actual bullshit, and which are cases where an uncharitable reader is missing the obvious intended use of the jargon.

Similar points go for what Tosi and Warmke (2016) call “moral grandstanding,” and what many others nowadays refer to as “virtue signaling.” Consider a corporate spokesperson expressing Coca Cola's disapproval of systemic racism, or Nike's objections to the war in Ukraine. As they say, “while moral discourse may unfold under the *pretense* of addressing injustice, many contributions are in fact intended to get others to believe that one is morally respectable.” I claim that such discourse counts as bullshit, in the sense that I interpret Frankfurt as describing, to the extent that the pretentious aim is primary. In the case of the corporate spokesperson, it is clear that the pretentious aim is primary so that the statement is, in my interpretation of Frankfurt's definition, bullshit. But many want to apply this judgment also to the statements of individuals on social media, sharing news stories or changing their profile picture in solidarity with the latest movement. I agree that in many such cases, the pretentious aim is at least part of what is going on. But I think it takes a cynical misinterpretation by the audience to deny that these cases *also* often involve a sincere and earnest belief by the social media user that the issue is important, and an equally sincere desire that the audience come to have this belief as well. Even more so than the use of academic jargon to impress, I suspect this sort of moral talk often has both intentions simultaneously.²

I offered an initial characterization of honesty, lying, and bullshit as three precise categories depending on which intentions and beliefs are primary. But I think it is just as reasonable to consider these categories as coming in degrees. One is honest or lying *to the extent* that the Gricean aim of getting the audience to come to believe the content is primary, and one is bullshitting *to the extent* that some other aim is primary. If we wanted to give a precise numerical quantification of the degree of bullshit, we might either measure the degree of strength of the non-Gricean intention, or the degree of absence of the Gricean intention, or the ratio of the strength of the two intentions. (These may not result in the same measure of degree of bullshit.) I will not pursue this thought further here, but it is possible that it may yield further insights.

There are also cases of bullshit of the first kind, where the motive does not involve the audience coming to believe the truth of what is said, that are not naturally characterized as “pretentious.” One example is described by John Searle (1965). An American soldier in the Second World War is captured by Italian troops. With only a hazy memory of high school German, he recites a line from a German poem, in a gruff tone of voice. The American soldier hopes the Italian troops

²I also think that this sort of grandstanding can occur in the complete absence of bullshit. Just as a person can earnestly and honestly state many pretentious claims about their own intelligence or wealth, they can engage in moral grandstanding by earnestly and honestly stating pretentious claims about their own moral virtue. But empirically it seems more common for people to grandstand by making claims about some already well-known object level moral outrage, in order to *demonstrate* their supposed moral virtue, rather than speaking explicitly about their own virtue.

know even less German than him, and will thus not come to believe the truth of what is said, but rather will form the belief that he is a German officer. Searle intends this as a challenge to Grice's purely intention based account of meaning, since what is said does not mean "I am a German officer," even though it is intended to get the audience to come to believe that. But I think it serves as a better example of bullshit. The speaker may or may not intend for his primary intention to be recognized, but his intention in any case does not involve the audience understanding what is said and coming to believe the content of that utterance. He is instead trying to bullshit his way out of becoming a prisoner of war.

A final example my account classifies as bullshit of the first kind occurs in the paper "Knowledge-lies" (Sorensen, 2010). These are intentionally false statements made, not for the purpose of getting the hearer to believe that they are true, but for the purpose of preventing the hearer from knowing something else. The initial example Sorensen gives comes from the climactic scene of the movie *Spartacus*. After the Roman army has defeated the rebellion of the people they have enslaved, the general demands to know which one of them is Spartacus, the leader of the rebellion, so he can be put to death. Spartacus comes forward and says, "I am Spartacus!" But then another enslaved person comes forward and says, "I am Spartacus!," followed by another, and another. After all these statements, the Roman general does not know who is Spartacus, and does not believe *any* of the statements. While Sorensen wants us to accept that these statements are in fact lies (though of a special sort), my account classifies them as bullshit. Someone who wants to classify these as lies might develop a different classification than mine (perhaps one like that of Stokke & Fallis, 2017 that allows the categories of bullshit and lie to overlap), but I hope to show with the rest of the examples in this article that my classification is useful, particularly in the way it extends to speech acts other than ordinary statements.

Bullshit of the second kind is not so common. Recall that this is bullshit where the speaker does have a primary goal that involves the hearer coming to believe the content of what is said, but the speaker neither believes this content to be true nor believes it to be false. The teller of tall tales and the embellisher of stories often fall into this category. For all they know, the details they add might be true, but they do not particularly care. If the primary goal of the embellisher is just for the audience to believe the embellished story and thereby better enjoy the experience of hearing the story, then they are bullshitting in only this second way.

However, in many cases, the embellisher might also fall into the first kind of bullshit. For instance, the primary goal of the embellisher might be for the audience to believe *that* the speaker is a good storyteller, rather than for the audience to believe *the story* and be entertained. It might also be that the embellisher tells the story merely out of love of telling the story, with no regard for whether the audience believes it or not – this would be akin to Augustine's "real liar" that Frankfurt mentions, whose goal is deception for its own sake. A bullshitter is a "real bullshitter" whose goal is just bullshitting for its own sake.

Another example of a case that falls more clearly under this second category of bullshit is suggested by some remarks towards the end of Frankfurt's essay. In a democratic society, citizens are in some sense expected to have opinions about everything that might be subject to the political process, from tax rates to abortion laws to school curricula and international relations. Sometimes, this might lead to bullshit of the first kind, where the point of speaking is not to convince the audience, but just to demonstrate that one has an opinion. But it can also lead to clear cases of bullshit of the second kind. A partisan of one political party, who prefers that party for their policies on health care, may be speaking to an undecided voter who cares about affordable housing. The partisan may have no idea how the policies of this party will actually affect the price of housing, but yet try to make a case to this voter that the stated housing policies

of their preferred party will make housing more affordable. Many of us instinctively recognize that partisan talking points are often bullshit of this second sort, where the speaker is trying to convince, and knows what they are trying to convince us of, but clearly does not actually have any information about the truth or falsity of these talking points. On my account, the die-hard partisan who truly believes the entire party line is not a bullshitter, but the slightly more moderate partisan, who knows what the party line is, and parrots it despite not having a settled view on the matter, is.

Frankfurt also gives an example from the life of Wittgenstein that may fit this second class, if we consider it to be bullshit at all. Wittgenstein's Russian teacher, Fania Pascal, was in the hospital and said, "I feel just like a dog that has been run over." This remark is said to have disgusted Wittgenstein, because she had no way of knowing the specific feeling a dog has when it has been run over. Wittgenstein seems to have diagnosed this statement as an instance of bullshit of the second kind, where the speaker expects the hearer to come to believe the content of the statement, without having an opinion herself as to whether it is true or false. Most of us would of course diagnose the case differently. We would recognize the assertion as obviously metaphorical in some way, and interpret the content of the assertion as something other than its literal meaning. We would then take the speaker to have believed this content, intended us to recognize this content, and intended us to come to believe it by recognizing their intention. We would take the speaker to be speaking earnestly, with a metaphor, rather than bullshitting. But it is helpful to see how one and the same utterance can be interpreted either earnestly or as bullshit, depending on what speech act it was interpreted as.

It is an inelegance of my account that "bullshit" ends up being a disjunctive category, with these two different kinds. But this is perhaps to be expected in the analysis of ordinary language words that have often been used without a formal theory. Because the belief in truth or falsity usually plays a significant role in the Gricean intention, it seems plausible that the paradigm instances of bullshit are instances of both kinds, and the word has come to be used indifferently for both. In British English, the word "tin" is paradigmatically used for a can that is made of the element symbolized Sn, but it can also be used for a can made of any metal, or for any object that is made of the element symbolized Sn. Sometimes, the development of a theory can lead the language to specialize on one particular usage, as with the word "fish" that historically denoted any animal that lived in the water (leading to phrases like "shellfish" and "fish tacos") but has now become more precisely defined. It still does not identify a biological clade – lungfish and coelacanths with their muscular limbs form a clade with tetrapods like frogs, lizards, and us, rather than with the ordinary bony fishes with no muscles in their fins, and these two bony groups together form a separate clade from the sharks and rays. People today are more familiar with the biological theory of fish than with actual daily experience of creatures in the water, so they have eliminated whales and shrimps from their extension (even though they have not fully adopted a cladistic biological category). Perhaps someday, when people are more familiar with philosophical analyses of bullshit than with actual daily experience of the phenomena, they may settle on using the word "bullshit" for one of both of these classes, and eliminate all other uses of the word from the extension.

The last example of Wittgenstein's teacher Pascal raises an important concern. The account seems to predict that whenever a statement is made that is not intended to be believed, then it is bullshit. However, metaphors are statements that are made, and not intended to be believed literally, and we would not want to say that metaphors are generally bullshit. And in fact, John Austin (1975, 1979) argued that there are plenty of other speech acts of this sort that aren't intended to be believed, but have some purpose that we would not think of as bullshit. Questions,

commands, promises, namings, marryings – these are all done with words by means of making statements that usually are not true or false, but have some other primary effect, and yet these are not generally bullshit. However, at least sometimes they *are* bullshit, and it would be helpful to have a theory that can say what this takes.

2 | BULLSHIT SPEECH ACTS

Austin characterizes most of these speech acts in terms of what he calls “felicity conditions.” For some speech acts, these are very parallel to Grice’s characterization of meaning, involving characteristic intentions. Where the intention of a statement is characteristically to get the hearer to believe the content of what is said, at least in part by means of recognizing the speaker’s intention, the intention of a command is to get the hearer to *do* what is said at least in part by means of recognizing the speaker’s intention, and the intention of a question is to get the hearer to reveal some information at least in part by means of recognizing the speaker’s intention. But for other speech acts, Austin notes that these “felicity conditions” may go beyond the presence of various intentions. In order for saying “I do” to get someone married, and for signing one’s name on a piece of paper to legally bind one to a contract, and for an umpire saying “yer out!” to get someone out in baseball, there are various social conventions that need to exist, like legal structures and the rules of a game, and these usually require the satisfaction of various other background conditions too, like the speaker not being already married or for someone else to hold a baseball in a particular location.

Austin distinguishes these “felicity conditions” from “sincerity conditions.” As he notes, while marrying, promising, apologizing, and other speech acts often take the grammatical form of a first person present tense verb,

one thing we must not suppose is that what is needed in addition to the saying of the words in such cases is the performance of some internal spiritual act, of which the words then are to be the report. ... If we slip into thinking that such utterances are reports, true or false, of the performance of inward and spiritual acts, we open a loophole to perjurers and welschers and bigamists and so on.

(Austin, 1979, p. 236)

That is, while the inward intention to actually do the promised act is essential for a *good* case of promising, and the inward intention to support the spouse in sickness and in health and so on is essential for a *good* case of marrying, the absence of such an intention does not mean that you did not promise or did not get married – it means that you have done it, but somehow badly.

My proposal is that the attitudes that go into the felicity conditions and sincerity conditions play the role of Gricean intentions and belief, respectively, in the categorization of honesty, lying, and bullshit. I will use the term “characteristic intentions” for the intentions that form a part of the felicity conditions characterizing the speech act, and the “sincerity attitudes” for the attitudes that form a part of the sincerity conditions. That is,

- An utterance put forward as an instance of a speech act is earnest (equivalent of honest) if it is made with the characteristic intentions of that class of speech act, and the speaker holds the sincerity attitudes.

- An utterance put forward as an instance of a speech act is insincere (equivalent of lying) if it is made with the characteristic intentions of that class of speech act, and the speaker holds contraries of the sincerity attitudes.
- An utterance put forward as an instance of a speech act is bullshit otherwise.

Depending on the details of the characteristic intentions and sincerity attitudes of the speech act, this may give rise to two kinds of bullshit.

- An utterance put forward as an instance of a speech act is bullshit of the first kind if it is made without the characteristic intentions.
- An utterance put forward as an instance of a speech act is bullshit of the second kind if it is made while the speaker neither holds the sincerity attitudes nor a contrary of them.

These definitions are best understood through examples. I will give various rough characterizations of the felicity and sincerity conditions for various kinds of speech act, and then show instances that my account classifies as bullshit.

Consider the speech act of promising. This act has been given many analyses by philosophers, and there is controversy about the details (Habib, 2022). However, I take a useful rough idea to be that a promise is characteristically made with the intention that the hearer come to have the expectation that the speaker will do what is promised. (There may be more to the characteristic intention, but my example concerns this part of the intention.) For instance, if someone has made it clear that they would like to talk over a paper idea with a friend, the friend might say “let’s meet up in the next few weeks” as a way to promise to talk over this paper idea. The sincerity attitude in this case is an intention to actually do what is promised. If the friend actually intends to meet to discuss the paper idea, this is an earnest promise, while if the friend intends to beg off, then this is an insincere promise. Intending to beg off is a contrary attitude to intending to meet up to talk over the paper. On my account, there are two ways for a promise to be bullshit – the speaker might not intend for the hearer to come to have the expectation that the speaker will do what is promised, or the speaker might lack both the intention to do what is promised and an intention to do otherwise.

As an example of the first kind, consider two old college roommates who live in the same city but have not been in touch in years, who are both out downtown with different friends and run into each other on the street. After a few moments of idle pleasantries, they might head off in their separate ways with their other friends, and one of them says “let us meet up in the next few weeks.” This has the same form as the earlier statement, and is at least intended to look like a promise with the same force as that one. However, depending on the relationship of the old roommates, it may be perfectly clear to both of them that neither has any expectation that the other will actually try to meet up. They may both intend not to meet up, but there is no deception or insincerity. The point of the statement is not to generate such an expectation in the other, but rather to be yet another idle pleasantry, to fit the approved social form of a friendship, and perhaps keep up pretenses in front of their other friends. To put it bluntly, this promise, mild as it is, is bullshit.

Examples of the second kind may be clearest when the person does not even know the content of what they have promised. When installing software on an electronic device, it is common to be required to sign an “End User License Agreement.” This purports to be a legally binding document in which the user promises not to use the app for various unwanted purposes, such as the violation of local or national laws, the transmission of pornography, the harassment of

other users, and who knows what else. Most users do not read the document, but just click “Yes” (or, in some annoying cases, scroll to the end before clicking “Yes” because the app requires at least the pretense of having read the document). Since the user did not even read the terms, they neither intend to follow them nor intend to break them. Thus, this is a bullshit promise of the second kind. (In fact, as Ray Briggs pointed out in personal conversation, because this behavior is so common, some courts have found that these contracts are unenforceable. If this is right, then one or both parties may have offered the agreement without the intention that the other formed an expectation about their behavior, and thus it might be bullshit of the first kind as well, perhaps on both sides.)

For another set of examples we can consider the speech act of warning. I take it that the characteristic intention of issuing a warning is intending for the audience to exercise caution around the thing in question, at least in part because of recognition of the intention behind the warning. When someone puts up a sign that says “beware of dog,” or says “that restaurant uses a lot of mystery meat,” they intend for the audience to exercise caution, and perhaps even entirely avoid trespassing in the yard or going to that restaurant. The sincerity attitude involves a belief that the thing in question actually does pose some kind of threat to the audience, if they are not careful. These warnings would be insincere if the dog in question is actually known to be very friendly (or if there is no dog at all!) or if the restaurant is actually known to serve well-sourced meats of high quality.

For examples of bullshit warnings of the first kind, we are looking for warnings issued for a reason other than to get the audience to exercise caution. We can think of cases where someone is legally required to issue a warning, so that they issue the warning in order to comply with the legal requirement, rather than with an intention that the audience exercise caution. California's Proposition 65, passed in 1986, provides for an agency that manages a list of chemicals and concentrations above which they are known to increase cancer risk by a 1-in-100,000 probability, and both bans businesses from discharging these chemicals into drinking water sources and requires businesses to post warning signs if such concentrations of chemicals are present on their facilities. The first part of this proposition has caused immense benefit, but for decades; the second part just led to a proliferation of signs saying

Warning: This facility contains chemicals known to the state of California to cause cancer and birth defects or other reproductive harm.

Since these signs were in gas stations, hardware stores, medical facilities, many apartment buildings and restaurants, and even at Disneyland, people often completely ignored them, and they did no good. Many businesses did not believe anyone would exercise any caution as a result of these signs, and thus they were posting them only for compliance with the law. However, starting in 2018, these signs are now required to list the relevant chemicals, so that customers and tenants can understand the difference between an apartment building whose fireplaces might occasionally exceed carbon monoxide limits and an apartment building whose flooring is emitting formaldehyde. Regardless of the presence or absence of benefit to the reader, if these signs are posted strictly out of desire for compliance with the law rather than any care about that benefit, the posting of such a sign is bullshit of the first kind.

In many cases, these warning signs are also bullshit of the second kind, where the issuer of the warning has no belief one way or another about whether the chemical being warned about in fact poses a threat. The state of California may have such a belief, and may even, as the signs claim, have knowledge, if a state is the kind of thing that can have knowledge. But the

state does not seem to be the one actually issuing the warning, and business proprietors often pay no attention beyond what they are legally required to do. This is often a characteristic feature of bullshit – one party requires a second party to make a speech act, and the first party holds the sincerity attitude for the speech act while the second party does not. A point that I will return to later is that while the absence of the sincerity attitude in the person who actually makes the speech act seems problematic, if there is in fact a reliable party who does hold the sincerity attitude, the bullshit may nevertheless serve a good function. Some customers are in fact usefully warned by modern Prop 65 warnings in California, even if most people just treat them as regulatory bullshit.

Warnings that are purely bullshit of the second kind can occur in other contexts as well. For instance, when playing a game of chess with a highly competitive friend, one might engage in trash talk to throw the opponent off. Some of this trash talk may consist of bragging about one's position, and warning the opponent about the risks posed to them by some of one's pieces. These warnings may be most effective as trash talk when they convince the opponent to worry about and watch out for pieces that are not actually threatening. But if one *only* warns one's opponent about pieces that *aren't* actually threatening, then the opponent may learn to ignore the pieces one warns about. The most effective sort of trash talk might thus be warnings issued without any belief one way or another about whether the relevant piece poses a threat. Thus, trash talk can often be bullshit of this second kind, made with the intention that the audience react to it in the characteristic way, but without the beliefs that make it earnest or insincere (and this is likely true when the trash talk consists of speech acts other than warnings as well).

I claim that the characteristic intention of a question is to get the hearer to reply with some information, at least in part due to recognition of this intention. Questions made without this intention are often called rhetorical questions. By asking a question, without expecting or stating the answer, one can often get an audience to feel some of the temptation of a salient potential answer, without recognizing the fundamental weaknesses they would be forced to confront if it were stated explicitly. I claim that this is a kind of bullshit of the first kind.

At this point, a reasonable objection might arise that rhetorical questions made in the course of a public speech are very clearly speech acts of a different sort than ordinary questions asked in a one-on-one conversation. I will discuss the difficulty of identifying speech acts in more detail later, but for now I will observe that at least sometimes, rhetorical moves of this sort are made in one-on-one conversations where they might masquerade as earnest questions. Someone who is “just asking questions” might ask, “do we really have any evidence beyond her accusation?” or “could this be a false flag operation?” not as a way to elicit information, but as a way to insinuate that their preferred interpretation of the situation is the natural one. Even if rhetorical questions in the context of an oration are not bullshit, because there is no pretense that the speaker might want the audience to give an answer to the question, this kind of “just asking questions” in an ordinary conversation more plausibly is. It masquerades as a request for information, but is actually being used as a way to instill fear, uncertainty, and doubt.

Bullshit questions of the second kind may be less conventionalized, but may still be possible, depending on how we understand the sincerity conditions for a question. I take it that something like a genuine interest in the answer to a question is the sincerity condition for a question. There are cases in which the asker already knows the answer but is trying to elicit the answer for another reason (perhaps as a test of an informant's honesty), and there are cases in which the asker does not know the answer and really does not care (as with some questions on a first date, or when a courteous host is trying to draw out a shy new guest into speaking but does not actually care about the answers). For further potential examples, see Whitcomb (2023).

In some cases, it may be hard to tell whether one or another of these kinds of cases is better classified as an insincere question or as a bullshit question. Intuitively, the strongest cases for bullshit may be when the point of eliciting the answer is sufficiently far removed from the content of what the answerer will say – for instance, if a pickpocket's accomplice stops a mark in the street, asking for complicated directions to a landmark, in order for the mark to be distracted while thinking about and explaining the answer so the pickpocket can work unnoticed. This distraction would work equally well if it used a bullshit story or a bullshit accusation rather than a bullshit question.

Nowak (2019) argued that different languages may develop many importantly different speech acts that their speakers learn how to use. If any of these have distinctive characteristic intentions, or distinctive sincerity conditions, then they also will be able to be bullshit of the first or second kind.

Academic speech communities have developed a specialized speech act of academic citation, which academics have a strong intuitive understanding of, whether or not it has been precisely specified. I take it that the characteristic intention in providing a citation is to point the reader to another academic source on the topic at hand, and perhaps get the reader to believe something about what that source says. The sincerity condition is a belief by the author that the source actually says what is suggested. There is a kind of academic dishonesty when the author believes that the source does *not* say what is claimed. But far more common is a kind of academic bullshit (of the second kind) in which the author suspects the source says *something* relevant to the topic but does not really have a belief either way about whether the source supports or opposes what is claimed.³ Bullshit citation of the first kind arises when the author is motivated by something other than pointing the reader to a relevant source – perhaps when a citation is included only to appease the referee, or to juice the statistics of oneself or a friend.⁴

Another interesting case of a class of speech acts without a formal analysis is poetry. Gerald Cohen notes that a challenge for his account of bullshit, on which anything unclarifiable is bullshit regardless of the intention with which it was made, is that “a piece of good poetry may be unclarifiable.” But in order to absolve poetry, he notes that:

An unclarifiable text can be valuable because of its suggestiveness: it can stimulate thought, it can be worthwhile seeking to interpret it in a spirit which tolerates multiplicity of interpretation, and which therefore denies that it means some one given thing, as a clarifiable piece of text does. So let us say, to spare good poetry, that the bullshit that concerns me is not only unclarifiable but also lacks this virtue of suggestiveness.

(Cohen, 2002, p. 334)

He then notes that this causes a problem for his main analysis, in that “many academic bullshitters get away with a lot of bullshit because *some* of their unclarifiabilia are valuably suggestive.” But I think it is more productive just to note that poetry has a different sort of aim than an academic text, and to use this “intentional encasement,” as he puts it, to identify bullshit within a given domain.

None of this is quite to identify the characteristic intentions that provide that “intentional encasement” to help us identify poetry, or even whether there is a single sort of speech act that poetry is. But it does give a sense of what it would take for poetry to be bullshit. Poetry should be made with some sort of characteristic intention. There is some further attitude that would make it sincere. A poem about loss written by someone who has not experienced loss, and specifically lacks the relevant

³In this sentence I resisted the temptation to add a citation to a famous book I should have read but have not.

⁴In this sentence I resisted the temptation to add a citation to my dissertation.

attitude, may be insincere, but may still be a good poem made with the characteristic intention. But a poem written without considering whether one holds the attitude at all would be bullshit of the second kind. If a poet has millions of followers on Instagram, and profits immensely from a line of merchandise, some readers might suspect that the poetry is written solely out of desire for profit, and not out of any poetic intention, and would constitute bullshit of the first kind.

My account of bullshit predicts that people will *call* such poetry bullshit, but it does not necessarily actually classify the writing of the poetry as bullshit. (Recall, my account is about the actual intention behind the speech act, and not the form or content of the speech act itself, though people will usually characterize the product on the basis of what intention they reasonably believe would have produced this form or content in this context.) If writing with an eye to commercial success is incompatible with the characteristic intention of poetry, then we might have to classify the poetics of writers from William Shakespeare to Cole Porter as at least partially bullshit. But since the commercial success of these artists depends at least in part on an audience having the appropriate reaction to the work, it seems theoretically implausible that this work would actually *be* bullshit of the first kind (though it could be bullshit of the second kind if the author pays no attention to the sincerity condition). In order to actually count as bullshit of the first kind, the profit would have to be intended to come about by some means *other* than a production of the characteristically intended poetic response.

An uncharitable reader might suspect that the response intended to be elicited by a poem that one scrolls past in an Instagram feed cannot possibly be the proper poetic response. But this then depends on a substantive theory of poetic intention. And it is clear that different genres of poetry intend different responses – Gertrude Stein, Alexander Pope, Homer, and Bashō sought reactions of very different scope and scale, and to classify Instagram poetry as bullshit one would have to show that the intended reaction is not just different in kind from all of these, but is in fact no poetic reaction at all. I do not want to take a stand on whether this is substantively correct, but I will give one clear example of bullshit poetry of the first kind. Consider a situation where a handbag designer wants poetry in a foreign language on their handbag, and commissions a poet whose language they cannot read. For several designs, the poet might write an earnest poem, but after eventually realizing the nature of the commission, the poet might just start churning out bullshit, whose purpose is to *look* like poetry, but not actually elicit a poetic response at all. Some Instagram poets might be doing the equivalent, but it seems that their situation is usually more complex, since audiences are at least intended to read some or all of the words.

The historic evolution of different characteristic intentions for different genres of poetry suggests ways that new classes of speech acts might arise out of old. A speech act originally intended to look like one type of speech act, but made without the characteristic intention, starts out as bullshit. But over time, this pattern of bullshit can become recognized, and conventionalized, and identified as a new type of speech act that is now made with its characteristic intention, and thus no longer bullshit. This is very clear with different kinds of questions. I have mentioned how rhetorical questions could have emerged from “just asking questions” kinds of questions. Many forms of politeness involve using conventionalized questions that then get interpreted as polite phrasings of a request (“could you pass the salt?”) or a greeting (“how are you doing?”). Before these were identified as conventions, it is likely that these emerged as a kind of bullshit. Speakers found it too direct to demand the salt, or to directly announce their willingness to engage in a conversation. Instead, they asked a question whose answer they cared about, such that hearers would then be drawn to notice the speaker's desire for salt or interest in a conversation. Eventually, speakers began to bullshit with the question – they asked it in order to get the salt or begin a conversation without actually caring to hear the answer. These days, English speakers

rarely even hear “could you pass the salt?” as a question, but people who are not familiar with the conventionalized nature of “how are you doing?” (or “what’s up?” or “are you ok?” or any of a number of other similar questions that are conventionalized in different dialects of English) might hear it as bullshit of the second kind, once they realize the speaker does not actually care to hear how they are doing, and just wants to start a conversation.

It may be that the kinds of questions asked on an oral exam, or at a trivia competition, also derive from bullshit questions of the second kind, where a speaker asks not to learn the answer, but to learn what the hearer thinks, or is willing to say. But perhaps more interestingly, the *answer* to such a question, when recognized as a test of this sort rather than as bullshit, provides a different kind of speech act that could itself be interpreted as a kind of bullshit. It has been noted even as early as Grice’s original work that giving an answer to a question of this kind does not involve the characteristic Gricean intention of getting the hearer to believe what is said, since the answerer assumes the hearer already knows the answer. I think it is useful to interpret the speech act of answering an exam question as having relevantly different characteristic intentions from the ordinary speech act of making a statement, so that we can avoid characterizing such an answer as bullshit.

However, there still are characteristic intentions and sincerity attitudes involved in the speech act of answering a test question. At least part of this intention involves aiming at the truth. It is unusual for a student or a trivia contestant to do the equivalent of lying, and give an answer they believe to be false. (Though once in a while someone who knows they do not know the answer will write an entertainingly false answer as a kind of joke.) But it is very common for a student to bullshit their way through a test, giving answers whose accuracy they have no idea of, in hopes of getting an acceptable score. Note that this is bullshit of my second kind, being made without a belief of its truth or falsity. Carson (2010, p. 62) points out that such bullshit is very much made with a *concern* for truth or falsity, and with the intention to produce the same sort of response in the intended audience as an earnest answer, and is thus not bullshit of my first kind. It is only the second kind of bullshit, in that it is not backed by a belief in its truth or falsity.

Being good at trivia involves something that sometimes feels a lot like this kind of bullshitting. One tries to figure out an answer that could possibly be right, and fits with the hints that have been dropped in the question, even where one is not that confident that the answer is correct. Just as the bullshit artist and teller of tall tales may convince themselves of some of the embellishments they make up to make the story better, the trivia contestant convinces themselves that something is the answer. But since this is the characteristic mode in a trivia competition, the process does not itself constitute bullshit, though it has similarities to bullshit in the context of a classroom exam or in the context of a storyteller. Perhaps more importantly, unlike when a teller of tall-tales engages in self-deception, in a well-constructed trivia contest (and especially in a crossword puzzle, where the solver can get good evidence that the answer is correct from the crosses), this process is in fact a way to come to have *knowledge* of the truth, even if they began without any belief one way or another.

Fiction can help us imagine scenarios where a speech act that seems like total bullshit to us has become conventionalized, so that the pretense disappears. In the 2013 movie *Her*, the opening scene shows the main character working at his job, for the company [BeautifulHandwritten Letters.com](http://BeautifulHandwrittenLetters.com). He has a pool of clients who commission him to write thoughtful letters to their loved ones, year after year. They supply some photos, and some basic facts and sentiments, but he supplies the details of the wording that make it seem heartfelt. (To heighten the absurdity, we see that he does not actually write the letters by hand – he dictates them to a computer, which prints them in a handwritten style.) To us, it would feel like bullshit if we learned that a loved

one had commissioned an apparently heartfelt letter, or used a chatbot to compose it, rather than actually writing it themselves. But when all parties are aware of what is happening, and we have stripped out the pretense, perhaps we could learn to see this the same way we see singing a Paul McCartney song to express your love for your loved one rather than a song of your own composition. (It would be even more impressive if you commissioned Paul McCartney to write and perform a song expressing your love!)

One difficulty my analysis raises is that, as Cohen puts it, the “intentional encasement” now matters. Whether an utterance constitutes bullshit depends not just on the content of the utterance (as Cohen suggests) or whether the utterance aims at the truth (as Frankfurt suggests), but on whether the utterance was made with the “characteristic intention” of the kind of speech act it constitutes. Some speech act theories identify the speech act of an utterance at least in part in terms of the intention with which it was uttered – depending on the details of such a theory, bullshit in my sense might actually be impossible! For my account of bullshit to work, we need a theory of the characteristic intentions of speech acts that allows for violations of this characteristic intention, such as that provided by Reiland (2020).

But I think it may be sufficient for my account to focus on the connection Frankfurt saw between bullshit and pretentiousness. With the examples I have identified, there is at least a *pretense* that the utterance is patriotically lecturing an audience about American history, or promising to meet up with someone, or warning someone about cancer risk, or asking about someone's day, or citing a relevant academic source, or providing poetic inspiration. What makes the utterance bullshit is either that there is a mismatch between the pretended intention of the act and the actual intention of the act, or an utter lack of care regarding the sincerity conditions that usually justify the characteristic intention.

An advantage I claim for my account of bullshit is that it can naturally extend to acts beyond speech acts. Some speech acts are done by means of behaviors that do not look a lot like speech – clicking a check box on the terms and conditions, or posting a sign provided by the government. But my analysis can go further, to any sort of act for which there are characteristic intentions, and perhaps sincerity conditions.

3 | OTHER BULLSHIT ACTS

One of the few philosophical analyses I have seen of bullshit that does not involve words is that of Leslie Howe (2017). She argues that certain deceptive practices in sport are analogous to bullshit, in particular “simulation” or “taking a dive,” where a player falls to the ground as if injured, hoping that the referee will call a penalty on the other team. However, even though this is a non-verbal act, she analyzes it in its role as a kind of communicative or informative act. The goal is to get the referee to believe that a violation has occurred. Some have thought of this act as like a lie, where the player does this in the knowing absence of a violation, to get the referee to believe that there was a violation. However, Howe notes that sometimes this act is done when a player recognizes that there has been some contact from the opposing player, and is unaware of whether it was a violation or not. Like Frankfurt's bullshitter, this diver is actually often unconcerned with the truth or falsity of the belief they bring about in the referee, and just care about the advantage they get.

I think this is a productive way to apply Frankfurt's original account of bullshit, but it only applies to acts undertaken with the intention of inducing a belief. With the bullshit Prop 65 warning, we have already seen an example of bullshit done out of a desire for compliance rather

than with the intention of inducing a belief. It is true that issuing a warning has a characteristic intention that involves belief, but the bullshit motivation does not. This is actually quite different from the case of taking a dive. In taking a dive, the player imitates the behavior of an injured person, who falls to the ground and cries out in pain. But the real behavior is not best thought of as an *act* at all – when it is real, it happens as a reflex, with *no* intention. By imitating a behavior that is typically done with no intention, the diver is able to induce a belief in another, and if done with no regard to the truth or falsity of the belief, it is in that sense analogous to Frankfurtian bullshit statements.

However, I aim to generalize Frankfurt's account in a different way. One of the central insights of action theory is the idea that what makes something an act, as opposed to a mere behavior, is the fact that it is done with some intention, under some description (Piñeros Glasscock & Tenenbaum, 2023). One type of act is a speech act, done with some characteristic intention or other, and one type of speech act is a statement, made with the intent of getting the hearer to believe what is said, either sincerely, when the speaker believes it, or as a lie, when the speaker disbelieves it. My reading of Frankfurt's characterization of bullshit is a statement made without regard to whether the hearer comes to believe it, or without regard to whether it is actually true or false. Rather than applying the label of bullshit to any act done with the aim of getting someone to believe something that the actor neither believes nor disbelieves, I apply the label of bullshit to any act done with some intention other than the characteristic intention of that act, or without regard to the sincerity conditions of that act.

As an example, consider the Transportation Security Agency. Passengers on all commercial flights that take off in the United States are required to undergo a security screening conducted by this agency, in which carry-on luggage is sent through an x-ray scanner and the passenger walks through a metal detector or millimeter-wave scanner. While this much might seem like reasonable safety precaution for people going into a crowded flying metal tube with several hundred others for a few hours, parts of the process that are often considered particularly galling are a requirement to take off one's shoes before going through the scanner, and a complete ban on carrying any liquids or gels in containers larger than 3 ounces (100 ml). To many, it seems that the characteristic intention of a security screening should be that it meaningfully reduces certain security risks. But many find it implausible that sending shoes through the x-ray rather than the metal detector has any meaningful benefit, or that allowing passengers to bring a dozen different small containers of cosmetics and medicines is actually safer than allowing them to bring a water bottle they are currently drinking from. Bruce Schneier (2003) argues that many of these rules are better understood as “security theater,” whose purpose is not to actually reduce risk, but to rather put on a performance of security. Whether this is intended to make worried infrequent flyers feel safer, or to convince Congress that the agency is doing something useful, if the point is the theater rather than the security, then it is bullshit on my account.

Many of us might have similar reactions to some precautions taken in light of the COVID-19 pandemic. If a restaurant makes a big show of sanitizing the pens that customers use to sign their checks, and requiring customers to wear a face mask while they walk to the table, but allows people to remove their masks while they dine indoors and makes no attempt to open the windows, then it seems that the restaurant is performing “hygiene theater” rather than actually attempting to effectively mitigate risk. This is bullshit on my account. But then again, so is a case where the proprietor actually improves the indoor air quality by upgrading the ventilation system, because they are required to by law, and not because they intend any health benefit. As I discuss at greater length later, the mere fact that something is bullshit on my account does not mean it has no value.

In the comedy *The Producers* (originally a 1967 movie, then a 2001 musical, and then a 2005 movie-musical), the title characters aim to put on a Broadway show that will flop. The idea is that if the show is unsuccessful, then investors will not realize that the shares sold total well over 100%, but if the show was successful, then each investor would want their cut of the profits. With a flop, the producers hope to pocket the difference between the total investment and 100%, but the outcome is an accidental success, and the investors come trying to collect their proceeds. If the characteristic intention of putting on a musical is intending for the audience to enjoy it, or intending for the show to be commercially successful, then their entire plan is bullshit. But if the enjoyment is a sincerity condition, and the characteristic intention is just that the performers will go on stage and create a performance, then this is more akin to a lie or a fraud, than to bullshit. In this case, the concept of bullshit helps us draw some distinctions, but is not quite as clear to apply without a clear sense of what the “characteristic intention” might be.

One challenge to the need for this characteristic intention comes from Wreen (2013). (I thank David Sosa for pressing me on a similar case.) Wreen describes the following case:

Imagine that after years of study I come up with a complicated system for beating the casinos that I sincerely believe is flawless. I travel across the United States lecturing about it to various groups, enthusiastically touting its virtues. In fact, I couldn't be more wrong: the system is seriously defective and contains multiple errors, silly even egregious errors. ... In short, my system is humbug or bullshit, and I've been bullshitting, even if I don't believe I've been bullshitting and certainly didn't intend to bullshit.

Wreen raises this as a challenge for Frankfurt's claim that the intention is relevant to whether the person is bullshitting, given that the intention in this case is earnest.

However, my diagnosis is slightly more subtle. I claim that the *lectures* aren't bullshit, but the *theorizing* is. While the person may have fooled themselves into thinking that they are theorizing earnestly, I claim that at least some of the characteristic intentions involved in theorizing involve checking for errors. A person might *think* that they are checking for errors, but if there really are silly and egregious errors, then it seems likely that the person has not in fact been checking. They have convinced themselves that they are theorizing, but this is bullshit. As long as a person's intentions are not completely transparent, it is possible for a person to do a bullshit job of theorizing without realizing that they are bullshitting. (And if on some level this person does not really believe what they are saying, or if they are lecturing not with the intent that audiences actually believe the theory but instead as an emotive display of distrust of the casino conglomerates, then the lecturing could be bullshit as well.)

One recent extensive theoretical discussion of the concept of bullshit, as applied to activities other than speech acts, is the discussion of “bullshit jobs” by the anthropologist David Graeber, first in a widely shared short popular article (2013) and then in a more theoretical book (2018). His primary characterization of a “bullshit job” is based on whether the job actually creates value for the world, and he spends much of the first chapter working out a definition. He starts with,

Provisional Definition: a bullshit job is a form of employment that is so completely pointless, unnecessary, or pernicious that even the employee cannot justify its existence.

Importantly, part of the reason this definition is provisional is that he wants the category of “bullshit job” to be objective, dependent on actual lack of positive value, and uses the employee's lack of

justification merely as a diagnostic tool. He thinks that people often have an inflated sense of the importance of their own work, so while there may be some people whose job is in fact bullshit but have not recognized it, he thinks it is unlikely that many people will have a job that provides positive value to the world but think it is bullshit. (I will challenge this diagnostic tool later.)

His other modification to this provisional definition brings it closer to the concept I discuss. After noting that it seems inappropriate to consider the job of being a mafia hitman as a “bullshit job,” despite its perniciousness, he notes,

Bullshit jobs are not just jobs that are useless or pernicious; typically, there has to be some degree of pretense and fraud involved as well. The jobholder must feel obliged to pretend that there is, in fact, a good reason why her job exists, even if, privately, she finds such claims ridiculous. There has to be some kind of gap between pretense and reality.

He notes that this criterion is relevant because bullshitting is a kind of deception, and adds a footnote discussing Frankfurt’s distinction between statements that are bullshit and lies:

one is intentional deception, the other, reckless disregard for the truth. I’m not sure the distinction entirely works in this context but I didn’t think entering a debate on the subject would be particularly helpful.

I agree with the tenuousness of the connection to reckless disregard for truth in this particular context. However, my generalization of Frankfurt’s account generalizes this to a disregard for *the characteristic intention* of the act, which *happens* to be about truth in the case of ordinary statements, but can be about other things in the case of other statements, acts, or employment. I thus think there is more of a connection between Frankfurt’s account and Graeber’s than Graeber realized.

Overall then, there is a strong similarity between Graeber’s account of “bullshit jobs,” and my account of bullshit applied to the category of employment. For Graeber, for a job to be bullshit is for there to be a gap between the pretense that it provides positive value to the world, and the reality that it does not. For me, for a job to be bullshit is for there to be a gap between the characteristic intention behind hiring someone in that job, and the actual intention with which they were hired. As I will show, there is some significant and non-accidental overlap in our accounts, but they are importantly different in some ways.

In the second chapter of the book, Graeber gives a typology of five primary categories of bullshit jobs: “flunkies,” “goons,” “duct tapers,” “box tickers,” and “taskmasters.” Flunkies are people who are hired primarily to make someone else feel important – for instance, he notes that some offices have receptionists or doormen that are not really needed for wrangling visitors, but are really there to make the bosses feel important, since visitors need to go through multiple steps to meet them. Goons are people who are hired primarily to make other people do what they do not want – he counts the military, but more importantly people like advertisers, corporate lawyers, and public relations workers. Duct tapers are “employees whose jobs exist only because of a glitch or fault in the organization; who are there to solve a problem that ought not to exist.” Box tickers are people who are hired in order to make it look like something has been done, rather than to actually do something, the way that many diversity officers and post-problem fact-finding teams are. Taskmasters are people whose job is only to assign actual work to others, rather than to do any work themselves.

On my account, Graeber's flunkies and box tickers are engaged in paradigmatic bullshit jobs. Flunkies are the human equivalent of pretentious verbal bullshit – you make it look like you know a lot by spouting meaningless jargon, and you make it look like you are important by hiring a meaningless receptionist. Jargon and receptionists really do important work for some people in some contexts, but here they are not filling their characteristic function, and instead make it look like the person is the kind of person who could use them. Box tickers are the human equivalent of a Prop 65 warning – the government said we have to put up a warning label or hire an inspector, so I guess we will do it, even though I see no point in it myself. (An important point that I will return to later is that on my account these box tickers are *always* bullshit, because they are operating out of obligation rather than out of the characteristic intention, while for Graeber it matters whether their work results in some actual positive value for the world.) Some duct tapers may also be engaged in bullshit, on my account. If the boss has hired his incompetent son as the office manager, and an “assistant,” who is ostensibly supposed to help out, but who actually does all the work that the son cannot do, then on my account *both* the son's job and the assistant's are bullshit. (It appears that Graeber counts the *assistant* as bullshit, because the assistant is usually the one who has the feeling that their job should not exist, but it seems to me that on substantive grounds, the *son* is actually the one who is useless.)

However, on my account, goons and taskmasters are usually not doing bullshit jobs. Graeber's substantive theories of human value suggest that these jobs are pernicious or pointless, but as long as the employer actually has the characteristic intention associated with the job, my theory says it is not bullshit. However, there can be exceptions. Sometimes what appears to be a company hiring a security guard is actually a company paying a local gang some protection money, so that their heavies just stand around by the door all day, rather than robbing the place. This would be bullshit on my account. And sometimes an apparent taskmaster for one level of employees is actually a flunky for someone two levels up. This would also be bullshit on my account.

One other important difference between my classification of bullshit as a gap between pretense and reality of what a job entails, and Graeber's classification as a gap between the claim of doing something valuable and actually being useless (or net negative), is that Graeber allows for a classification of some jobs as “second-order bullshit.” As he notes, a janitor or a plumber is usually doing a real and valuable job. But when the janitor and the plumber maintain the offices of a business that is entirely bullshit, Graeber counts them as bullshit as well. On my account, if the janitor and plumber are hired for the characteristic reasons that a janitor and plumber are hired, their jobs are not bullshit, whether they are serving a useless office or a useful one.

Graeber's use of the employee's opinion of the value of their job as a diagnostic tool disagrees with his substantive theory in the case of these second-order bullshit jobs. An employee could easily think their job is valuable without realizing that they are serving a useless office. But it can go wrong the other direction as well. As Marx noted, when a physical task has been broken down into the component parts that make for an efficient factory process, each part often feels pointless, and workers are alienated from any understanding of the value their work makes possible. In many cases, the feelings of uselessness that Graeber identifies in much modern office work may well be traceable to a similar division of labor that has been applied to knowledge work in recent decades. As Soffia et al. (2021) note, Graeber's methodology may still be important in understanding the human consequences of the work environment, even if it does not properly track his concept of bullshit jobs – “feelings of usefulness at work are not a direct indication of the social value of that work but are tied to the degree to which the social relations under which that job is undertaken enable individuals to realize their human potential.” On my account, this

diagnostic tool is not relevant to the concept of bullshit, but we instead need an account of the characteristic intention for any job.

One important feature of Graeber's account not shared by mine is that for Graeber, being bullshit is tantamount to a particular negative moral evaluation of a job – it does nothing of value for the world. I keep a more evaluatively neutral conception of “bullshit,” though like Frankfurt, I can show why bullshit is *often* pernicious. If the characteristic intention of a job is positive, then anyone who is hired into it without the intention of doing something positive is doing a bullshit job on my account. Conversely, if the uncharacteristic intentions with which an employer hires someone for a job are of no value for the world, then anyone doing that bullshit work is doing something of no value. But we have seen that these conditions are not always satisfied. If goons and taskmasters are characteristically hired for reasons of no positive value, then they are not bullshit on my account despite satisfying this negative evaluation. Compare: even though a pretentious statement might not be bullshit on Frankfurt's account if it is honest and intended to be understood, it might have no more value than pretentious bullshit. Conversely if the uncharacteristic intention with which someone is hired actually results in positive value, then the job can be bullshit in my sense while still producing positive value in the world. I claim that this is very often the case for Graeber's box tickers, and box-ticking bullshit of all sorts.

One discussion Graeber quotes of a box-ticking bullshit job is from a permitting consultant for a construction firm. She lists many kinds of reports that she is required to write up and submit in order to get permission for a construction project, on topics like environmental impact, landscape impact, sunlight analysis, tree impact assessment, flood risk assessment, and so on. She notes, “Each report is about 50 to 100 pages, and yet the strange thing is, the resulting buildings are ugly boxes remarkably similar to the ones we built in the sixties, so I don't think the reports are serving any purpose!” It is definitely possible that these reports serve no useful purpose (and if the reports end up discouraging dense walkable developments in urban areas while allowing sprawling car-oriented developments by the highway, they may even be net negative). But the look of the buildings and the opinions of the employee are not actually a good diagnostic for this.

If, say, the flood risk assessment is something that no one on the construction team cares about, and no user of the building will complain about on 99% of days, and no one at the city permitting office cares about except to tick the box of having done the assessment and ensuring that the building plan fits the legal criteria, it can be bullshit by all the parties involved according to my account. But it can still be of positive value for the world! If the regulations were written well, then denying permission to building projects that do not fit the flood risk rules while permitting the ones that do can be an excellent way for a city to, over time, gradually reduce the amount of dangerous runoff that fills the streets whenever there is a heavy rainstorm, and thus reduce the property damage and loss of life that occurs. No one involved may have any interest in or awareness of this effect, but the box-ticking ritual can result in good anyway.

This occurs with many of the examples I gave of bullshit speech acts as well. Prop 65 warnings may be made merely from a grudging desire to comply with the law, but they may nevertheless actually help some members of the public minimize their risk of exposure to harmful chemicals (particularly with the newer rules that help the public understand which chemicals are relevant at each location). Sometimes the citation that the referee insisted an author include is actually valuable for the reader, even if the author thinks that it is bullshit. James McAuley and Harold Stewart wrote bullshit poetry under the name “Ern Malley” in an attempt to discredit their rival Max Harris for publishing gibberish, but ended up producing work that later generations have

found more worthy of study than the work produced under their own names (I thank Ray Briggs for this example).

And I think this can happen with bullshit statements of Frankfurt's original sort too. Consider Jennifer Lackey's "creationist teacher" example:

Suppose that a Catholic elementary school requires that all teachers include sections on evolutionary theory in their science classes and that the teachers conceal their own personal beliefs regarding this subject-matter. Mrs Smith, a teacher at the school in question, goes to the library, researches this literature from reliable sources, and on this basis develops a set of reliable lecture notes from which she will teach the material to her students. Despite this, however, Mrs Smith is herself a devout creationist and hence does not believe that evolutionary theory is true, but she none the less follows the requirement to teach the theory to her students. Now assuming that evolutionary theory is true, in this case it seems reasonable to assume that Mrs Smith's students can come to have knowledge via her testimony.

(Lackey, 1999, p. 477)

As written, this case is clearly not an example of bullshit – Mrs Smith teaches, and thus presumably satisfies the characteristic intention of making a statement (i.e., intending for the students to believe what she says on the basis that she said it), but disbelieves what she says, and is thus not bullshitting but lying. However, we could very easily modify the example to turn it into an example of box-ticking bullshit.

Suppose that an elementary school requires that all teachers include sections on evolutionary theory in their science classes and that the teachers conceal their own personal beliefs regarding this subject matter. Rather than trust the teachers to develop their own lesson plans, the school develops a set of reliable lecture notes and requires teachers to rehearse and use these notes with apparent sincerity. Mrs Smith is an utterly incurious person and has never paid particularly close attention to the lecture notes provided by the school or formed opinions about the matters contained in them, but she is a skilled actor and convincingly teaches any lesson plan she is given. Assuming that evolutionary theory is true, in this case it seems reasonable to assume that Mrs Smith's students can come to have knowledge via her testimony.

In this case, assuming the school has done its job of developing lecture notes well, and assuming Mrs Smith has done her job of performing those notes well, the students end up with knowledge, even though Mrs Smith was bullshitting the entire time, just doing her job rather than considering whether what she was saying is true or false.

In many of these cases of box-ticking bullshit, the end result might have been better if it had not been bullshit, that is if everyone involved had been motivated to *actually* care about the activity they are *apparently* engaged in. But given a large and complex society, where tasks are broken up among dozens, hundreds, or even millions of individuals, it seems that some sort of Marxian alienation is inevitable. Not everyone can know all the purposes of all the things they are doing as parts of large organizations, so not everyone will actually care about everything that actually matters in their jobs or as members of society. Various levels of policy that require people to do their part anyway will lead to lots of bullshit, but if these policies are designed well, this bullshit

can sometimes be valuable nonetheless. The work may be bullshit from the point of view of the worker, but if we think of the work as being done indirectly, by the people who set the rules requiring this work to be done, then it is no longer bullshit.

There are surely many harms of bullshit. Frankfurt writes about how “bullshit is a greater enemy of truth than lies are,” because the liar at least must respect the role of truth in order to oppose it, while the bullshitter can just ignore it. The bullshit promiser or warner or receptionist may similarly come to lack the respect for promising or warning or being a receptionist that the insincere one has. Erik Olsson (2008) argues, using formal models from Hegselmann and Krause (2006), that some amount of bullshit is compatible with society as a whole maintaining its connection to the truth, though high enough levels would cause fundamental problems. Similar results seem plausible for the characteristic intentions of other acts.

Similarly, Graeber writes about the “psychic wound running across our society,” of people performing tasks that they secretly believe do not need to be performed. This is akin to the harm that Marx observed of alienation from the product of one’s labor. Soffia et al. (2021) use a range of quantitative measures to argue that while Graeber is right to observe that there is a harm here, he is wrong in his quantitative estimates of it. (He estimates that the prevalence is 37% and rising while they estimate that it is 9% and falling. He estimates that it is most widespread in the Anglosphere while they find it is more widespread in Eastern Europe.) This “psychic wound” is surely harmful to the individual in ways beyond the harms Frankfurt observes where people come to lose a belief in the relevance of truth (or whatever the task is). But I think that for understanding this harm, it is more helpful to have a concept focused on the mismatch between the characteristic intention and the real intention with which something is done, rather than a concept focused on actual lack of value of the work. This harm is real even if the work has actual value, and even if the value of the work means that it is better to bullshit this work than let it go un-done.

REFERENCES

- Åkerman, J. (2009). A plea for pragmatics. *Synthese*, 170(1), 155–167.
- Austin, J. L. (1975). *How to do things with words*. Clarendon Press.
- Austin, J. L. (1979). Performative utterances. In J. O. Urmson & G. J. Warnock (Eds.), *Philosophical papers* (pp. 233–252). Oxford University Press.
- Buekens, F., & Boudry, M. (2015). The dark side of the loon. Explaining the temptations of obscurantism. *Theoria*, 81(2), 126–142.
- Carson, T. (2010). *Lying and deception: Theory and practice*. Oxford University Press.
- Cohen, G. A. (2002). Deeper into bullshit. In S. Buss & L. Overton (Eds.), *Contours of agency: Essays on themes from Harry Frankfurt* (pp. 321–339). MIT Press.
- Easwaran, K. (2009). Probabilistic proofs and transferability. *Philosophia Mathematica*, 17(3), 341–362.
- Frankfurt, H. (1986). On bullshit. *Raritan*, 6(2), 81–100.
- Graeber, D. (2013). On the phenomenon of bullshit jobs: A work rant. *Strike Magazine*, 3, 10–11.
- Graeber, D. (2018). *Bullshit jobs: A theory*. Penguin Random House UK.
- Grice, H. P. (1957). Meaning. *The Philosophical Review*, 66, 377–388.
- Grice, H. P. (1975). Logic and conversation. In D. Davidson & G. Harman (Eds.), *The logic of grammar, chapter* (pp. 64–75). Dickenson.
- Habib, A. (2022). *Promises*. Stanford Encyclopedia of Philosophy.
- Hegselmann, R., & Krause, U. (2006). Truth and cognitive division of labour: First steps towards a computer aided social epistemology. *Journal of Artificial Societies and Social Simulation*, 9(3), 1–33.
- Howe, L. (2017). Simulation, seduction, and bullshit: Cooperative and destructive misleading. *Journal of the Philosophy of Sport*, 44(3), 300–314.
- Lackey, J. (1999). Testimonial knowledge and transmission. *The Philosophical Quarterly*, 49(197), 471–490.
- Mahon, J. E. (2015). *The definition of lying and deception*. Stanford Encyclopedia of Philosophy.

- Nowak, E. (2019). Multiculturalism, autonomy, and language preservation. *Ergo*, 6(11), 303–333.
- Olsson, E. J. (2008). Knowledge, truth, and bullshit: Reflections on Frankfurt. *Midwest Studies in Philosophy*, 32, 94–110.
- Piñeros Glasscock, J. S., & Tenenbaum, S. (2023). *Action*. Stanford Encyclopedia of Philosophy.
- Reiland, I. (2020). Constitutive rules: Games, language, and assertion. *Philosophy and Phenomenological Research*, 100, 136–169.
- Schiffer, S. (1972). *Meaning*. Clarendon Press.
- Schneier, B. (2003). *Beyond fear: Thinking sensibly about security in an uncertain world*. Springer-Verlag.
- Searle, J. (1965). What is a speech act? In M. Black (Ed.), *Philosophy in America* (pp. 221–239). Allen and Unwin.
- Soffia, M., Wood, A. J., & Burchell, B. (2021). *Alienation is not 'bullshit': An empirical critique of Graeber's theory of BS jobs*. *Work, Employment and Society*.
- Sorensen, R. (2010). Knowledge-Lies. *Analysis*, 70(4), 608–615.
- Stokke, A., & Fallis, D. (2017). Bullshitting, lying, and indifference toward truth. *Ergo*, 4(10), 277–309.
- Tosi, J., & Warmke, B. (2016). Moral grandstanding. *Philosophy and Public Affairs*, 44(3), 197–217.
- Whitcomb, D. (2023). Bullshit questions. *Analysis*, 83, 299–304.
- Wreen, M. (2013). A P.S. on B.S.: Some remarks on humbug and bullshit. *Metaphilosophy*, 44(1–2), 105–115.

How to cite this article: Easwaran, K. (2023). Bullshit activities. *Analytic Philosophy*, 00, 1–23. <https://doi.org/10.1111/phib.12328>