

Toward a Theory of Offense: Should You Feel Offended?

CHANG LIU

Abstract

The feeling of being offended, as a moral emotion, plays a key role in issues such as slurs, the offense principle, ethics of humor, etc. However, no adequate theory of offense has been developed in the literature, and it remains unclear what questions such a theory should answer. This paper attempts to fill the gap by performing two tasks. The first task is to clarify and summarize the questions of offense into two kinds, the descriptive questions (e.g., what features differentiate offense from similar moral states like anger?) and the normative questions (e.g., what are the conditions for taking offense to be apt?). The second task is to answer these questions by developing what I call ‘the violated norm theory of offense’. According to this theory, feeling offended entails that the norm one endorses is judged to be violated by the offender. Appealing to the violated norm enables this theory to answer the descriptive questions (e.g., taking offense differs from anger because of features like not requiring victims and the difficulty of animal offense) and the normative questions of offense (e.g., taking offense is apt only if the violated norm is universalizable).

1. Introduction

In an age of clashing values, the feeling of being offended gives rise to many controversies and philosophical questions. A controversial Dutch Christmas tradition involves ‘Black Pete’, a helper of Santa Clause played by performers in blackface and costumes (Morse, 2012). Some proponents argue that Black Pete, merely a holiday tradition for kids, is not intended to offend people of African descent and the Netherlands is historically irrelevant to blackfacing in American minstrel shows. However, many accuse this character of racism on the ground that blackfacing promotes racial injustice, even if it does so unintentionally. Cases like this bring up philosophical questions about the emotion of offense. When is it appropriate to take offense? Should we refrain from taking offense if the offender has no bad intention? Is the world a better place if no one feels offended?

Many philosophical debates involve questions about offense, but a theory of offense has yet to be developed in the literature. For instance, Feinberg (1985) argues for the Offense Principle (i.e., Behaviors can be legally prohibited to prevent offense) as an

alternative to the Harm Principle. Theories of slurs have been explaining the offensiveness of slurs and how to regulate offensive words (Anderson and Lepore, 2013; Bolinger, 2015). In the ethics of humor, feminist philosophers have discussed the nature of the offense caused by sexist jokes (Bergmann, 1986). Nevertheless, these discussions have not produced a theory to answer the questions about offense systematically. It also remains unclear what questions should be asked and how they could be answered.

This paper aims at filling this gap by performing two tasks. The first task is to clarify two groups of questions that a theory of offense has to answer. On the one hand, there are the *descriptive questions* such as Q1) the nature question (e.g., ‘What is it to feel offended?’), Q2) the demarcation question (e.g., ‘What differentiates between offense and similar emotions like anger?’), Q3) the taxonomy question (e.g., ‘How should varieties of offense be categorized?’). On the other hand, there are the *normative questions* such as Q4) the aptness question (e.g., ‘Under what conditions is it appropriate to take offense?’) and Q5) the value question (e.g., ‘Does taking offense have any value?’).

The second task is to provide a moral psychology theory of offense to answer these questions. I will call this theory ‘*the violated norm theory of offense*’. Here is the basic idea: to feel offended entails that the *norm* one endorses is judged to be *violated* by the acts of the offender. For instance, when people are offended by blackfacing in the Dutch Christmas tradition, they judge that it violates their endorsed principle of racial equality.

Appealing to the violated norm enables this theory to answer the questions I introduced. This theory’s answer to Q1) the nature question is that taking offense entails that a norm one endorses is judged to be violated. As for Q2) the demarcation question, I argue that the norm involved in taking offense gives rise to special features (e.g., not requiring victims and the difficulty of animal offense) that differentiate offense from other moral states. My theory’s answer to Q3) the taxonomy question is to categorize offense by what kind of norm is violated and how the norm is violated. As for Q4) the aptness question, my theory holds that someone’s taking offense is apt only if the norm she endorses is universalizable. As for Q5) the value question, I believe that taking offense has an instrumental value to promote the norms.

It is helpful to clarify the term ‘offense’ at the beginning, given its ambiguity. First, I will use ‘*offense*’ to refer to the mental states of feeling offended or taking offense (rather than offenses in the military or the legal senses). Therefore, ‘offense’, ‘taking offense’, and ‘feeling

offended' will be used interchangeably. Second, my usage of 'offense' does not mean *acts of offending people*. Since the term 'offense' is reserved for a kind of mental state, I will call these acts '*offending acts*'. Likewise, my theory of 'offense' intends to address questions about certain *mental states* (e.g., Is it appropriate for me to feel offended?), not questions of certain *acts* (e.g., Is it morally wrong to offend others?). Questions about offending acts are relevant but different issues. A theory of offense as mental states will be helpful to address them.¹ Third, my usage of 'offense' does not include being offended in the sense of having displeasing experiences (e.g., being offended by the offensive odor of rotten fruits). I will explain this in detail in section 3.2.

Here is the plan for this paper. In section 2, I will argue that a theory of offense matters for theories of slurs, ethics of humor, and the offense principle. To fulfill the first task of this paper, the questions about offense will be clarified and summarized in the following two sections. In section 3, I will introduce the descriptive questions of offense, including the nature question, the demarcation question, and the taxonomy question. In section 4, the normative questions, such as the aptness question and the value question, will be presented. To achieve the second task, I will develop the violated norm theory to answer these questions. Section 5 will illustrate its core idea: taking offense entails that the norm one endorses is judged to be violated. In section 6, this theory will answer the descriptive questions. In section 7, it will provide answers to the normative questions. Finally, section 8 will conclude this paper.

2. Why Offense Matters

In this section, I will illustrate the need for a theory of offense with its implications for debates in theories of slurs, feminist ethics of humor, and the offense principle of legal prohibition.²

¹ See § 7.1 for more details.

² A theory of offense also has implications for theories of forgiveness and theories of anger. For instance, forgiveness as overcoming negative emotions is often taken to be a good thing (Darwall, 2006; Moore, 1989; Richards, 1988). However, how can forgiving offenses be good if taking offense has a certain value? Moreover, theories of anger have characterized the conditions of apt anger (Srinivasan, 2018). Is offense a kind of anger? Does the norm of anger apply to offense?

Theories of Slurs: A theory of offense matters for theories of slurs (i.e., offensive words like ‘chink’ and ‘queer’) in two ways. First, it helps to explain an essential feature of slurs, i.e., their offensiveness (Anderson and Lepore, 2013; Bolinger, 2015). The prohibition theory, for instance, holds that slurs cause offense because *using the word itself* violates the prohibition on those words (Anderson and Lepore, 2013, p. 353). By contrast, Bolinger’s contrastive choice account proposes that slurs are offensive words because using them *communicates an offensive message*, e.g., using ‘chink’ instead of ‘Chinese’ signals that the speaker endorses a negative attitude toward them (Bolinger, 2015, p. 9). Therefore, a theory of offense helps to explain how slurs cause offense (e.g., Is the offense caused by the slur itself or the message expressed by the slur?).³ Second, a theory of offense is useful for understanding the moral status of slurs. Sometimes, slurs are used without intentions to derogate or to offend. For example, Tupac Shakur, an African American rapper, takes the N-word to mean ‘Never Ignorant Getting Goals Accomplished’ in his rap song (Kennedy, 2003, p. 36).⁴ Should people feel offended, if a slur is used without bad intentions? Questions like this need answers from a theory of offense.⁵

Feminist Ethics of Humor: Bergemann (1986) gives a feminist analysis of the two ways in which sexist jokes are offensive. First, some sexist jokes are offensive because they disparage and belittle women. Second, even if sexist jokes are not meant to disparage women (‘No offense is meant’), they are still offensive for presupposing sexist beliefs. Bergemann argues that holding sexist beliefs is the prerequisite of appreciating the fun of sexist jokes. Feminists are offended, not just by the disparagement, but also by the sexist beliefs as the grounds of the fun. Bergemann’s distinction between the two kinds of offense requires a theory to clarify the varieties of offense.⁶ What distinguishes between the two kinds of offense? Why is the first kind excusable by ‘no offense intended’ but not the second kind?⁷

³ See § 6.3 for my analysis.

⁴ Examples like this are often called ‘non-derogatory uses’ or ‘non-offensive uses’ in theories of slurs. For explanations of the non-derogatory uses, see Liu (2019; 2020), Diaz-Legaspe, Liu and Stainton (2019).

⁵ For the answer from my theory, see § 7.1.

⁶ A theory of offense also helps the incongruity theory in philosophy of humor. See § 6.1.

⁷ See § 7.1 for my answers.

Offense Principle: A theory of offense matters for Feinberg's Offense Principle. Mill's Harm Principle holds the only justification for limiting liberties is preventing harm to others (Mill, 1977, p. 223). By contrast, Feinberg (1985, p. xiii) argues that causing *offense*, instead of doing *harm*, can warrant legal prohibition: 'It is always a good reason in support of a proposed criminal prohibition that it is probably necessary to prevent serious offense to persons...'⁸ However, this Offense Principle faces the challenge that not every instance of offense should be given equal considerations. If white supremacists are offended by the civil rights movements, they should not be given equal consideration as those who are offended by racial discrimination. To address this, the offense principle needs a theory of offense to differentiate between appropriate offense and inappropriate offense.⁹

3. The Descriptive Questions of Offense

The first task of this paper is to outline the two kinds of questions of offense, i.e., the *descriptive questions* about the nature and the features of offense and the *normative questions* about the aptness and the value of offense. In this section, I will present the descriptive questions of offense, including the nature question (§3.1), an objection against the nature question (§3.2), the demarcation question (§3.3), and the taxonomy question (§3.4).

3.1 The Nature Question

Perhaps the most central question of a theory of offense concerns the nature of offense. A personal example may be helpful here.

A group of drunken white undergraduates approached me on Saint Patrick's Day. Perhaps they were looking to have some fun or to make some trouble. One of them asked for my name and I told him my Chinese name. He then bowed to me and greeted me with 'Konichiwa'. I did not feel threatened, as they were not aggressive. Nonetheless, I was slightly offended. By greeting me in Japanese, the drunk undergraduate acted as

⁸ I will argue that Feinberg's account of offense in criminal law does not apply to a moral psychology of offense in § 3.2.

⁹ See § 7.1 for a detailed analysis.

if he did not care if I was Chinese or Japanese. For them, Asian identities did not deserve to be recognized.

Cases like this give rise to philosophical questions: What is it for someone to take offense? In this example, what is going on when I am offended by the greeting in Japanese? Call this the Q1) '*nature question*' of offense:

Q1) The Nature Question: What is it for a person to feel offended?

To address this question, we have to answer whether offense is relational or not. Is taking offense a person's intrinsic property, just like other raw feelings? If so, the Japanese greeting offended me simply by inflicting something like pain in me. Otherwise, taking offense is perhaps something relational. Maybe when I am offended, I am offended about *something*. But what does the relation of offense consist of? Am I offended by the person, the person's attitudes, or his acts or behaviors? To answer this question, we also have to specify which aspect of me is involved in the offense. Is it my belief, my dignity, or my personhood that is being offended?

3.2 The Heterogeneity Objection against the Nature Question

A common objection against Q1) the nature question is that 'being offended' is not a unified phenomenon to theorize about. What we call 'offense' is a wide spectrum of things, ranging from being offended by an odorous smell, being offended by nudity, to being offended by racist remarks online, etc. It makes no sense to ask for the nature of taking offense because the varieties of taking offense have nothing in common.

Here is the short version of my response: although feeling offended *in the sensory sense* is indeed too heterogeneous, a unified moral psychology theory of offense is possible because it explains feeling offended *in the normative sense*.

The first step of this response is to disambiguate the term 'offense'. 'Being offended' appears disunified because it is ambiguous between a sensory sense and a normative sense in English. On the one hand, one can 'be offended' in the *sensory sense* when something causes displeasing sensations in her. This is the sense in which one is said to be 'offended by a bad smell' and a flavor can be 'offensive'. On the other hand, a person is offended in a *normative sense* when values or norms are involved, e.g., being offended by someone's statement of their

political view.¹⁰ The ambiguity of 'offense' can be supported by cross-linguistic examinations. For instance, the Chinese word for 'being offended' (*bei mao fan*) has only the normative sense. It makes no sense to say one feels 'mao fan' from an odorous smell in Chinese. Similarly, the German translation 'beleidigung' lacks the sensory sense of 'offense'.

The second step is to argue that a moral psychology theory of 'offense' explains the normative sense, not the sensory sense of taking 'offense'. The heterogeneity objection assumes that a moral psychology of offense explains being offended in the sensory sense, e.g., by an odorous smell, disgusting taste, etc. Offenses in the sensory sense are indeed too heterogeneous to theorize; they have little in common besides the fact that they are all displeasing. By contrast, taking offenses in the normative sense is a unified phenomenon. The answer to Q1) the nature question from a moral psychology of offense should specify the common component of offenses in the normative sense. As I will argue in section 5, the normative sense of offense shares a unifying component, i.e., the judgment about violating an endorsed norm.

One might challenge my response to the objection with Feinberg's account of offense, which primarily focuses on its sensory sense. In formulating his offense principle of criminal law, Feinberg (1985, p. 1) uses the term 'offense' to cover 'the whole miscellany of universally disliked mental states', including 'passing annoyance, disappointment, disgust, embarrassment, and various other disliked conditions such as fear, anxiety...'. Although Feinberg does not have a full-blown theory of 'offense', his notion of offenses as disliked mental states supports his task in criminal law. He emphasizes that offenses are not harms 'occupying the lower part of the same scale'. Moreover, focusing on the sensory sense of offense allows him to address offenses regulated by nuisance law, such as emitting odors and making noise.

Developed as an account of offense for criminal law, Feinberg's account does not work for moral psychology. In particular, what Feinberg counts as 'offense' for the offense principle is both too broad and too restrictive for a moral psychology theory of offense. First, it is too broad because it primarily focuses on 'offense' in the sensory sense. However, many cases of being offended in the sensory sense should not be within the domain of moral psychology. Suppose I am offended by the disgusting smell of rotten fruits.

¹⁰ This should not be confused with Feinberg's technical definition of the 'normative sense' of offense (Feinberg, 1985, p. 1).

Despite the fittingness of this feeling, it cannot be evaluated as right or wrong, appropriate or inappropriate.¹¹ Second, it is also too restrictive for a moral psychology of offense. His notion of offense focuses only on the offense caused by wrongful (right-violating) conducts (Feinberg, 1985, p. 1). This is because criminal law cannot regulate offenses when no one is wronged. However, this leaves out cases like taking offense from aesthetic affronts and breaking etiquettes. A moral psychology of offense should evaluate the appropriateness of taking offense at someone's bad aesthetics, even if no one is wronged by it.

3.3 The Demarcation Question

Question Q1) matters for a related question, i.e., what differentiates offense from other similar moral states? Let us call this Q2) '*the demarcation question*' of offense. Feeling offended appears to be similar to many moral emotions, such as annoyance, anger, resentment, and hatred. Therefore, we need to know whether there is a boundary between offense and other states. Is offense a sui generis kind of moral state? Can offense be ultimately reduced to other moral states such as anger? If not, what differentiates offense from them?

Answering this question of demarcation takes two steps. First, we have to figure out if offense has distinctive features that other moral states lack. If offense does have distinctive features, it is a sui generis moral state. Otherwise, offense can be reduced to other moral states. Second, after identifying these features, we have to explain why offense exhibits them. An account of the nature of offense helps to explain it. This is why one's answer to the question of demarcation depends on one's answer to the nature question of offense. Let us summarize the question of demarcation as follows:

Q2) The Demarcation Question: Does offense have features that differentiate it from other moral states such as anger? If so, why does offense exhibit those features?

3.4 The Taxonomy Question

Even with a good answer to Q2) the demarcation question, we still need to taxonomize the many varieties of offense. There seem to be

¹¹ For the fittingness of feeling offended, see § 4.1.

many kinds of offense, and they differ in at least two dimensions. First, offense can vary in what I call '*domains*', i.e., what the act offends. For instance, I may be offended *morally* when I see someone is enjoying bullying others. Moreover, a person can take offense *politically* when they are discriminated against because of their race. It is also possible to be offended *aesthetically*; an art lover can be offended by someone's bad taste. Second, offense can vary by what I call '*modes*', i.e., how an act offends. Being offended by someone's bullying others is a '*direct*' offense about the act itself. By contrast, being offended by someone's laughing at the act of bullying is a '*communicative*' offense about the message that the laughter sends. Moreover, asking the victim about their experiences of being bullied may cause '*consequential*' offense by producing traumatic psychological effects.¹² There may be more domains and modes than what I have introduced here.

These complexities give rise to Q3) 'the taxonomy question' of offense, i.e., how should varieties of offense be taxonomized? Let us summarize this question as follows:

Q3) The Taxonomy Question: How should the varieties of offense be categorized? Why do they fall under these categories?

Answering the taxonomy question takes two steps. The first step is to identify the dimensions by which offense varies. Are there domains other than moral, political, and aesthetic offense? Are there other modes than direct and communicative offense? Are domain and mode the only two dimensions to categorize offense? The second step is to explain why offense can be categorized along with these dimensions. On the one hand, we have to explain the differences within a given dimension. For instance, what differentiates being offended morally from being offended politically? Are they differentiated by the kinds of acts (e.g., moral acts, political acts, etc.) that cause offense? On the other hand, we need to explain the nature of different dimensions. For instance, what are the domains and modes of offense?

4. The Normative Questions of Offense

In addition to the descriptive questions, a theory of offense has to answer normative questions about offense. In this section, I will introduce the aptness question (§4.1) and value question (§4.2) of offense.

¹² See § 6.3 for the detailed analysis.

4.1 *The Aptness Question*

Aptness (or appropriateness) is one of the crucial aspects in evaluating mental states. Not all mental states are subject to the evaluation of aptness. For instance, it is absurd to say ‘It is inappropriate for you to feel the pain’ to a person who is in pain. By contrast, offense, like many moral states, can be apt or not. When people take offense in ordinary life, we usually evaluate their feelings as appropriate or inappropriate. It is common to say things like ‘He should not be offended by such a trivial issue’, or ‘She has a good reason to feel offended by that statement’. Someone who constantly gets offended by trivial things is usually described as ‘oversensitive’. Moreover, the aptness of offense is often the core issue of many controversies. For instance, people debate over whether we should feel offended by blackfacing in the Dutch Christmas tradition. It is also controversial whether people should feel offended by the MAGA hats worn by Trump supporters.

Here are two caveats about my usage of ‘aptness’ (or ‘appropriateness’). First, ‘aptness’ does not mean the fittingness of taking offense. Like many emotions, taking offense have representational contents. Feeling offended is fitting when the contents are accurate, i.e., the norm one endorses is indeed violated by the offender’s act. Nevertheless, aptness is an evaluative notion beyond fittingness. Taking offense can be fitting without being apt. For instance, a white supremacist’s being offended by racial justice is fitting when their racist norm is indeed violated, but it is not appropriate.

Second, to say taking offense at something is *apt* or *appropriate* is to say that there is a reason or justification to feel offended. Correspondingly, feeling offended is inapt or inappropriate when it is unreasonable or unjustified to be offended. Moreover, the reasons must be intrinsic, rather than instrumental. Suppose someone pays me 100 dollars each time I feel offended. My taking offense is not made apt by this instrumental reason.

The aptness of offense gives rise to a normative question: What does it take for feeling offended to be apt (or appropriate)? A theory of offense has to identify the conditions for taking offense to be apt. Call this Q4) ‘the aptness question of offense’:

Q4) The Aptness Question: What are the conditions for taking offense to be apt?

I believe an answer to Q4) involves answering at least three sub-questions.

Toward a Theory of Offense

First, an answer to Q4) has to specify the good ‘reasons’ (or ‘warrants’) to take offense. When we are offended because of certain beliefs or values, which one of them warrants taking offense? When a Chinese person is harassed with the Japanese greeting ‘Konichiwa’, racial equality seems like a good reason to take offense. After all, no one should be discriminated against because of their racial identity. However, there are trickier cases. Suppose a person is offended by criticisms of their country because they cannot tolerate any insult to its greatness. Is this nationalism a good ‘reason’ to feel offended? If not, what differentiates between insulting one’s national identity and one’s racial identity?

Second, an answer to the aptness question has to address a question about standings: Who is in the right position to take offense? That is, does the aptness of offense require special standings such as victimhood or certain relations to the victims? When an act directly harms someone, it seems appropriate for the victims to take offense. But how about those whose interests are unaffected by the offending acts? Is it appropriate for a resident of China to be offended by slavery in the U.S., even if she is historically unrelated to the injustice?

Third, answering the aptness question has to take intentions into consideration: Does the intention of the offender matter for the aptness of taking offense? Offending acts are often given excuses such as that the act is unintentional, or ‘No offense is meant’. For instance, many argue that no offense should be taken from the act of blackfacing in the Dutch Christmas tradition. This is because the performers of Black Pete, who entertain children during the holiday, do not mean to offend black people. Is taking offense inapt when the offender does not intend to offend?

4.2 The Value Question

Being apt and being good should be distinguished from each other. A moral state can be apt without being good, e.g., hating a cheating partner is appropriate, but it may be bad because of damaging relationships. Similarly, the value of taking offense is a separate issue from its aptness. It is possible for taking offense to be apt even if it has a negative value. It may be argued that oppressed minorities are justified in taking offense from racist remarks, even if their feeling of-fended does not make anything better.

Therefore, a separate question about the value of offense arises: Does taking offense have any positive value? Let us call this ‘the

value question of offense'. On the one hand, feeling offended seems to be a bad thing. We often avoid offending others, and we apologize if they get offended. On the other hand, the feeling of offense may have certain values in one's moral life; there seems to be something wrong with a person who never gets offended, even by the most monstrous acts.

What does it take to answer the value question? An answer to the value question has two preliminary options: taking offense either has no positive value or it has some positive value. If offense has a positive value, we have to specify whether its value is intrinsic or instrumental. Is taking offense valuable by itself? Is it good because it helps to promote other values such as social justice? Let us summarize the value question as follows:

Q5) The Value Question: Does taking offense have any positive value? If it does, what kind of value is it?

5. The Basic Idea of The Violated Norm Theory

After the first task of introducing the descriptive and the normative questions, the second task of this paper is to develop a theory to answer them. This section will begin with a naïve relational theory of offense. By showing its flaws, I will illustrate the basic idea of what I call 'the violated norm theory of offense'.

I will start with a naïve relational theory of offense. Consider again Q1) the nature question: What does it take to feel offended? The naïve theory holds that taking offense is a binary relation between a person and an act. That is, for a person to feel offended is for them to feel disrespected by the offender's act. In other words, a person p feels offended only if p judges that person q 's act A disrespects p . For example, when a Chinese person gets offended by someone's saying 'Konichiwa', they judge that the offender's greeting disrespects them.

However, this naïve relational theory faces apparent counterexamples. People sometimes take offense even if they are not targeted by disrespectful acts. For instance, non-Jewish people can be offended by anti-Semitic speeches, and men can be offended by sexist comments against women. Suppose the Chinese person offended by the Japanese greeting tells their experience to their white friend, who has never met the offender. The white friend can be offended by this racist incident, although they are not a direct target of disrespectful greeting in Japanese. If this is possible, taking offense should be more than a simple binary relation.

Toward a Theory of Offense

My violated norm theory analyzes offense as a ternary relation between the offended person, the offending act, and the norm that is judged to be violated. The basic idea is that for someone to take offense is for them to feel that the offender's acts violate the norms they endorse. This explains why people can take offense when they are not the target of the offender's acts. Although they are not the direct targets, their values or cherished norms are targeted by the disrespectful acts. For instance, when the white friend gets offended by the racist greeting to their Chinese friend, they judge that their principle of racial equality is violated, even if they are not the target. Similarly, when male feminists are offended by sexist remarks against women, they endorse a norm of gender equality and they feel that the remarks violate their endorsed norm.

Let us summarize the core idea of the violated norm theory as follows:

The Basic Idea of the Violated Norm Theory: A person p feels offended by person q 's act A only if 1) p endorses a norm M , and 2) p judges that A violates M .

Concepts like endorsement, violation, and norms have to be clarified in characterizing this theory. First, to *endorse* a norm is to be in a kind of mental state that approves of the norm. Therefore, my usage does not include endorsements such as saying that 'I support gender equality' or signing a statement of gender equality. Second, my usage of 'norm' does not mean 'established social practices'. What I have in mind is similar to Gibbard's account of norms in his norm expressivism (Gibbard, 1990, p. 46). That is, a norm is a set of possible rules or prescriptions (e.g., 'Do not use racist language'). They evaluate and prescribe acts. It is possible for a norm to never be instantiated in society, e.g., gender equality in a patriarchal society. Third, an act violates a norm when it is an act prohibited by the norm. For instance, someone's sexist remarks violate the norm of gender equality because gender equality prohibits making discriminatory remarks against women.

Here are four more caveats about the theory. First, taking offense involves one's *judging* that their endorsed norm is violated, rather than the actual violation of the norm. This is because taking offense often arises from mistaken judgments about violating norms. Suppose a white friend kindly offers their Chinese friend a ride by asking 'Do you have a car?'. The Chinese friend can take offense by misunderstanding the offer as a racist message that the Chinese are too poor to afford cars. Second, taking offense does not require the offended to be conscious of the violated norm. Sometimes people endorse certain norms without being aware of

them (e.g., implicit racism). They can be offended even if they cannot explicitly articulate their 'reasons' to feel offended.¹³ Third, this theory does not exclude the possibility of an empirical approach to offense. Conceptual analysis is only half of the whole picture because it remains silent on the functions, neural mechanism, and evolutionary origins of the mental state of feeling offended. Therefore, empirical research on offense would complement the violated norm theory. Fourth, I believe taking offense is (at least partially) a cognitive or evaluative mental state. That is, taking offense is not a pure feeling. It involves a cognitive component, i.e., the judgment that the endorsed norm is violated.

Moreover, the basic idea of the violated norm theory could be further refined in many directions. I will briefly walk through two of these possibilities, given the limit of this paper. First, the necessary condition of taking offense could have restrictions on the *capabilities of offenders*. That is, feeling offended could require that the offender is capable of following norms. This would allow the violated norm theory to explain *not being offended* by violations of norms by inanimate objects, animals, and even children. For instance, I can be unoffended by a barking dog that breaks the etiquette norm of keeping quiet.¹⁴ This is because I do not take dogs to be capable of following human etiquette. Second, the basic idea could be refined in terms of the *proportionality of violation*. That is, the *weaker* the endorsement of the norms is, the *higher* the degree of violation is required for offense to happen. This could allow the violated norm theory to accommodate cases where the violation of norms is too insignificant to offend. Consider the example of persons who are amused (rather than being offended) by offensive jokes but find them morally objectionable (D'Arms and Jacobson, 2000). They do not feel offended because the joke's violation is not significant enough, given their weaker endorsement of the norms. Finally, notice that these further refinements complement the basic idea of the violated norm theory, instead of contradicting it. In other words, they do not prove the necessary condition specified in the basic idea to be unnecessary.

¹³ This does not conflict with my claim that taking offense involves judgment; the offended can consciously judge that the offender violates their endorsed norms without being aware of the exact norm that is violated. This is analogous to the conscious judgment that something is wrong with my car without knowing which part is broken.

¹⁴ I may be offended by its owner for not controlling the dog. But this is different from being offended by the dog.

This basic idea of introducing norms as a component of offense plays a key role in the violated norm theory. As the following sections will show, many questions of offense can be answered by appealing to the violated norms.

6. Answering the Descriptive Questions

So far, I have introduced the basic idea of the violated norm theory of offense. In the following two sections, this theory will be expanded to answer the two kinds of questions of offense. This section will address the descriptive questions, including the nature question (§6.1), the demarcation question (§6.2), and the taxonomy question (§6.3).

6.1 Answering the Nature Question

Q1) The nature question of offense asks for what it is to feel offended. The basic idea of the violated norm theory has already answered this question. My answer can be summarized as follows:

- Q1) The Nature Question: What is it for a person to feel offended?
- A1) A person feels offended only if the norm they endorse is judged to be violated by the act of the offender.

The violated norm theory has implications for philosophy of humor. According to the incongruity theory of humor, we find something funny when we perceive incongruity and our expectations are violated (Clark, 1970). My theory helps the incongruity theory to explain why offensive jokes can be funny. Jokes are offensive because they violate the norms people endorse, e.g., etiquette and social taboos. Members of society often expect these norms to be upheld. Consequently, there is congruity when jokes violate them. For instance, people are usually expected to conform to the etiquette norm of avoiding the F-word. Therefore, jokes about the French word for seal ('phoque') can be funny because of the incongruity between an endorsed norm and an unexpected way to bypass it.

6.2 Answering the Demarcation Question

The demarcation question (Q2) asks whether offense has special features to distinguish it from other moral states and why offense has

those features. In this paper, I will focus on the distinction between *offense* and *anger*, given the similarity between the two.

Consider an 'identity view' that identifies offense with anger. When we are offended by something, we are usually angered by it. This is because offense is not a *sui generis* kind of moral state; it has no unique feature to distinguish it from anger. To be offended is to be angered, and vice versa.

However, this identity view struggles to explain a special feature of taking offense, i.e., there is *no victim required* for someone to be offended. Anger lacks this feature; for someone to be angered, there must be a victim and damages done to the victim. For instance, when marginalized groups are angered by hate crimes, they consider themselves the victims who suffer the physical and mental damages. Nussbaum's theory can explain this by taking anger to involve the belief that 'there has been some damage to me or to something or someone close to me ...' (Nussbaum, 2004, p. 188). Taking offense, unlike anger, does not always require victims who are damaged. It is possible to be offended even if there are no clear victims. For example, suppose a coffee lover is offended by the bad taste of someone, who drinks coffee with too much sugar and milk. The coffee lover is not the victim because they do not consume the bad coffee. Nor is the offender a victim of themselves, since they enjoy this way of drinking coffee. It remains unclear who has suffered damages.

The violated norm helps to explain why there is *no victim required* for someone to be offended. This is because it is a norm, rather than a person, that is felt to be 'damaged' when someone is offended. When we take offense, we think that our norms or principles are under attack. So long as an endorsed norm is violated, it is possible to take offense even if no person is damaged. In the example above, the coffee lover judges that his norm against unnecessary condiments (not himself) is violated by the offender.

Another problem of reducing offense to anger is the *difficulty of animal offense*. Despite the controversy over animal minds, we have no problem attributing anger to animals (even if they have no mental states). If I trespass into the territory of a nesting goose, it can hiss at me and chase me. It is natural to say that the goose is angry at me. By contrast, it seems far more difficult to attribute offense to animals.¹⁵ We can hardly find conservative cats offended by veganism or liberal dogs offended by violations of animal rights.

¹⁵ Animals can be said to be offended in the sense of being attacked. But this is a different sense.

Perhaps it is unfair to use human values as examples. Consider again the example of the nesting goose. It is more difficult to say that the goose feels offended than to say that it is angry at me. Taking offense appears to be a uniquely human capability. If offense can be reduced to anger, why is it harder to attribute offense, unlike anger, to animals?

Again, my theory has no problem with explaining the *difficulty of animal offense*. We find it difficult to attribute offense to animals because it is unclear whether they can endorse certain norms or values. When the angry goose chases me out of its territory, it is probably out of its territorial instincts rather than its insistence on rights to privacy. By contrast, it is easy to attribute anger to animals because it does not require the capability to uphold norms.

The answer to Q2) can be summarized as:

- Q2) The Demarcation Question: Does offense have features that differentiate it from other moral states such as anger? If so, why does offense exhibit those features?
- A2) Offense has special features, such as not requiring victims and the difficulty of animal offense, because taking offense entails that one's endorsed norms are judged to be violated.

6.3 Answering the Taxonomy Question

The violated norm theory answers Q3) the taxonomy question (i.e., 'How should the varieties of offense be categorized?') by explaining the domains and modes of offense respectively.

One of the dimensions to categorize varieties is their *domains*. Therefore, we can taxonomize offense into *moral offense*, *political offense*, *aesthetic offense*, etc. The kinds of offense are not limited to these few domains. Since taking offense involves a norm, there should be as many domains of offense as the kinds of norms. What is the difference between these domains, e.g., being offended morally and being offended politically? According to the violated norm theory, the difference is in the kind of norm that is judged to be violated. A person is morally offended when they judge that a moral norm or principle they endorse is violated, whereas a person is politically offended when they judges that their endorsed political norms are violated. Therefore, the violated norm theory can explain this taxonomy by appealing to the kind of the violated norm.

Another dimension to taxonomize offense is *the modes of offense*.¹⁶ There should be at least three modes: *direct offense*, *communicative offense*, and *consequential offense*.¹⁷ According to the violated norm theory, these three modes differ with respect to *how* the offender's act violates the endorsed norm. First, direct offense happens when *the act itself* violates the endorsed norm. Consider the example of bullying again. Taking offense from an act of bullying is a direct offense, in which the person judges that this act violates their moral norm against bullying. Second, communicative offense happens when *the act expresses messages* that violate the endorsed norm. For instance, someone's laughing at the acts of bullying sends the message that bullying is fun and should be condoned. This message violates the anti-bullying norm endorsed by the offended person. However, the act of laughing *itself* is not directly violating the norm against bullying (which does not prohibit laughter). Third, consequential offense occurs when the *act produces certain effects* that violate the endorsed norm. Suppose asking a victim about their experience of being bullied (even out of kindness and sympathy) reminds the victim of their traumatic experiences. Asking the question itself, as well as the message it communicates, does not violate the norm against bullying. However, producing traumatic psychological effects violates a certain moral norm that protects the victims from being traumatized.

My answer to the taxonomy question can be summarized as follows:

- Q3) The Taxonomy Question: How should the varieties of offense be categorized? Why do they fall under these categories?
- A3) Offense can be categorized at least by their 1) domains (e.g., political, ethical offense, etc.) and 2) modes (e.g., direct, communicative, and consequential offense). This is because taking offense can vary by 1) the kinds of norms that are violated and 2) how the norms are violated.

The violated norm theory helps to adjudicate between competing explanations of the offensiveness of slurs. The prohibition theory argues that slurs are offensive because *using these words themselves*

¹⁶ Notice that my theory allows the combination of the domains and the modes of offense. For instance, there can be direct moral offense and communicative moral offense, in addition to direct political offense and communicative political offense.

¹⁷ Interestingly, these three modes mirror the distinction between locutionary, illocutionary, and perlocutionary acts in speech act theory (Austin, 1962).

violates prohibitions, whereas theories like Bolinger's contrastive choice account take the *message communicated by the slurs*, rather than slurs themselves, to be the source of offensiveness (Anderson and Lepore, 2013; Bolinger, 2015). I believe both views are true to an extent. This is because slurs can cause three modes of offense and each of the two views is correct about one of the modes. Theories like Bolinger's are true when slurs cause *communicative offense*, e.g., hearers are offended by a racist's utterance of 'chink' because their deliberate choice of words communicates their contempt toward the Chinese. By contrast, the prohibition theory is true for *direct offenses*. For instance, unintentionally mentioning the N-word in a TV show can cause *direct offense* for violating the social norm that prohibits the N-word, even if the speaker does not intend to express racist messages.

7. Answering the Normative Questions

In addition to the descriptive questions, the violated norm theory also answers the normative questions of offense, including the aptness question (§7.1) and the value question of offense (§7.2).

7.1 Answering the Aptness Question

What are the conditions for taking offense to be apt? The violated norm theory answers Q4) the aptness questions by specifying a necessary condition of the aptness of taking offense. Again, my theory does not rule out the possibility of further conditions.

A necessary condition of apt offense is the *universalizability* of the endorsed norm. According to the violated norm theory, to feel offended is to feel that the norm one endorses is violated. I believe that taking offense is apt only if the violated norm is universalizable, i.e., both the offender and the offended can have reasons to accept the norm. For instance, when a Chinese person is offended by the racist greeting in Japanese, they judge that this greeting violates their principle of racial equality. It is appropriate to feel offended because the violated norm of racial equality is universalizable; everyone (including the offender) can have a reason to accept racial equality (even if they do not actually accept it). For instance, the white offender should accept racial equality for reasons such as that race is morally irrelevant to one's worth, and that the offender would not want to live in a society that discriminates against them because of their race.

Moreover, the universalizability condition explains the cases of inapt or inappropriate offense. Taking offense is often inappropriate because the violated norm is not universalizable; the offender cannot reasonably accept the norm of the offended party. Suppose a coffee purist gets offended because I put sugar and cream in my coffee. My act violates their principle that coffee should be consumed without condiments to appreciate its true flavor. However, their feeling offended seems like an overreaction to other ways of drinking coffee. What makes their taking offense inapt? According to the violated norm theory, this is because other people have no reason to accept their norm of coffee purism. Perhaps the coffee purist has personal reasons against milk and sugar in coffee, e.g., their ability to appreciate coffee, their dislike of sugar, etc. Nevertheless, it is unreasonable for others to accept their high standard, since this is only a matter of taste.¹⁸

Let us summarize the violated norm theory's answer to the aptness question as follows:

- Q4) The Aptness Question: What are the conditions for taking offense to be apt
- A4) Taking offense is apt only if the norm endorsed by the offended person is universalizable.

One of the objections against the universalizability condition is that apt offense can arise from non-universalizable norms like religious prohibitions. Suppose the host offers the guest foods that are against the religious beliefs of the guest. It seems appropriate for the guest to take offense even if her religious prohibition is not universalizable (that is, there is no reason to expect non-believers to accept her prohibition).¹⁹ However, my response is that the universalizability condition is nonetheless met in cases like religious prohibition. After all, there can be universalizable norms about non-universalizable norms. Despite the non-universalizability of religious prohibition, it is a universalizable norm that no one should be offered foods against their religious prohibition. For instance, even the host has a reason to accept this because they would not want to be offered foods against their prohibitions.

¹⁸ The violated norm theory also explains why the coffee purist thinks that their taking offense is appropriate. This is because they take the norm of coffee purism to be universalizable. However, they are mistaken in thinking so, for the reason I have given.

¹⁹ Thanks to an anonymous reviewer for raising this objection.

Toward a Theory of Offense

After specifying the universalizability condition, I will argue that the conditions of apt offense do not include *intentions*. Is it inappropriate to take offense when the offender does not intend to offend? My view is that unintentional offending acts usually do not make taking offense inapt. Consider again the example of consequential offense. A friend of the victim asks them about the experience of being bullied, and the question offends the victim. Even if the friend does not mean to offend, it is still appropriate for the victim to feel offended. This is because what violates the norm does not depend on the intention. Whether the question produces traumatic psychological effects is out of the control of the friend. The same can be said about direct offenses.

There is nevertheless an exception. The aptness of communicative offense does depend on the intention of the offender. This is because the message expressed in communicative offense is often expressed via communicative intentions (also called ‘speaker’s meaning’), which are determined by the intentions of the offender (Grice, 1957).²⁰ Consider the example of getting offended by a bystander’s laughing at the act of bullying. The offended person judges that the laughter sends a message that violates their principle against bullying. Suppose the bystander laughs because they are tickled. This at least makes taking communicative offense less appropriate. ‘No offense. I didn’t mean to laugh’ is a good excuse for the bystander.²¹ If the bystander lacks the relevant communicative intentions, they are not actually sending a message that violates the anti-bullying norm.

Apt offense at unintentional acts is possible because of the possibility of being offended by unintentional acts. If taking offense at unintentional acts were impossible, discussing its aptness would make little sense. The basic idea of the violated norm theory can explain this possibility. According to my theory, to be offended requires one’s judgment that one’s endorsed norm is violated, and norms can be violated regardless of the offender’s intention. For example, a victim can be offended by the request to repeat their traumatic experiences because it violates the norm prohibiting traumatizing victims, no matter what the offender intends.

²⁰ A good example of communicative intentions is someone’s showing Mr. X a photo of Mr. Y’s intimacy with Mrs. X (Grice, 1957, p. 382). The communicative intention of this act is that 1) Mr. X forms the belief that Mrs. X is having an affair; 2) Mr. X recognizes his intention; 3) Mr. X’s belief is at least partially caused by the recognition of his intention.

²¹ Notice that the lack of ill intentions does not excuse other modes of offense. In this example, taking consequential offense is still appropriate because of the negative psychological effects caused by the bystander.

Although unintentional offending acts do not usually make taking offense inapt, the intentions of intentional offending acts do affect the degree of the aptness of taking offense. Suppose someone intentionally asks a victim to repeat their traumatic experiences, in order to protest this norm against traumatizing victims. Taking offense at this is more apt than taking offense at a similar request intended to satisfy one's curiosity without due respect to the norm. Given the complexity of 'intentional' offending acts, there is a spectrum of intentions that give rise to the most apt offenses and the less apt (but still apt) ones. The following is a list of intentional acts in the *descending* order of aptness (of taking offense at them).

1. Intentionally asking traumatizing questions for the sake of violating the norm against traumatizing victims itself.
2. Intentionally asking traumatizing questions because violating the norm against traumatizing victims helps the agenda against PC culture.
3. Intentionally asking traumatizing questions with indifference to the norm against traumatizing victims.
4. Intentionally asking traumatizing questions without knowing if this violates the norm against traumatizing victims.

The violated norm theory helps to explain the difference in their aptness. Despite the varieties of intentions, it is a general rule of thumb that taking offense is more apt when violating the norm is a more dominating reason of intentional offending acts.

Moreover, the violated norm theory helps to address issues in *slurs*, *feminist ethics of humor*, and *the offense principle*. First, it answers the question 'Should we be offended by slurs used without bad intentions?' According to my theory, it can be apt to be offended by slurs used without bad intentions. This is because, although the aptness of communicative offense caused by slurs (e.g., communicating racist messages) depends on intentions, the aptness of their direct offense (e.g., violating the social norm that prohibits slurs) does not. Second, it explains the difference between the two kinds of offense in sexist jokes. The disparagement of women in sexist jokes causes communicative offense. If the speaker does not mean to disparage women, then taking offense is less apt. By contrast, holding sexist beliefs causes a direct offense, which is not excusable by the intention of the offender. Third, this theory allows the offense principle to distinguish between appropriate and inappropriate offense. Civil rights movements should not be legally prohibited from offending white supremacists. This is because the norms of white supremacy are not reasonable to the activists.

The violated norm theory of offense (as mental states) also helps to develop an ethics of offending acts. When it is appropriate to take offense is one of the factors that explain when it is wrong to offend. First, I believe if it is wrong to offend someone, it is appropriate for her to take offense. For instance, if it is wrong to offend the Chinese with racist remarks, it is also appropriate for them to feel offended. Second, the reverse of the principle above is not true; it is not the case that if it is appropriate for someone to take offense, it is wrong to offend her. Imagine a confrontation in which protestors from a marginalized group offend the white supremacists with obscene expressions like F-words. It seems appropriate to be offended by the F-words since no one should be subject to obscenity. However, it may not be wrong to offend the white supremacists. This is probably because they occupy the dominant social status and offense can hardly harm their dignity and self-respect, etc. In summary, a moral psychology theory of taking offense is one of the considerations of an ethics of offending acts.

7.2 Answering the Value Question

The second normative question to be answered is Q5) the value question (i.e., does taking offense have any positive value?). According to the violated norm theory, taking offense does have a certain positive value. Imagine an 'offense-free' possible world, which is an exact duplicate of the actual world. It contains the same offensive things such as hate speeches and discriminations, except that no one ever feels offended by them. For its residents, hate speech feels like mathematical mistakes; they are wrong but there is no need to be emotional about them. Is such an offense-free world a better world than ours? I believe that something is wrong in such a world. The proper reaction is to take offense and do something to stop these injustices. However, the residents' reactions are inadequate; they are not emotionally motivated to defend the norms they endorse.

If taking offense has a certain value, what kind of value is it? The violated norm theory takes the value of offense to be *instrumental*. We can quickly rule out the possibility of the intrinsic value of offense. No one would pursue offense for its own sake.²² Offense is instrumentally valuable because it is a moral state that motivates us

²² Offensive materials like pornography and hate speech are often made, not for causing offense per se, but for other things like pleasures or advancing certain political views.

to defend our norms. It helps to bring out the values that the endorsed norms aim to achieve. If people get offended by hate crimes, they are motivated to stand up to the haters and promote social justice. If someone gets offended by bad aesthetics, they are more likely to do something to promote beauty. Human norms can shape the world because we, as normative agents, are equipped with emotions like offense.

My answer to question Q5) can be summarized as follows:

- Q5) The Value Question: Does taking offense have any positive value? If it does, what kind of value is it?
- A5) Taking offense has a certain positive value. Its value is instrumental to defend the endorsed norms.

8. Conclusion

In this paper, I have shown the need for a theory of offense from many areas in philosophy. For instance, it matters for theories of slurs to explain the offensiveness of slurs. A theory of offense will help the offense principle of legal prohibition by differentiating between appropriate and inappropriate offenses.

Since no adequate theory of offense has been developed, this paper has completed *two tasks* to fill the gap. *First*, I have introduced the descriptive questions and the normative questions that a theory of offense has to answer. An example of the descriptive questions is Q1) the nature question: What is it for a person to feel offended? An example of the normative question is Q4) the aptness question: What are the conditions for taking offense to be apt? *Second*, I have developed a 'violated norm theory' to answer these questions. For instance, the answer to Q1) is that a person feels offended only if the norm they endorse is judged to be violated by the offending act. Appealing to the violated norms also answers other questions like Q4): taking offense is apt only if the norm endorsed by the offended person is universalizable.

This preliminary theory aims at starting a conversation on the nature and the ethics of offense. I hope a more robust theory of offense will arise from objections and contribute to other areas such as slurs, the offense principle, ethics of humor, etc.²³

²³ For helpful discussions and comments, I am grateful to Wei Cheng, Yiwon Zhang, Weitao Huang, Matteo Ravasio, Julius Schönherr, the participants of the CPSSA online workshop, and an anonymous reviewer for the journal.

References

- Luvell Anderson and Ernie Lepore, 'What Did You Call Me? Slurs as Prohibited Words', *Analytic Philosophy*, 54(3) (2013), 350–363.
- J. L. Austin, *How to Do Things with Words* (Clarendon Press, 1962).
- Merrie Bergmann, 'How Many Feminists Does It Take to Make A Joke? Sexist Humor and What's Wrong with It', *Hypatia*, 1(1) (1986), 63–82.
- Renée Jorgensen Bolinger, 'The Pragmatics of Slurs', *Noûs*, 49(1) (2015).
- Michael Clark, 'Humour and Incongruity', *Philosophy*, 45(171) (1970), 20–32.
- Justin D'Arms and Daniel Jacobson, 'The Moralistic Fallacy: On the "Appropriateness" of Emotions', *Philosophical and Phenomenological Research* 61(1) (2000), 65–90.
- Stephen Darwall, *The Second-Person Standpoint: Morality, Respect, and Accountability* (Cambridge, MA: Harvard University Press, 2006).
- Justina Diaz-Legaspe, Chang Liu, and Robert J. Stainton, 'Slurs and Register: A Case Study in Meaning Pluralism', *Mind and Language*, 35(2) (2020), 156–182.
- Joel Feinberg, *Offense to Others* (New York: Oxford University Press USA, 1985).
- Allan Gibbard, *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Harvard University Press, 1990).
- H. P. Grice, 'Meaning', *Philosophical Review*, 66(3) (1957), 377–388.
- Randall Kennedy, *Nigger: The Strange Career of a Troublesome Word* (New York: Vintage Books, 2003).
- Chang Liu, 'Slurs and the Type-Token Distinction of Their Derogatory Force', *Rivista Italiana di Filosofia del Linguaggio*, 13(2) (2019), 63–72.
- Chang Liu, 'Slurs as Illocutionary Force Indicators', *Philosophia*, (2020).
- John Stuart Mill, *On Liberty*. Vol. XVIII, in J. M. Robson (ed.) *The Collected Works of John Stuart Mill* (Toronto and Buffalo: University of Toronto Press, 1977).
- Kathleen Dean Moore, *Pardons: Justice, Mercy, and the Public Interest* (New York: Oxford University Press, 1989).
- Felicity Morse, 'Zwarte Piet: Opposition Grows To 'Racist Black Pete' Dutch Tradition', *Huffington Post UK*, December 6, (2012). https://www.huffingtonpost.co.uk/2012/12/06/zwarte-piet-black-pete-christmas-dutch-netherlands_n_2249412.html (accessed January 19, 2021).
- Martha C. Nussbaum, 'Emotions as Judgements of Value and Importance', in R. C. Solomon (ed.) *Thinking about feeling: Contemporary philosophers on emotions* (New York: Oxford University Press, 2004).
- Norvin Richards, 'Forgiveness', *Ethics*, 99(1) (1988), 77–97.
- Amia Srinivasan, 'The Aptness of Anger', *Journal of Political Philosophy*, 26(2) (2018), 123–144.

CHANG LIU (ch.liu@live.com) is a Boya Postdoctoral Fellow at the Department of Philosophy and Religious Studies at Peking University. His research focuses on social and political philosophy of language, especially slurs. His recent publications include 'Slurs as Illocutionary Force Indicators' (*Philosophia*, 2020).