Incitement: Some questions and problems from a philosophical perspective

Lloyd Carr
Philosophy (Emeritus)
Rivier University
Summer, 2021

1. We often use language to influence other people. While not all language use and perhaps not the most central part of language use is to influence others, a large part goes to getting others to believe certain things, or to experience certain emotions, or to behave in certain ways. Some of this language use can be judged good: we try to get others to believe what we believe is true, to do what we believe is right, or to experience what we believe are virtuous emotions. But some language use can reasonably be judged wrong: we try to get others to believe lies, to act in criminal ways, or to experience vicious emotions. One interesting case of language use falling within this broad category of "wrongful influence" is incitement – someone, the inciter, communicates with a target audience, the incitees, in the attempt to get them to commit wrongful actions. Incitement is interesting (in at least one sense) because it turns out to be unexpectedly difficult to understand. It is a crime and in this respect the legal world has and is continually subjecting incitement to legal scrutiny, test cases, and refinement. My interest in incitement, however, is philosophical; specifically, (i) how might we conceptualize the relation between inciter and incitee, and (ii) on what ethical grounds do we judge the inciter and incitement to be morally wrong, independent of its illegality?[1]

2. The causal model

At first view, incitement seems to be an instance of causality: the utterances[2] of one language user causes certain actions to be performed by those who hear the language, and these actions are criminal and typically morally wrong. In this view, because the language user *causes* others to commit a wrongful action, incitement is itself a wrongful use of language.[3] The general structure is: A causes B to do wrong and the wrongness of B's action backtracks to A. Applying this to the case of incitement, the inciter (A) is taken to bear a measure of the moral blame and legal culpability for the wrongful actions of the incitees (B), having caused them. The (consequentialist) moral principle at work here is: to cause a wrong is itself wrong.[4]

But upon scrutiny, this causal model of incitement has problems. First, it does not seem possible for one person's language directly to cause another person to act. If, for example, the target audience didn't hear the inciter's words, or if they weren't paying attention, or if in the

extreme case the target was dead, the inciter's words alone could not directly cause the target to act; the target audience must meet certain conditions (e.g. be alive, understand the language used, listening, etc.) in order for the inciter's words to have any effect. Second, the causal connection between inciter and incitees appears problematic. If, on the one hand, the causal connection from inciter to incitees is sufficiently *strong*, then it would seem that it better fits the case of coercion rather than incitement. For example, if someone uttered threats that forced (i.e. strongly caused) someone else to do a wrongful action, then the former would be guilty of coercion while the latter acted under coercive conditions and would typically not be considered morally blameworthy. If, on the other hand, the causal link is *weak* such that the incitees retain their autonomy, then it would seem that the inciter can't be blamed for the wrongful actions done by the incitees; the latter acted completely on their own.

Perhaps the causal connection from inciter to incitees shouldn't be thought of as direct, having only one link as it were: strong or weak. We might try to envision incitement as requiring a complex causal connection with several steps. First, the inciter utters language intended to cause certain beliefs and emotions in the incitees. (The inciter might but need not hold these beliefs or experience these emotions.) Next, once the incitees accepts the beliefs and experiences the emotions, this mental state causes the incitees to commit wrongful actions. The recent example that seems to fit this model is that of former President Trump's impeachment for inciting the crime of insurrection. Trump addressed his supporters and convinced them that his loss of the 2020 Presidential Election was actually a case of theft; his audience accepted the belief that a Second Term Presidency was stolen from Trump. Trump then used language intended to cause anger in his listeners that the election was stolen from someone (Trump) they supported. Once their mental state was sufficiently strong, it caused his listeners to act criminally: they violently attacked the Capitol Building trying to stop the Senate process for officially making Biden US President.

This multiple-step causal connection between inciter and incitees seems more realistic than the single step of direct causality, but it also introduces problems of a different sort. If incitement is the cause of a cause of a wrong, then if one causal step fails do we still have incitement? Here are three cases. (1) The inciter succeeds in causing the beliefs and emotions to arise in the incitees, but this mental state fails to cause the incitees to commit any wrongful action (say, the sudden presence of a police force keeps the incitees from acting). There was no wrong done on the part of incitees; do we still have the crime of incitement? Not in this (consequentialist) causal model, yet the inciter seems to have acted wrongfully. (2) The incitees already are in a mental state sufficient to cause them to commit wrongful actions, and they do so, but the inciter's words fail to cause or to increase this mental state (say, the inciter was not a very powerful speaker and the inciteful words didn't influence in any way the already heated mental state of the incitees). The inciter's words had no influence on the incitees' actions; do we still

have the crime of incitement? Not in this (consequentialist) causal model, yet the inciter seems to have acted wrongfully. (3) A third way causality might fail is the case of a mismatch between the actions the inciter intends to cause the incitees to commit, and the actual wrongful actions the incitees end up committing. Suppose, for example, someone tries to incite a mob to close down a store (say the store is selling Nazi insignias). The mob reacts to the inciter's words by forming the appropriate negative beliefs and emotions, but then they do not close down the store; their mental states cause them to burn the store down instead. The inciter never meant to incite arson; do we have the crime of incitement and if so, incitement to what? Again, this (consequentialist) causal model confronts a problem; where cause and effect don't line up, incitement becomes at least ambiguous, and yet the inciter seems to have acted wrongfully.[5]

If incitement is defined causally, then if the causal connection running from inciter to incitees fails as in cases (1) - no effect means no cause, (2) - no cause means no effect, and (3) - cause and effect don't line up, we seem to lose any consequentialist grounds for holding the *inciter* morally responsible (as well as criminally culpable) for incitement, though the *incitees* can be held morally responsible for the wrong they do. In this causal model, the inciter's language appears to be neither causally necessary nor sufficient for the incitees to commit wrongful actions.

Perhaps we don't have the right kind of causal connection in mind between inciter and incitees to understand in what way incitement is wrong. A basic distinction made in the philosophy of language is between: (a) the perceivable physical marks or sounds that must occur in language use, and (b) the meaning or semantic content conveyed by means of those marks or sounds.[6] The inciter, in using language, will produce physical marks or sounds (e.g. will pronounce words, will write sentences, will use colored images, etc.) that cause, in a complex causal chain with many steps, certain brain events to take place in the incitees who hear or read the inciter's words. These physical events running from the inciter's body movements (neural brain activity, speech organs events, writing mussel movements, etc.) to incitees' body movements (physiological ear or eye stimulation, neural activity, brain events, etc.) "support'' a level of meaning running from the inciter's cognitive state to the incitees' cognitive state of understanding what the inciter has said or written. The exact relation between the semantic and the physical levels of language use is a contested question, but it will be convenient for my purposes here to employ the metaphysical concept of supervenience: we will say that in communicative language use, and thus in the case of incitement, the semantic level *supervenes* on the physical level.[7]

If, on the physical level, the connection between inciter and incitees is factual causation, as it surely is, what about the supervenient (semantic) level of meaning? Certain cognitive events in the mind of the inciter clearly "influence" (through language use) cognitive and emotive events

in the minds of the incitees who understand what the inciter is saying, but is this "influence" itself causal in the way that the subvenient physical level is causal? And if not, is it nevertheless a sufficiently strong connection to make the inciter morally blameworthy, by consequentialist principles, for contributing to the wrong done by the incitees to their victims?

I believe that the mental "influence" from inciter to incitees is not causal in the way the subvenient (physical) level of inciteful language use is factually causal. The meanings expressed by the inciter are intended (by the inciter) to have a *psychological* effect on the target audience and this mental effect is further intended to *motivate* action on the part of the incitees. Let's call this type of supervenient causality "*semantic*," i.e. causality by way of the expression and understanding of meanings. The inciter wants to alter the mental state of the incitees by using language in a way intended to bring about this change. This typically includes the following: (i) giving the target audience a set of beliefs or strengthening such beliefs if the target already accepts them, (ii) arousing certain hostile/heated emotions in the incitees, (iii) providing the incitees a sort of rationale, a kind of permission and encouragement to act in wrongful ways. These three "effects" together serve the inciter as a strategy of *persuasion* and the incitees are the target being persuaded by this strategy.

One important part of this inciteful language use is to convince the incitees that the wrong they are being incited to do is not really wrong but justified from some alternative perspective, some "bigger picture" revealed by the inciter. By working to remove moral/legal inhibitions in the incitees, the inciter frees them to do things that now appear to them as "justified" by a different standard. For example, an inciter might try to incite a mob to murder someone (A) by presenting the action not as murder but as a form of delivering swift justice, giving A only what he deserves. The mob forms the belief that they wouldn't be committing murder; they would be doing a good thing and thus be justified in killing A. In the heat of, say, anger, vengeance or hatred, the mob murders A having been "permitted" this action by the rationale provided by the inciter. In this typical case of incitement, depicted in numerous films and novels, we find the three alterations in the target's mental state listed above: belief, emotion, and permission.

Granted that the "causal" connection between inciter and incitees is one of (supervenient) semantic causation, does this advance the attempt to understand the connection between inciter and incitees such that the inciter bears moral blame for the wrongs committed by the incitees? It seems not. One problem is this: semantic causation allows incitees their autonomy and their agency. Even though semantically caused by the inciter, it is still *their* beliefs which they are free to doubt or reject (but don't), it is still *their* heated emotions which they are free to control (but don't), and the incitees must still permit *themselves* to act in wrongful ways (which they do by accepting the inciter's alternative rationale). Semantic causation, in effect, seems to blur the distinction between incitement and persuasion, between inciter and

motivational speaker. There ought to be a difference between A *persuading* B to do something wrong, and A *inciting* B to do wrong. Semantic causation, however, appears not to allow for such a distinction by allowing incitees their autonomy; the semantic-causal link from inciter to incitees is too "thin." Thus, by consequentialist standards there is a strong sense in which the inciter can be presented as blameless for the wrongs done by the incitees, who after all acted it would seem of their own accord.

Another problem for this variation of the causal model is that when there is no effect, it is hard if not impossible to hold a cause morally responsible. An inciter might craft language to cause semantically the three mental effects (beliefs, emotions, permission) in the target audience but these mental effects might not happen. For example, the target audience might have a contravening psychological state (say, deeply held religious beliefs) that allows them to resist the inciter's language. Or, if the psychological effect does happen in the incitees as semantically caused by the inciter's language, it might not issue in wrongful action. For example, having been aroused to the point of action the incitees might have a last moment change of heart that overrides their psychological state or perhaps the presence of police might strengthen their inhibitions and keep them from acting. If this were the case, by the structure of the (semantic) causal model, incitement would not have occurred. Yet the inciter still seems to have done wrong by the very act of incitement, even though it didn't result in wrongful actions on the part of the incitees.

Whether the model is one of factual or of (supervenient) semantic causality, the basic consequentialist moral principle set forth above (to cause a wrong is itself wrong) leads to the conclusion: *if a wrong is not caused then there is no wrong done*. To apply this principle to incitement, when we try to understand incitement as some kind of causality, if the inciter's language has no consequences (i.e. effects) then it seems that we lose the consequentialist grounds for holding the inciter morally blameworthy for incitement. Yet it seems in such cases that the inciter has, nevertheless, done something wrong. Let's, then, try a different approach.

3. The subjective model

We might try to define "incitement" as essentially subjective: a language user (S) engages in linguistic behavior such that (a) S *believes* certain utterances will motivate a target audience to act in ways S knows to be criminal, (b) S *intends* the utterances to motivate the target audience to commit such actions, and (c) S has a *reasonable expectation* that the target audience will act in the ways S intends in response to S's utterances.[8] Incitement is defined as a three-part mental state of the inciter: a belief, an intention, and an expectation. What is noteworthy in this model is that there need be no objective causal chain running from the inciter to the

incitee. So, for example, an inciter can be in the 3-fold mental state (a), (b), (c) and it might be impossible for the target audience to commit the illegal actions; say unbeknownst to the inciter the target audience have all died. Nevertheless, the inciter would be subject to moral judgment and presumably legal prosecution for incitement; the wrong of the inciter is *independent* from the wrong of the incitees.[9]

As with the causal model, this model of incitement in terms of mental contents has problems. One large problem is subjective privacy; it is impossible for anyone directly to know the mental state of another person. One's own mind is a private world. So, for incitement (in this model) to be subject to moral judgment, not to mention a prosecutable crime, there must be objective evidence that the defining subjective state exists in the mind of the inciter. The only possible objective evidence is the inciter's words, either at the time of incitement or in the inciter's past; if someone never uttered or written anything, it would not be possible to accuse the person of incitement for there would be no objective evidence on which to base the accusation. So, given objective evidence, there will have to be an inference from this evidence (i.e., the inciter's words) to the presence of the requisite mental state in the mind of the inciter prior to moral judgement.

And here lies the problem. Someone could utter inciteful words and yet not be in the mental state defining incitement. A common example is found in all the films and plays that contain scenes in which an actor, playing a character, incites a mob to commit illegal actions. The inciteful words are uttered, but the actor playing the part of the inciter isn't in the actual mental state that would define incitement; the actor is only acting. Yet the audience of the play or the viewers of the film might be (and historically have been) incited by the inciteful words of the actor actually to commit wrongful actions. Here we have the complete absence of the defining mental state; do we have a case of incitement? It seems so, yet we can't within this subjective definition of incitement. By deontological standards, we lose the grounds to judge anyone other than the incitees guilty of wrongful actions.

Another problem with this subjective concept of incitement is that it makes it impossible for a (ro)bot to incite, yet intuitively this clearly seems possible. Imagine a (ro)bot operating with a sophisticated AI learning algorithm; it could easily learn to utter words that incites listeners to commit wrongful actions, yet the (ro)bot would not have the "mental state" that subjectively defines incitement, for the (ro)bot would lack what we mean by "subjectivity." Again, we can turn to motion pictures for examples. Take the film "Her" (2013) in which a lonely person in the future falls in love with the "woman" behind a voice he communicates with, but the "woman's" voice is actually an AI system designed to offer pleasing responses to lonely men. We can easily picture this AI program inciting listeners to do wrongful things; yet by the subjective definition of incitement, this couldn't be a case of incitement.

We need not turn to movies or our imagination for a realistic example. In the 2016 US Presidential Election, Russian bots spread false information to target consumers of news from certain websites. The false information on these sites was believed by many and it so angered some in the intended audience that at least one committed the illegal action of attacking a Washington, DC pizza house with firearms. This seems to be a clear case of incitement, yet the bots involved completely lacked subjectivity. To the argument that the true inciters in this case were the people who "programed" the bots there is the response that bots operating with AI learning algorithms function "autonomously" and "learn" which words and pictures affect consumers in the desired way; the more they learn, the better they get at generating wrongful behavior in the target audience.

The subjective model of incitement was examined here to see if it could avoid the problems that arose for the causal model. For the causal model, if the causal relation from inciter to incitees fails, then we are not able to apply consequentialist moral theory to judge incitement wrong; consequentialism requires an act to have consequences, and thus requires the inciter to cause an effect in a target audience. The subjective model succeeds in avoiding this problem, but it appears to have a fatal flaw of its own: namely, without the existence of the right mental content we can't use deontological moral theory to judge incitement wrong; it appears that we have cases of incitement without the subjective state deontology requires. So, once again, it looks like we lose the grounds (at least in these cases) for incitement to be judged morally wrong.

Both models of incitement, it would seem, fail to capture cases that appear to be acts of incitement, and thus they also fail to support judgments that the would-be inciter has done something morally wrong (at least in these cases). For incitement to be morally wrong, one would think that where the right subjective conditions are absent, there would then be causal conditions in place on which to base a moral evaluation by consequentialist principles; and where the right causal conditions are absent, one would think that the right subjective conditions would be present on which to base a moral evaluation by deontological principles. Such is not the case, combining the causal and the subjective models of incitement still leaves us with the same gaps in moral evaluation. We appear to be driven to the conclusion that, while incitement is legally a crime, neither of these two major ethical theories provides a way by which we can understand incitement as something that is morally wrong. So, let's try another approach.

4. The cultural model

Perhaps incitement is not (simply) a relation between an inciter and a target audience, a relation that only occurs when someone initiates an inciteful use of language. Instead, we might try thinking of incitement as a social event that draws upon and is conditioned by certain cultural practices and tradition that overall contributes to forming a population's way of life. Two examples might help clarify what I have in mind. When we use the phrase "a climate of distrust" to describe the ongoing negotiations between, say, a labor union and management, we are not referring to an actual instance of overt distrust on anyone's part. What we mean is something more nebulous, namely, that the negotiations are being carried out against a background of attitudes that make it likely the negotiations will fail. We attempt to capture by "a climate of distrust" a context that increases the probability of failure; it describes something about the history and the current milieu of the negotiations that anticipate and often eases the way for them to break down.

We use the phrase "systemic racism" to characterize the ongoing struggles of a racial minority, not to refer to any specific instance of racism. "Systemic racism" doesn't describe a one-time event; rather it describes something hard to pin down - an ambient, diffuse cultural context, a milieu and a tradition of attitudes, which contribute to making certain racial practices and interactions of a population the social norm. Systemic racism primes, as it were, individuals and institutions within that culture to accept and partake in racist attitudes, policies, and behaviors.

Generalizing, we can say that the culture in which an individual grows and develops will impart certain ways of seeing their world, of judging events and other people, and inculcates a style of reacting to challenges and opportunities. Much of this is not recognized as such, it is simply "our world," "the way we live," and "the way things are." A typical structure that develops in cultures is that sub-groups see themselves as dividing up into "us" and "them," into "we" and "they/others." Biases, prejudices and often hostilities target "them" and "others," especially if "us" and "we" belong to a relatively homogeneous dominant ethnicity within the population and the "others" belong to a minority group. For example, European cultures have a well-known history of anti-Semitism and anti-Roma prejudices. In the United States there is a tradition of anti-immigrant (or anti-foreigner) bias within the primary (European) ethnicity: anti-African since the founding of the country, anti-Asian and anti-Irish in the 19th Century, and anti-Southern and anti-Eastern European immigrants in the early 20th Century. As in the case of the two examples described above ("climate of distrust" and "systemic racism"), such cultural conditions tend to desensitize a population to certain wrongs; they prepare the way, normalize, and increase the probability that one group perceives another with suspicion, animosity and even a target of hostility.

Incitement seems to fit this cultural model; perhaps we might view incitement as a social interaction against a cultural background, an *ethos*, which increases its probability. In this view,

incitement would not be socially possible unless the incitees already have a propensity, were culturally "primed" as it were, to do wrongful actions to their victim. The inciter acts as a "trigger," an actualizer, of a pre-existing propensity for the incitees to act in wrongful ways, but the inciter is not the cause. The following two analogies might help us picture the role of inciter in this cultural model of incitement.[10]

We ordinarily say that the probability is zero of an unflipped coin landing heads up. Imagine such a coin is fair, and now someone flips the coin and suppose it lands heads up; has the one who flipped it *caused* the coin to land heads up? Clearly not, that's a chance event. But the agent did increase the probability of the coin's landing heads up from 0 to .5 by flipping it.

Another analogy: we would typically say that the probability is zero of a car not being driven to be directly involved in a moving traffic accident. Now someone drives the car and it gets into a moving traffic accident, but the driver is in no way at fault (say another driver rear-ends the car while it is sitting at a traffic light). This traffic accident has a certain probability ($1 > n > 0$) of happening. While the driver did not *cause* the accident, the driver did increase the probability from 0 to .n that the traffic accident would happen by driving the car.

Within the cultural model of incitement, the ethos of animosity makes for a cultural background that eases the way for the wrongful actions on the part of the incitees towards the person or property of their victims. The inciter's language, then, is like the agent who flips the coin or the agent who drives of the car; the inciter does not *cause* the incitees to commit wrongful actions (as the causal model would have it), and the inciter need not even have any intentional subjective content (as the subjective model would have it), but rather increases their probability. The inciter functions opportunistically, as it were, taking advantage of a propensity, a preexisting background of cultural hostility toward the incitees' victims.

The cultural model gives us, I believe, a better way of understanding incitement than either the causal or the subjective models. It is hard to understand an instance of incitement as an interaction that happens in isolation from a social context or a cultural ethos. Both actual and fictional cases of incitement seem to occur against a background of animosity toward the incitees' victims; they occur within a social history that prepares the way for them to "erupt."[11] In addition, the cultural model offers an improvement in so far as it avoids the counterexamples that present problems for the causal model and the subjective model of incitement (as described above). The problems with the causal model don't apply in the cultural model because the relation between the inciter and the incitees is not causal in this model. Also, the problems with the subjective model don't apply to the cultural model because the inciter need not be in any specified mental state or even have any subjectivity at all; we can understand how an actor playing a part but lacking the subjective content that defines incitement, or a bot

that lacks subjectivity altogether, can function as an inciter. But does the cultural model advance our understanding why incitement is morally wrong? It would seem that it does not.

For one, we lose the ability see the inciter and incitees as full moral agents. Both inciter and incitees, within this model, are equally subject to the cultural forces that make a sub-group into a "them" and an "other," and prepares the way for their victimization. In this respect, the cultural model decreases both the inciter's and the incitees' individual agency to the degree that their cultural tradition plays a role in incitement. And, importantly, to the degree that agency is reduced, moral blameworthiness is equally reduced and becomes shifted to and diffusely absorbed by the cultural.

Another problem for this model is that it seems to let the inciter off the hook when compared to the wrongful actions of the incitees. If the cultural background of the incitees "primed" them to do a moral wrong to their victims, then the inciter's language seems to have only a minimal role in inciting them to action and incitement itself either would not be considered a specific moral wrong, or perhaps only a "lesser" moral wrong, compared to what the incitees do. The cultural model allows the inciter's language to be a mere excuse for the incitees to do what they all along were primed to do, given their culture, to their victims. Moral judgement tends to focus on the wrong done by the incitees to their victims and away from the wrong of incitement. In this model, it becomes increasingly difficult to see what justifies accusing the *inciter* of moral wrongdoing in incitement.[12]

Perhaps the biggest problem confronting the cultural model of incitement centers around the question: from what ethical point of view is it possible to judge that incitement is wrong? Both act consequentialism and deontology take an agent perspective in evaluating the wrongfulness of an act or an intention to do one's duty, but the cultural model of incitement is not agent centered and even downplays the role of individual agents. These two ethical theories appear to be a poor fit with this model. Rule consequentialism (RC) and virtue ethics (VE) might each initially look promising in so far as each allows for the role of culture in our moral evaluations; culture in the form of social norms (moral rules) in the case of RC, and culture in the form of character formation, education, and social practices in building virtuous habits in the case of VE. But RC is still primarily agent centered in requiring an agent freely to select the moral rules that function as action-guiding. Also, forming a specific rule making incitement impermissible appears ad hoc, making up a cultural norm to "fit the bill." VE likewise remains agent centered once the agent's virtuous actions become second nature. An additional problem for VE is that it seems perfectly possible that an inciter in the act of incitement is exercising a host of virtuous habits (e.g., courage, leadership, loyalty, civic pride, honesty, etc.). Within the cultural model of incitement, then, neither act or rule consequentialism, nor deontology, nor virtue ethics gives us an ethical basis that would justify evaluating incitement and the inciter as morally wrong.

The cultural model of incitement would seem to require a communitarian ethical theory as the basis for its moral evaluation. However, while communitarianism has produced a political/social philosophy and a philosophy of the self, little work so far has gone into developing an ethics.[13]

In the absence of a communitarian ethics, we might consider relativism to be the best available ethical view point by which we might judge incitement, specifically moral relativism (MR). There are many versions of moral (or ethical) relativism, and many distinctions within each, but for my purposes here it is sufficient to characterize MR as making two central claims, one negative and one positive. (1) In making moral evaluations it is not possible to leave one's cultural perspective; there is no standard "universal" point of view, no "view from nowhere," no "neutral" perspective. (2) Humanity is multicultural, and cultures vary in their norms; a population's moral system is part of that population's cultural norms, and all moral evaluations are made from within the perspective of one's culture. A contrasting ethical point of view is that of moral absolutism or universalism (MA). Again, there are several versions of MA and important distinctions within each, but it will be sufficient here to characterize it by two central claims, each the negation of those characterizing MR: (1) In applying moral standards, it is possible to make moral evaluations from a universal point of view, a "view from nowhere," a perspective that is "neutral." (2) There is a core of moral values and norms that are universal; that is, they are transcultural, and thus are independent of any population's social history or culture.[14]

Since the time of the ancient Greek philosophers the relativism/absolutism debate has posed challenges for arriving at objective criteria in making moral evaluations, and represents a deep divide in ethical theory. As characterized here, it is impossible to accept both; to accept one requires the denial the other. Our question, then, is: from what point of view are we to judge that incitement is morally wrong and that the inciter is morally blameworthy? It seems that the cultural model of incitement continues to have this problem: from the perspective of both MR and MA it is hard to justify a moral evaluation that incitement is wrong and that the inciter is morally blameworthy.

Suppose for the moment that we accept MR and are confronted with an incident of incitement. How would our moral evaluation go? In the cultural model, incitement is a culturally conditioned event. If a culture is such that it increases the probability of incitement, then any actual instance of incitement happening in that society must be able to appeal to something in its cultural tradition as justification. This point was argued above; both inciter and incitees believe that their culture provides them a "permissive background," an implicit rationale, to act as they do. From the cultural perspective within which the incident of incitement occurs, it would appear to be not wrong but a justified event. After all, its culture prepared the way for it and provided a permissive milieu. From within their culture, both the inciter and the incitees

will be judged to have done a moral wrong *only* if their cultural tradition has *not* prepared the way for the wrong in question; but then this would not be the model of incitement under consideration. According to MR, the moral system by which incitement is judged is *part* of the culture that has eased its way. Thus, to the degree that culture has prepared the way, it is hard to see how incitement can be judged morally wrong from the perspective of that culture's moral standards, a perspective MR requires us to take.

Perhaps the idea of "cultural" as a homogeneous and internally consistent tradition is unrealistic. Most cultures contain countercultural institutions and practices. One such institution might be a population's legal system which could make incitement illegal even though there is a cultural tradition that increases its probability. Another might be a population's dominate religion that condemns incitement as sinful. From a legal or a religious perspective, one can see that incitement might be deemed illegal or sinful even as the broader culture contains animosities among sub-groups that make for a propensity for incitement. However, the focus here is on the *morality* of incitement, not its illegality or its sinfulness. Also, it would be odd to have a culture's *moral* values and norms that condemn incitement to be based in a countercultural movement, given that according to MR a population's moral system is *part* of the population's culture.

It would seem, then, that from the perspective of MR incitement would not appear to be morally wrong and the inciter not only not to have done something wrong but perhaps even to have done something his/her culture would reward. To see incitement as wrong in the cultural model, it appears necessary that we "stand outside" the culture, and judge its wrongfulness from the (detached) "moral point of view."

Suppose, then, that we deny MR and accept MA, the ethical position that morality, at least in its core, is a set of universally applicable principles that are transcultural. Does it provide a base for judging incitement morally wrong? Assuming that there are moral absolutes, the problem here is finding those that would apply to the cultural model of incitement. Moral absolutes would presumably be about the most fundamental aspects of human existence such as: life and what is necessary to sustain human life, and bodily integrity and protection from physical harms. Moral values and norms that qualify as universal would presumably be about those conditions of human existence that all humans share in common, no matter the historical circumstances, the linguistic tradition, and the cultural practices by which humanity is divided into different ethnicities. As a linguistic convenience, let's call such moral absolutes "human rights," which of necessity must be conceived at a sufficient level of abstraction and generality so as to apply universally.

Where the *victims* of the incitees are to have their human rights violated by the *incitees*, it is clear that MA (as characterized here) provides a base for condemning such actions as morally

wrong. In addition, to the degree that a culture represents a milieu of permissiveness with respect to the wrong done by the incitees to their victims, MA would justify condemning that culture (in that respect). So, for example, a culture that provides a positive background for anti-Semitism, or one that makes systemic racism the social norm, can be morally evaluated from the perspective of MA as deficient or "toxic" for its easing the way for incitees to violate the human rights of their victims.

But what about the inciter's act of incitement? Within the cultural model under consideration, it is hard to see how the inciter's use of inciteful language alone could be a violation of anyone's human rights. It seems that the inciter's act of incitement would be absorbed into his/her culture and be counted as one more "influence" that works to increases the probability of the wrong done by the incitees. MA, it seems, provides no *special* ground for evaluating incitement as morally wrong or holding the inciter morally blameworthy for incitement. The focus, from the perspective of MA, is on the wrong committed by the incitees to their victims, and the inciter seems to be let off the hook by becoming part of, and being absorbed in, the larger "toxic" cultural background.

If we suppose, for sake of argument, that a sub-group (A) within a culture that has a tradition of animosity toward another sub-group (B) one day violates the human rights of B. This would be morally wrong according to MA (as characterized here). Now add to A's victimization of B the fact that A was incited to violate B's human rights by inciter S. Is the wrong that A does to B any *more* wrong by having been incited? In this picture, the inciter S does not *partake* in the wrong done by A to B, so A is not being increased by one more agent in the wrong they do to B. S's inciteful use of language however would increase its probability (given the cultural background). If incitement is morally wrong by MA standards, then the cumulative moral wrong of the incitees' actions (to their victims) *plus* the moral wrong of the inciter's actions (of incitement) toward A should be greater than the moral wrong of the incitees' violation of B's human rights alone. But this is not the case; the combined moral wrongs are not more wrong than the incitees' (A's) wrong considered alone by MA standards. There seems to be no basis within MA for claiming an increase in moral wrongfulness by the addition of S's incitement. The moral wrongdoing A does to B before our imaginary addition of S's incitement, and then after this addition, appear to be equal cases of moral wrongdoing, with perhaps the latter case containing an increase in cultural influence represented by the inciter's act of incitement.

Suppose, continuing the same line of reasoning, that A victimizes B and we ask A why? A responds: we were incited by S do it. How would we judge this? If we took "incite" to be a case of "strongly cause" (i.e., A responds "S made us do it."), then we would believe that S somehow *forced* or *coerced* or *made* A violate B's human rights, and the moral wrongdoing would shift from A to S. But this would no longer be the cultural model of incitement; we would be back to

the causal model with all its problems. However, if we took "incite" to mean "culturally primed," as the cultural model of incitement requires, I very much doubt that we would consider the wrong A did to B less wrong; we might even accuse A of offering a feeble excuse in the attempt to avoid moral guilt. In other words, we would take S's role to be morally minimal if not negligible.

In sum, neither MR nor MA seems to provide an ethical perspective for evaluating incitement as morally wrong, and finding the inciter morally blameworthy, within the cultural model of incitement. There are good grounds within MA for accusing the incitees of a grave moral wrong to the degree that the incitees have violated the human rights of their victims, but MA does not appear to contain any moral principles that are both universal and applicable to incitement itself.

5. <u>Concluding remarks</u>

Incitement is by definition a crime and there are specific criminal standards in most countries for prosecuting incitement according to the applicable laws. My exploration, however, concerns the moral evaluation of incitement and the inciter. To find incitement morally wrong, and the inciter morally blameworthy, I explored two lines of inquiry: first, I developed a series of three frameworks or models by which we might conceptualize the relation between the inciter and the incitees, testing the strength of each with counterexamples to discover failures to capture cases of incitement. Second, relative to a given framework or model of incitement, I attempt to apply the principles of a conceptually related ethical theory to discover the theory's limits in morally evaluating incitement as wrong.

The first model of incitement places it in a causal framework and the ethical theory that naturally applies is consequentialism. The causal model, in several variations, failed to capture clear cases of incitement and as a result consequentialism could not be used in these cases to evaluate incitement as morally wrong. The second model places incitement within a subjective (mental content) framework and deontology the relevant ethical theory. While the subjective framework avoids the problems that weaken the causal model, it likewise fails to capture cases that clearly appear to be incitement; in such cases, deontological ethical theory can't be applied to show that incitement is wrong and that the inciter is morally blameworthy. A third framework, the cultural model, appears to be the strongest of the three in capturing cases of incitement and in avoiding the problems that weakens the causal and the subjective models. However, while the cultural conceptualization offers both an insightful and a realistic picture of incitement, there isn't an ethical theory that naturally aligns with it (as consequentialism does with the causal model and deontology with the subjective model). In the absence of a specific

ethical theory, the general theories of moral relativism and moral absolutism were tailored to align with the cultural model of incitement and were applied in the attempt to provide a foundation for evaluating incitement as morally wrong and the inciter morally blameworthy. Neither theory, however, offered principles that would justify evaluating either *incitement* itself or the *inciter's* use of inciteful language as morally wrong, though moral absolutism clearly contains norms condemning the actions of the *incitees* to the degree that incitees violate the human rights of their victims.[15]

By way of a brief analogy, suppose we have three moral agents: A, B, and C. B hurts C such that B can be morally judged to have done a wrong. As we condemn B, suppose B tells us: (1) A *told* me to do it. This would ordinarily not lessen the wrong B did to C. However, we might wish to discover A's motives to see if they warrant an *independent* moral judgment; but even so, it would not be anywhere near the moral wrong that B does to C. Now suppose B tells us: (2) A *made* me do it. If we find that this was true, then it would remove B's moral agency and switch moral blame from B to A. Finally, suppose that B tells us: (3) my society, including A's words, *primed* me to do it. If A = inciter, B = incitees, and C = B's victims, is incitement a "telling" as in (1), a "making" as in (2), or a "priming" as in (3)? (1) seems too weak to be incitement, (2) seems too strong to be incitement, and (3) seems to be the right condition for incitement, but it becomes uncertain how to justify our judgment that A (or B) had done a moral wrong.

The surprising (at least to me) result – namely, that in many cases the inciter's inciteful use of language, independent of free speech considerations, and incitement itself escapes moral condemnation – reflects, I believe, two problems. The first is that the exact relation between inciter and incitees needs further examination and conceptualization (how exactly does the "influence" of inciter on incitees happen?). The second is, given that incitement is not persuasion, how best to form moral norms that condemn all cases of incitement and inciteful language as wrong, but are not so strong or broad that they incorrectly apply to instances of *persuasive* language to do wrong.

 At present, I do not see any way of combining the three models into one overall, comprehensive account of incitement as a morally blameworthy act. Yet, it might be worth exploring some ways in which the virtues of each of the models examined above could be integrated into such a framework. Alternatively, there may be a yet unanticipated model or scheme that could in itself, or in conjunction with the three models examined here, provide a coherent and cogent account for why incitement is a morally blameworthy act, in addition to being a crime in our legal system.[16]

Notes

1. In setting up the problem, I make two assumptions. First, with respect to the law, clearly not everything that is criminal is immoral and not everything legal is moral; there is a well-known gap between law and morality. While incitement is by definition criminal behavior, there might be unusual cases of incitement that are morally justified. However, I will restrict my focus in what follows to cases in which incitement intuitively seems morally wrong, neglecting unusual examples where it might be argued that the inciter is on good moral grounds. Second, I acknowledge that with respect to morality, an inciter in the act of inciting might simultaneously commit several actions that we ordinarily consider to be morally wrong. For example, the inciter (in addition to inciting the target audience) might also be lying or breaking a promise or stealing from time at work. I don't want any such simultaneous immoral exogenous actions to be grounds condemning incitement as morally wrong. My question is: Why is incitement itself morally wrong? A response of the form: Because the inciter is lying and lying is morally wrong, will not be, given the way I am restricting the topic, an acceptable answer.

2. With respect to the inciter I will use "utterances" to simplify the exposition, but I wish to include under "utterances" not only the spoken word, but the written word and the use of images/pictures as well. Likewise, for the incitees I will use "hear" in the case of the spoken word but wish to include "read" for the case of the written word, as well as "see" for the case of images/pictures.

3. We recognize this view to be act consequentialist: the moral theory asserting that the moral status of an action depends on the act's consequences. To simplify, if an act's consequences are broadly categorized as either "good" or "bad" (in a sense that should be specified) and if acts are broadly categorized as either morally "right" or "wrong" (in a sense that should be specified), then consequentialism asserts: right depends upon good being produced and wrong depends upon bad being produced. The relation between an act and its consequences is clearly one of causality. Within consequentialism, if an act has no consequences, then the act cannot be judged morally right or wrong. That incitement appears to be causal is captured by the standard dictionary definitions: *incite* - (1) to move to a course of action, (2) to bring into being or make occur (*Webster's Third International*); *incite* – (1) to set in rapid motion, rouse, stimulate, (2) to put in motion, (3) to stir up (*OED*, 1971). It is natural to understand "to move," "to make," "to set," "to put," "to stir," and similar phrases as various ways of saying "to cause."

4. This, of course, is too briefly stated. The principle assumes several conditions: to cause a wrong *knowingly*, to cause a wrong *voluntarily*, and so forth for the context within which consequentialist moral norms are applied.

5. I realize that this third case goes too quickly and does not do justice to consequentialism. Rule consequentialism has the benefit over act consequentialism in so far as it can argue for a social norm that serves as a moral rule to the effect "no one should incite" or more broadly "no one should break the law." Also, rule consequentialism has the flexibility to accommodate such causal mismatches as in case 3, and perhaps act consequentialism might be broadened to include both double effects as well as relatively distant divergent effects. If defenders of consequentialism appeal to rules or don't see this third case as a problem, there are still the two other cases that I believe can deliver the challenge.

6. See, for example, the entries "Word Meaning" and "Theories of Meaning," especially their opening sections 1.2 and 1.0 respectively, in the *Stanford Encyclopedia of Philosophy* (*SEP*). (Accessed 3/2021)

7. *Supervenience* used here as a term of convenience should not, and I believe does not, introduce any metaphysical problems into my topic of incitement.

8. A summary of the *Wikipedia* article on incitement. The article covers these three subjective states under their legal names: *Mens rea* on both the inciter's and the incitee's parts, and *Actus reus* on the incitee's part. While not all *Wikipedia* entries are of equal quality, the entry on "Incitement" approaches it as a legal concept and appears to be a well-documented and a reliable entry from a legal point of view. Because my interest in incitement here is from a moral point of view, even though it is not a peer-reviewed source, I take it to be appropriate to cite this entry in setting up the subjective model.   (Accessed 3/2021)

9. We recognize this view to be Kantian or deontological: the moral theory asserting that the moral status of an action depends on the conformity of the agent's mental state, specifically the agent's intention, to moral norms, values, or duties. Even if an act has no consequences, the act can be judged morally right or wrong according to what the agent intends to do by acting, given a set of moral norms or duties. (See, for example, Kant's *Groundwork* for the classic statement of this ethical theory).

10. I take the following two analogies to be more intuitive than the famous 3-box problem in which showing one of 3 boxes to be empty increases the probability from 1/3 to 2/3 that switching gains the prize, but it does not *cause* the prize to be in the switched box; the prize was randomly already in one of the three boxes each with 1/3 probability originally.

11. A horrific example occurred in Rwanda (1994). Against a cultural background of class animosity between a Tutsi minority, herders perceived by Hutus as a "wealthy" upper class, and a Hutu majority, farmers perceived by Tutsis as a poorer "lower class," Hutu civilians were incited by media broadcasts and influential political leaders to join the military in a massacre of Tutsis. This attempted genocide of Tutsis by Hutus was in part a civil war between armed

forces, but the Hutus' civilian participation appears, by all reports, to have been a case of incitement. (See the *Wikipedia* entry "Rwanda genocide." Accessed 4/2021.)

12. While beyond to scope of this examination of incitement, if the culture in question should happen to contain a legal tradition of free speech, as the US does with its Constitutional First Amendment, incitement becomes in addition a difficult legal case to prosecute. If there is no free speech tradition in a population's legal history, incitement can be made a crime of wide scope that's relatively easy to prosecute. Indeed, in some countries some forms of graffiti can be prosecuted as a crime of incitement (see the entry "Incitement" in *Wikipedia* for several examples) whereas graffiti in the US typically falls under either free speech or property damage.

13. See Daniel Bell's entry "Communitarianism" in *SEP* (revised 2020) for a presentation of its political philosophy and philosophy of the self. Bell presents the debate between *universalism* and *particularism* as a broad issue that covers both ethics as well as political judgments; however, this debate more concerns the relative importance of developing a communitarian perspective over against the traditional universalist assumptions in Western ethics, political philosophy, and concepts of the self than concern for developing a communitarian ethical theory that could be applied to evaluate actual moral situations. See also Chapter 5 "Communitarianism" in Jean Hampton's *Political Philosophy* (1997, Westview Press). Hampton presents the case for a communitarian political theory contrasting it with the liberal, individualistic perspective of traditional Western political philosophy. Again, no attempt is made by Hampton to develop a communitarian ethical theory.

14. See, for example, the entries "Relativism" in the *SEP*, especially section 4.5 on Moral Relativism, and "Moral Relativism" in *SEP*. (Accessed 4/2021). As both entries explain, MR is a major theory within ethics that is not easily refuted. Similarly, MA (the "view from nowhere" theory) is equally a major ethical theory that is not easily refuted. In the history of (western) philosophy, powerful arguments have been developed pro and con for each theory with no broad consensus among philosophical ethicists that one is a superior theory.

15. It might be apposite to recall here that *not* being able to show that incitement is wrong is *not* the same as showing that it is *not* wrong. And arguing that an action is *not* morally wrong is *not* the same as arguing that it is morally right or morally permissible.