What should 'safety' mean in the classroom?

Alice Monypenny

Forthcoming in: J. Gatley and C. Norefalk (ed.) Philosophical Analysis for Educational Problems: Engineering and ameliorating educational concepts. Leiden: Brill.

Introduction

The topic of safe spaces in education, particularly higher education, has generated a great deal of debate for a number of years, garnering input from both within and beyond the institutions themselves and touching on topics such as freedom of speech, censorship and the role of the university in contemporary society. Since the early 2010s numerous commentaries on safe space policies and their effects have been published by academics, journalists and student activists. On the one hand, the idea that educators have a responsibility to create safe classroom environments can be found in many initiatives to make education more accessible and inclusive. But on the other hand, many concerns have been raised that safety is antithetical to education's role in challenging deeply held beliefs and encouraging critical thinking and resilience.

What these debates often make clear is that the concept of safety can be used in vague and polymorphous ways that make its practical application or debate about its value or lack of value in the classroom difficult. Robert Boostrom (1998, p. 398) observed how the language of safety and safe spaces in the classroom became common even though the terms themselves had largely been neglected as a subject of educational inquiry. The terms were widely used, he argued, but were undertheorized – very rarely were explanations given as to what a safe space is or justification offered as to why creating a safe space is a good thing (*ibid*.). There has been some definitional work since the time of Boostrom's writing (for example Callan, 2016; Holley & Steiner, 2005; Stengel, 2010; The Roestone Collective, 2014), but still a lack of clarity pervades many conversations about safety. Therefore, entering into this debate requires that we first answer some fundamental questions, including: what does it mean for a classroom to be safe, and is it a good thing, or even necessary, for a classroom to be safe?

Thinking about safety in the classroom often has practical aims which relate to either (1) responding effectively when it is claimed that a particular classroom is unsafe, or (2) improving classrooms in the future. What can sometimes be lost or side-lined is the explicit liberatory roots of the concept of safety in the classroom. The term 'safe space' as it is most commonly used in U.S. and U.K. educational contexts, has its roots in liberatory movements (Kenney, 2001), including the creation of spaces designed for liberation or to protect members of oppressed groups from the harms and dangers of oppression. In the engineering of the concept of safety in the classroom, I intend to retain this connection to liberatory aims.

1 – Why conceptual engineering?

For educators, the most immediate questions will often concern what makes a classroom safe – what conditions are required in order for a classroom to be safe, and how these are achieved. But much of the existing lack of clarity stems in part from a tendency to conflate conditions which create safety with the nature of safety itself. Understanding safety in the classroom fully will require empirical study – be it formal empirical research or the more informal work by individual educators of coming to understand the demands and challenges of their particular students and classroom contexts.

Conceptual engineering plays a vital role in preparing the ground for such work. Otherwise, attempts to understand safety in the classroom can lead to a conflation of questions of 'what' and 'how'. This can be observed in the ways in which we elicit student perspectives on safety. Existing work (Arao & Clemens, 2013; Ellsworth, 1989; Holley & Steiner, 2005; Winans, 2005) on the nature of safety in the classroom, often pays attention to students' perceptions and descriptions of safety.

Understanding what makes a classroom safe for different students in different contexts will require attention to the specific needs and vulnerabilities of individual and groups of students. We might question whether students are well placed to give insight into the concept of safety, especially when they are not trained to think critically about the practice and aims of pedagogy. But despite these worries about the unreliability or ambiguity of student testimonies, there may also be ways in which student testimonies regarding the means of achieving safety are more reliable than educator or third-person perspectives. In many contexts students are likely to be in a better epistemic position than educators to know and understand their own vulnerabilities and therefore may have better insight regarding what must be done to make the classroom safe for them. However, if we are to learn from student perspectives we must elicit them in a way which is not ambiguous and this requires a clear distinction between practical and conceptual questions.

Consider an anecdote from feminist educator and theorist Berenice Malka Fisher in which she reflects on her experiences of addressing the issue of safety in college classrooms in the USA. She recounts an incident in which she explicitly asked her students what 'safety' meant to them:

> After a go-around in which students talk about experiences with violence, Vera, a white and apparently middle-class woman, comments on how white, middle-class assumptions pervade several of the stories. Sarabeth, who is also white, becomes so agitated she can hardly stay in her seat. "I thought this was a safe space," she says, "that we were not supposed to judge each other." Vera leaps to her own defense: "I thought it was safe to say what we think!" I become anxious about my responsibility toward each of the students and toward the class as a whole. "What is 'safety' really about?" I wonder out loud. As we go around the class, Felice says that safety has to do with knowing that I as a teacher will not let things get "out of control." Lourdes comments that she always feels safer when people are speaking Spanish, as she does with her family and friends. Bruce remarks that as the only man in the room he would feel safer if there were "more of us." It is clear that "safety" does not mean the same thing to everyone. (2000, p. 137)

Fisher asks her students a broad question – 'What is safety really about?' – and arrives at the conclusion that 'safety' means something different to each of her students. One way her question can be interpreted (and her reflection on the students' responses suggests that this is her interpretation) is as a question about the nature of safety. In other words, her question is about what it is to be safe. However, it is not so clear that we can interpret all of the students' answers in the same way. When they report that safety is about knowing that the teacher is in control or that they feel safer when Spanish is being spoken, they may be describing conditions which create safety or make them feel safe.

It is possible that some or all of the students in Fisher's classroom held the same understanding of the nature of safety and are reporting the conditions, which for them, appear most important and relevant to achieving that end. For instance, Felice, Lourdes and Bruce could all understand a safe classroom as one in which participation does not carry significant risk. The answers they give could reflect the conditions which for them lower the stakes of participating in the classroom: knowing that the teacher will step in to defuse hostility; being able to speak in a language that they feel more comfortable and able to express themselves fully; and not feeling vulnerable to being singled out on the basis of their gender. The differences in the students' answers could reflect differences in their sensitivity and vulnerability to different sorts of risk.

Another example in which practical and conceptual questions are intertwined is in Holley and Steiner's (2005) often cited study investigating understandings and perceptions of safety in undergraduate and postgraduate level social work classrooms. Students were asked what they perceived to contribute to safe and unsafe classrooms. One of the stated aims of Holley and Steiner's study is to investigate the ways in which safe classrooms *are created* (*ibid*., p.50). However, Holley and Steiner also state that the study aims to investigate what a safe classroom *is* (*ibid*.), whilst also exploring the roles of instructors, students and physical classroom environments in 'creating an environment that encourages honest and open dialogue' (*ibid*.).

Furthermore, their survey made no explicit mention of safety. As they report on their method of data collection, Holley and Steiner imply that the questionnaire used contained no explicit mention of 'safety'. Instead, questions were framed in terms of how willing and able students felt in raising controversial ideas or sharing personal experiences related to the course content. For example:

> The third section of the questionnaire asked questions to determine the importance that students place on a safe classroom environment…. [T]hey were asked to rate the general importance of creating an environment where they could honestly express themselves. They were then asked to indicate… if such a classroom environment changed what they learned (*ibid*., p.53)

At most, the results of the questionnaire can tell us about the conditions that promote participants' feelings of safety when they understand a safe classroom to be one in which they can honestly express themselves. But, if the aim of the study is to investigate what a safe classroom is, then we cannot take this for granted. As shown in Fisher's anecdote, there are good reasons to think that students may have different understandings of what 'safety' means. Therefore, we cannot conclude from Holley and Steiner's study that the conditions

which created classrooms where students were able to honestly express themselves are the conditions which create classrooms which are safe (see also Redmond, 2010).

In order to do these kinds of work effectively and come to understand how to create safe classrooms, we must first settle the question of what it means for a classroom to be safe. This task is one of conceptual engineering for at least two reasons. Firstly, we cannot merely look to current uses of the term 'safe' in order to determine a single consistent account of the concept, since the current usage is varied and inconsistent. Secondly, educational concepts are useful tools which can be designed to serve some purpose. As I will argue below, the concept of safety can be designed in such a way as to serve the purpose of tackling oppression within education.

2. Methodology: Ameliorative Analysis

Now that I have argued that a conceptual engineering of the concept of safety is needed, I will describe the methodology by which this should be done. The conceptual engineering I undertake here is ameliorative (Haslanger, 2012): it considers what we want the concept to do for us, and how it might best serve our purposes. The first step, then, is to identify our purpose in engineering the concept of safety. Broadly, the aim of such a conceptual engineering project is to produce concepts that are useful tools for the creation of effective educational systems. The purpose of my engineering of the concept of safety is a little more specific. The term 'safe space' arose in the context of liberatory movements, specifically the creation of spaces that served a liberatory function. Oppression is dangerous: it threatens and causes harm, it escalates risk, and it can limit members of oppressed groups to mere surviving rather than flourishing. This is also true within education. In order to create liberatory classrooms, in which all students are equal, are respected and have equal opportunities and abilities to flourish, we need to take seriously the dangers of oppression. The concept of 'safety' within the classroom is a tool to recognise the measures which must be taken in order to respond to and overcome the dangers of oppression. That being said, it is worth noting that like with many liberatory concepts this project of conceptual engineering and the applications of its output may have much wider benefits, including for those who are not members of marginalised groups. After all, there may be other kinds of dangers which are not rooted in oppression but which may still have negative impacts in the classroom.

The term 'unsafe' plays a particular role within the discourse. When we describe something as 'unsafe' we indicate not just dislike, but some sort of threat - that it places us or others at risk. The language of safety is important in characterising barriers which impact students' ability to access the classroom and take full advantage of the learning opportunities within it. It demonstrates how such tools and practices are necessary because without them students are at risk and may be unjustly excluded from the classroom (Carter, 2015). They are not just preferences or things which make students feel more comfortable.

There are two important points to address: (i) to what extent should our engineering of the concept be revisionary; and (ii) should the resulting concept be normatively loaded? Firstly, there is a great deal of ambiguity in the way in which the term 'safety' is used to describe classrooms, something partly inherited from the messy nature of the public debate around the issue. Therefore, any attempt to spell out the concept in a consistent and practicable way will involve some revision by designating some current uses as incorrect or non-literal. However, the aim is not to revise completely current usage of the concept but rather to

preserve some of the core roots of its meaning. As already noted, the aim is to aid liberation and there is something important captured when the terms 'safe' or 'unsafe' are used.

This links to the second point: when we describe a classroom as 'safe' or 'unsafe' are we, by definition, describing something of value, and of disvalue, respectively? To meet the liberatory aims set out above, the concept should be engineered in such a way that it indicates an evaluative judgement. The concept of safety helps us to make an evaluation of classrooms in terms of their role in addressing and overcoming the dangers of oppression.

The conceptual engineering of the concept will proceed as follows. The first step is to identify the value(s) being indicated when a classroom is described as safe. Then, we must determine whether these values that have been identified are genuinely valuable in the classroom. If they are, we can then define the concept so that it best captures those values.

3. The value of safety

The first step is to identify the value being indicated when a classroom is described as 'safe'. Discussions of safety have often occurred in contexts in which safety is lacking or is under threat, and as a result, we can learn a lot about safety from looking at instances in which it is lacking. Therefore, I'll focus on the disvalue which is indicated when a classroom is described as 'unsafe'.

As an example of the ways in which 'unsafe' is used to describe a classroom, consider the following passage in which Elizabeth Ellsworth (1989) reflects on her experiences teaching an undergraduate course on Racism in the Media. Ellsworth observes that her classroom, at least to begin with, was not a safe space in which students felt able to speak openly and honestly about their experiences. She speculates about the reasons for this:

> 'Our classroom was not in fact a safe space for students to speak out or talk back about their experiences of oppression both inside and outside the classroom. […] Things were not being said for a number of reasons. These included fear of being misunderstood and/or disclosing too much and becoming too vulnerable; memories of bad experiences in other contexts of speaking out; resentment that other oppressions (sexism, heterosexism, fat oppression, classism, anti-Semitism) were being marginalized in the name of addressing racism – and guilt for feeling such resentment; confusion about levels of trust and commitment surrounding those who were allies to another group's struggle; resentment by some students of colour for feeling that they were expected to disclose "more" and once again take the burden of doing the pedagogical work of educating White students/professor about the consequences of White middle-class prejudice; and resentment by White students for feeling that they had to prove that they were not the enemy.' (1989, pp. 315–316)

Interpreting Ellsworth's comments here, we can separate out a number of different ways in which things go wrong in her classroom. Firstly, her students believe, perhaps rightly or perhaps not, that should they speak out they will face negative consequences – being misunderstood or becoming too vulnerable. They protect themselves by remaining silent. By speaking out, they believe that they would relinquish this barrier and be forced into a state of vulnerability.

The extent to which students are actually at risk of harm – what I will call *objective risk of harm* – and the extent to which they perceive themselves to be at risk of harm – what I will call *subjective risk of harm* – may come apart. For example, it is not difficult to imagine a scenario in which students would not face any backlash were they to speak out, but because of a lack of trust, the students still end up believing that they are at risk of harm. In such a case, students are not at risk of harm but perceive themselves to be. On the other hand, it is also plausible that students could be at risk of hostile backlash when speaking out without being aware of this risk. It is worth noting here that I am understanding harm in a broad sense – including physical, psychological, emotional, and social harms. Given the liberatory aims of this project, these harms will also include unjust exclusions and disadvantages which hinder students' learning and have an impact upon their educational outcomes.

Furthermore, the task of speaking out elicits a range of negative emotional responses from the students, including guilt, resentment, and confusion. These emotional responses can be distressing or draining and use up mental resources. The classroom context, with its diverse students, low levels of trust and situation within a larger institution, enables or perhaps encourages the students to have these negative emotional reactions when invited to speak openly and honestly.

So, we can identify three potential different 'disvalues' that are being indicated when Ellsworth's classroom is described as 'unsafe': a high level of objective risk; a high level of subjective risk; and the production of negative emotional responses. The next step is to establish whether each of these three are genuinely and always disvaluable in the classroom.

Risk in the classroom: Objective and Subjective

A high level of objective risk of harm is disvaluable because of the increased likelihood of students being harmed. This disvalue is independent of students' subjective level of risk. The disvalue of objective risk of harm will depend on the severity of the harm and the level of risk. There are many risks of harm that we tolerate everyday, either because the harms are so mild or the risk is so minimal. But over a certain threshold, the risk of harm is disvaluable.

A high level of subjective risk of harm can also be disvaluable, independent of the level of objective risk of harm. In Ellsworth's classroom, the feeling of being at risk of harm prevented students from speaking openly and honestly. If students believe that they are at risk of harm then they may pre-emptively act to protect themselves, especially by withdrawing from activities or becoming defensive. If Ellsworth wants her students to be able to speak openly and honestly, then she needs to create a classroom environment in which students feel that they are safe.

But what about cases in which students unjustifiably (from an outside perspective) perceive themselves to be at a high level of risk of harm? For example, imagine a student who has a severe phobia or perceives themselves to be at significant risk of harm from something which is not present, or does not genuinely pose a threat to them – such as a student who believes without good reason that there is a high level of risk that she will be harassed by her peers. In such cases, the high level of subjective risk is disvaluable only in so far as it has concrete negative effects on the student – such as causing her distress, leading her to attempt to protect herself in ways that cause her harm, or prevent her from fully participating in or making the most out of the learning activities in the classroom. Note that this does not entail that the appropriate response would be for an educator to try and reduce the risk of the student being harassed. Instead, the risk of harm would be reduced by addressing the

unjustified fear itself or, if appropriate and possible, removing things which prompt the student to believe she is at risk of harassment. Compare this to a case in which there is a high level of subjective risk but no negative impact on the student: a student falsely believes that they are at a high level of risk of harm but this does not cause them to feel any fear or distress, nor does it lead them to change their behaviour in any way. Thus, the high level of subjective risk has no negative impact on the student. If this is the case, then the high level of subjective risk is not disvaluable.

The disvalue of high levels of subjective risk therefore comes from the negative effects on students. More specifically, it comes from the *harms* caused by a high level of subjective risk. These include the emotional and psychological harms arising from feelings of distress, physical harms that may be caused by pre-emptive protective measures taken by the student, and harms caused by the disadvantaging effect of not being able to participate fully in and make the most of the learning activities of the classroom. If a high level of subjective risk does not cause any harm to students, then it is not disvaluable in the classroom.

To support this argument, consider the account of safety developed by Eamon Callan (2016). To address concerns about the relationships between safety, freedom of speech, and the ability to confront students with challenging views in the classroom, Callan narrows in on the nature of the safety with which we are concerned in the classroom. He distinguishes between 'intellectual safety' and 'dignity safety'. To be intellectually safe is to be protected from having one's deeply held beliefs and commitments challenged in a way that may be disorienting or even distressing (ibid. p.65). This kind of safety, Callan argues, is not to be expected in the classroom. Intellectual growth sometimes requires the challenging of deeply held beliefs, and therefore is an important part of education.

On the other hand, the classroom is 'dignity safe' for some student or group of students when they can 'participate without reasonable worries that they are likely to be humiliated by others' (ibid, p.67). Callan is clear that by 'humiliation' he does not refer to the feeling of indignation or distress caused by another's actions. Instead, to be humiliated in Callan's sense is to be treated in a way that implies that one has no valid claim to be considered equal (ibid.). Note that through this distinction, Callan demonstrates how the safety of the classroom is not dependent on how the conduct of others makes students feel regardless of whether that conduct is being misrepresented or misinterpreted. The only response on the part of the student which determines the level of safety is their belief about the extent to which they are at risk of being humiliated (the level of subjective risk of harm).

By incorporating the notion of a 'reasonable' worry, Callan pre-empts the concern raised above regarding subjective accounts of safety – namely, that a students' level of subjective risk of harm may be entirely detached from reality or irrational and therefore place an impossible burden on educators in reducing it. The subjective level of risk is determined by a students warranted beliefs regarding the extent to which they are at risk of humiliation.[1]

What Callan's account appears to show is the potential for a subjective account to be adapted to avoid the worry mentioned above. However, Sigal Ben-Porath's development of the account demonstrates that the disvalue of a high subjective risk of humiliation can be

[1] We might think that by placing this constraint on student's beliefs, Callan's account actually constitutes a 'hybrid' account of safety: the level of safety is determined by a combination of students' beliefs regarding the extent to which they are at risk and some facts about the relationship between those beliefs and the objective level of risk. This would depend on exactly how we were to spell out 'reasonable'.

explained in terms of objective risk of harm. Ben-Porath (2016) argues that dignity safety is an issue of access. A lack of dignity safety can, wholly or partially, prevent students from being able to fully take advantage of learning opportunities (*ibid*., p.80). A student who faces constant attacks on their dignity safety – by witnessing physical and verbal assaults on people like them or feeling vulnerable to violence and abuse, or being surrounded by messages that people like them do not belong in the institution or have equal status in the community – faces the wearing task of constantly responding to and resisting threats. This can be extremely draining and, especially when it goes unacknowledged, can direct attention away from full and fruitful engagement in learning. Thus, we care about dignity safety in education because when a classroom is not dignity safe for a student, they are at a greater objective level of risk of being prevented from taking full advantage of the learning opportunities in the classroom (and therefore being disadvantaged in their education).

Negative Emotional Responses

I now turn to the third disvalue we can see in Ellsworth's unsafe classroom. When asked to speak openly and honestly, the students had a range of negative emotional responses – feelings of guilt or resentment, feeling burdened with the task of educating others, or anxious that they must 'prove they are not the enemy' (Ellsworth, 1989, p. 316). Even if a student chooses not to speak openly and honestly and therefore feels protected from hostility, she may still encounter feelings of guilt and resentment. Simply being invited to speak openly about their experiences of oppression, in that context, may lead some students to engage in a costly mental process that involves weighing risks and benefits.

An individual student may have this kind of response as a result of general or specific anxieties or worries. Perhaps they have had bad experiences in the past, or they are generally not trusting of others or of their peers in particular. Perhaps they experience social anxiety and speaking in the classroom, especially speaking about their own experiences, heightens that anxiety.

However, students may face these questions as a result of the context of the classroom and the subterranean power structures and social patterns that operate within it. Take for example, the feeling of resentment from students of colour that they must educate their white peers about their experiences of racism whilst getting little benefit from the discussion themselves. The choice of whether to speak openly and honestly for students of colour is likely to be more fraught in classrooms in which many of their peers are relatively ignorant about the effects of racism and the educator does little to prevent ill-informed views from dominating the discussion. They may, for instance, have to consider whether to and how to speak, who will benefit, who their testimony might reach and how it might be interpreted, while questioning whether it is fair that they take on the task of 'educating' white students – especially when doing so might open them up to harm and bring few, if any, positive effects. Furthermore, invitations can be experienced by those who are relatively powerless as demands when coming from those in positions of power. That is, whilst the educator assumes that students have the option to stay silent if speaking out is too risky for them, the students themselves may not perceive that they are genuinely being offered that option.

The disvalue of negative emotional responses in the classroom (and likewise, the classification of such a classroom as 'unsafe') is disputed. For example, Brian Arao and Kristi Clemens (2013) recount an episode during which students were given an activity designed to demonstrate

differences in privilege and opportunity. The students responded negatively to the activity and challenged its worth in the classroom. Some students reported that they felt 'persecuted, blamed, and negatively judged', guilty about their privilege, and helpless when recognising the lack of privilege and opportunity others experienced. Others reported that the activity was a 'painful reminder of the oppression and marginalization they experience on a daily basis' and that they were placed in the role of educators for other students which caused them to feel angry and 'sorrowful' (*ibid*., p.137). Many of the students objected that the 'profound feelings of discomfort many of them experienced were, in their view, incongruent with the idea of safety.' (*ibid*.)

Arao and Clemens challenged their students' resistance to experiencing these negative emotions on the grounds that sometimes such experiences are necessary parts of the challenging and risky task of engaging with and acknowledging one's own privilege or sharing experiences of oppression (*ibid*., pp.139-140). Interestingly, this leads Arao and Clemens to question the value of safety when attempting to have honest discussions about issues of social justice (*ibid.*, p.139). However, this jump is too quick and instead we should question whether the negative emotional responses experienced by the students in this example do constitute the disvalue identified when a classroom is described as unsafe.

It's worth separating two worries here. The first concern relates to members of dominant groups (mis)using the concept of safety as a means to maintain their privilege by labelling uncomfortable acknowledgements of their privilege as 'unsafe' (Arao & Clemens, 2013, p. 140; Ludlow, 2004). Students may misidentify uncomfortable feelings and experiences with a lack of safety. In some cases the classification of such feelings as a lack of safety may be used to silence or sanitise discussion which challenges or undermines privileged viewpoints, or as a means to abdicate responsibility for the consequences of their speech. Being held accountable for the consequences of one's speech can be said to bring about negative emotional reactions such as guilt and embarrassment. Recall Sarabeth, one of the students in Fisher's classroom who said, "I thought this was a safe space… that we were not supposed to judge each other." Such an appeal to safety casts being 'judged' – that is, having her 'white, middle-class' assumptions pointed out – and the negative emotional response that it prompts as inimical to safety.

Secondly, there is the more fundamental concern that the kinds of negative emotional responses which are classed as unsafe cannot be excluded from a genuinely liberatory classroom. Boostrom (1998, p. 405) argues that the term 'safe space' when used to describe a classroom has the tendency to be conflated with a space which is without stress. 'Stress' itself is a vague concept which can incorporate both deep existential anxiety and vague discomfort. But such a broad notion casts as unsafe a whole range of feelings which will inevitably arise if students are to be made aware of their own ignorance and pushed to reflect critically on their perspective. As Boostrom writes: 'Understood as the avoidance of stress, the 'safe space' metaphor drains from classroom life every impulse toward critical reflection.' (ibid., p.406).

Similarly, as part of her 'pedagogy of discomfort' Megan Boler (1999) emphasises the need for students, as well as educators, to sit with and examine their emotional responses (including the difficult and painful ones) in order to uncover and challenge their deeply held beliefs, habits and values. What we see and pay attention to (and more importantly, what we do not see or pay attention to) is partly guided by emotions such as fear. Recognising what

we do not see requires that we are willing to confront and sit with this fear (*ibid*., p.182). Boler writes:

> The aim of discomfort is for each person, myself included, to explore beliefs and values; to examine when visual "habits" and emotional selectivity have become rigid and immune to flexibility; and to identify when and how our habits harm ourselves and others. (*ibid*., p.185)

It is not only critical reflection that may be ruled out. As Williams (this volume) notes, teachers may find there to be a tension between their commitment to achieving certain academic standards and promoting the emotional well-being of students because of the negative emotional impact which criticism may have on students.

What both of these concerns highlight is that at least some negative emotional responses in the classroom are not disvaluable (and may even be valuable). However, as noted above there are some instances which are disvaluable. What is needed, therefore, is a means of drawing a distinction between those which are and are not disvaluable.

To see how we might draw this distinction, consider two different examples. Firstly, imagine a student who experiences some embarrassment and guilt when it is pointed out that he is ignorant of some of the ways in which women adapt their behaviour to protect themselves from male violence. This embarrassment and guilt are to some degree unpleasant for the student at the time but does not have any lasting negative impact on him (and hopefully has the positive effect of prompting him to reflect on his male privilege). Compare this to a student who feels humiliated and extremely guilty when she is aggressively challenged by her peers for using incorrect terminology when discussing gender identity due to her lack of familiarity with the topic. This experience is genuinely distressing and harmful to the student and has a lasting impact on her confidence in discussing this topic. The negative emotional responses are disvaluable in the latter case but not the former. They may even have educational value in the former case. The difference between the two cases is the harm caused to the student in the latter case.

The distinction therefore should be drawn between negative emotional responses which are genuinely harmful and those which are not. Of course, drawing this line is by no means an easy task, in part because it will involve answering complex questions such as what constitutes a psychological or emotional harm. I will not attempt to answer these questions in this chapter, but which negative emotional responses will cause psychological or emotional harms will vary depending on the particular vulnerabilities of individual students. Limiting the connection between negative emotional responses and a lack of safety in this way alleviates the second worry raised above. Even when the classroom is a place for critical discussion and challenging privilege, students should not be at genuine risk of harm. The kinds of negative emotional responses which cause harm are still disvaluable.

The disvalue (if any) of negative emotional responses comes from their causing of harm: those which cause harm are disvaluable and those which do not cause harm are not disvaluable. As shown in the previous section, the same can be said for high levels of subjective risk: it is disvaluable insofar as it causes harm. Therefore, the disvalue being identified when a classroom is described as unsafe is, at the base level, the occurrence or objective risk of harm. Likewise, the value identified when a classroom is described as safe is, at root, the absence of harm and the low level of objective risk of harm.

Building an account of safety

I have argued in the previous section, that there are three candidates for the kind of disvalue which are captured by the use of the term 'unsafe' to describe a classroom: a high level of objective risk, a high level of subjective risk, and negative emotional responses. I argued that both a high level of subjective risk and negative emotional responses are not always disvaluable, and when they are they cause harm.

I will now turn to outlining an account of safety in the classroom which captures this value. I will not have the space here to fully develop the account, but will sketch an outline and highlight the factors which will feed into this definition.

> Safety[1]:   A classroom is safe for a student S when S is not harmed and there is a low level of objective risk that the student will be harmed as a result of their presence in the classroom or participation in the classroom activities.

This definition of safety allows us to capture the various values of safety. Firstly, it straightforwardly captures the disvalue of a high level of objective risk of harm. Secondly, it captures the disvalue of the cases of high subjective risk of harm and negative emotional responses because those which are disvaluable are those which cause harm.

This account is by no means complete: there are details to fill out. Conceptual analysis can only get us so far and other methodologies are required to fill in these gaps. For example, we will need to determine the threshold level of objective risk at which the classroom becomes unsafe. The level of objective risk present in the classroom and what kinds of subjective risk and negative emotional responses cause harm will be relational to each student. Likewise, the level of objective risk of harm which a student is able to tolerate may also vary. Therefore, the safety of the classroom should always be thought of as relational to a particular student. A classroom which is safe for one student, may not be safe for another.

Understanding the nature of safety will require consideration of its relationship to protection. The level of objective risk of harm will vary depending on the protective measures put in place. These might include equipment to protect against physical harm as well as skills, practices, character traits, behaviours and ways of thinking that protect against physical, emotional, psychological and social harms. Protective measures may be implemented by educational institutions, educators or students themselves. When concerns around safety arise, they often highlight a lack of sufficient protective measures. Often, the claim that a classroom is unsafe for some student or students is a call for the educator or educational institution to implement greater protective measures. In some instances, the institution, members of the institution, or even outsiders, respond by objecting that the institution and educators are not responsible for putting in place the requested protective measures. Instead, the responsibility falls on students to protect themselves.

We might think that, if protective measures are available for a student to implement herself and this would significantly reduce the risk of harm then the classroom will be safe for that student. However, think back to Ellsworth's classroom in which the classroom was unsafe, in part, because students needed to stay silent to protect themselves from the negative consequences of speaking out. In a sense, there was a means of protection available to the students: staying silent. But this did not suffice to make the classroom safe. This is because staying silent had a significant negative impact on the students' learning experience in the

classroom. They could not fully participate in the discussion. Therefore, if safety is to be a useful tool for making the classroom more inclusive and accessible, then a safe classroom cannot be one in which students must protect themselves in ways which exclude them from the classroom or prevent them from taking full advantage of the learning activities in the classroom. Neither will a classroom be safe if students are required to protect themselves in ways which cause or make them vulnerable to other harms.

This leads to a modification of the account:

> Safety[2]: A classroom is safe for a student S when a S is not harmed or there is a very low level of objective risk (or no objective risk) that the student will be harmed as a result of their presence in the classroom or participation in the classroom activities without S having to protect themselves in such a way that makes them vulnerable to further harms, or significantly disadvantages them in their learning.

Despite these complications and yet to be answered questions, the proposed account is a starting point which makes considerable headway in helping us to think about the requirements of making a classroom safe. Firstly, it provides a response to the concern that safety undermines valuable teaching experiences. Negative emotional responses – such as feelings of stress and discomfort which may play an educational role – do not make a classroom unsafe. That is, unless they genuinely harm the student, at which point they no longer play an educational role.

Secondly, it gives us an outline of the criteria by which we adjudicate whether a classroom is genuinely safe or unsafe. When a classroom is described as unsafe, we can test that claim by determining whether harm is caused by negative emotional reactions or a high level of subjective risk, or if there is a high level of objective risk. We can thus screen out misuses of the term 'unsafe' to maintain privilege as described above.

What I have demonstrated in this chapter is an ameliorative conceptual engineering of the concept of safety in the classroom. What makes this methodology ameliorative in Haslanger's sense is that I have identified the role which the concept of safety can play for us - identifying a condition that allows students to participate fully and effectively in the learning activities of the classroom and is therefore valuable (or even necessary) for the project of creating classrooms that are accessible and beneficial for all - and then engineered the concept to best play this role. By considering existing usage in light of liberatory aims, I have identified the valuable condition which is to be captured by the concept of safety. Then I have used the identified value to begin a re-engineering of the concept. This re-engineering is revisionary to some extent - it will require modifications to current usage of the term - but it captures the values indicated through current usage of the term which is motivated by liberatory aims.[2]

References

Anderson, D. (2021). An Epistemological Conception of Safe Spaces. *Social Epistemology*, *35*(3), 285–311. https://doi.org/10.1080/02691728.2020.1855485

Arao, B., & Clemens, K. (2013). From Safe Spaces to Brave Spaces: A New Way to Frame Dialogue Around Diversity and Social Justice. In L. M. Landreman (Ed.), *The Art of Effective Facilitation: Reflections from Social Justice Educators* (pp. 135–150). Stylus Publishing.

Ben-Porath, S. (2017). *Free Speech on Campus*. University of Pennsylvania Press.

Berenstain, N. (2016). Epistemic Exploitation. *Ergo*, *3*(22), 569–590.

Bergamini, M. (2018). *Are "Safe Spaces" In Universities Breeding Intolerance?* The Huffington Post. https://www.huffingtonpost.co.uk/entry/are-safe-spaces-in-universities-breeding-intolerance_uk_5aa6871be4b06b6a749ce3c4

Boler, M. (1999). *Feeling Power: Emotions and Education*. Taylor & Francis.

Boostrom, R. (1998). 'Safe spaces': Reflections on an educational metaphor. *Journal of Curriculum Studies*, *30*(4), 397–408.

Bouattia, M. (2018). *The Narrative On University Safe Spaces And No-Platform Policies Couldn't Be Further From The Truth*. The Huffington Post. https://www.huffingtonpost.co.uk/malia-bouattia/free-speech-campus-safe-spaces_b_14678920.html

Callan, E. (2016). Education in Safe and Unsafe Spaces. *Philosophical Inquiry in Education*, *24*(1), 64–78.

Carter, A. M. (2015). Teaching with Trauma: Trigger Warnings, Feminism, and Disability Pedagogy. *Disability Studies Quarterly*, *35*(2).

Ellsworth, E. (1989). Why doesn't this feel empowering? Working through the repressive myths of critical pedagogy. *Harvard Educational Review*, *59*(3), 297–324.

Fisher, B. M. (2000). *No Angel in the Classroom: Teaching through Feminist Discourse*. Rowman & Littlefield Publishers.

Haslanger, S. (2012). *Resisting Reality: Social Construction and Social Critique*. Oxford University Press.

Holley, L. C., & Steiner, S. (2005). Safe Space: Student Perspectives on Classroom Environment. *Journal of Social Work Education*, *41*(1), 49–64. https://doi.org/10.5175/JSWE.2005.200300343

John, S. (2011). Security, Knowledge and Well-being. *Journal of Moral Philosophy*, *8*, 68–91.

Kenney, M. R. (2001). *Mapping Gay L.A.: The intersection of Place and Politics*. Temple University Press.

Ludlow, J. (2004). From Safe Space to Contested Space in the Feminist Classroom. *Transformations: The Journal of Inclusive Scholarship and Pedagogy*, *15*(1), 40–56.

Palfrey, J. (2017). *Safe Spaces, Brave Spaces: Diversity and Free Expression in Education*. The MIT Press.

Pells, R. (2017). *University "safe space" policies levase academics in fear of losing their jobs, claims professor*. The Independent. https://www.independent.co.uk/news/education/education-news/university-safe-

spaces-academics-professors-fear-lose-jobs-students-free-speech-political-correct-pc-dennis-hayes-a7815991.html

Ramsay, P. (2017). Is Prevent a safe space? *Education, Citizenship and Social Justice*. https://doi.org/10.1177/1746197917693022

Redmond, M. (2010). Safe Space Oddity: Revisiting Critical Pedagogy. *Journal of Teaching in Social Work*, *30*(1), 1–14.

Roth, M. S. (2019). *Don't Dismiss "Safe Spaces."* The New York Times. https://www.nytimes.com/2019/08/29/opinion/safe-spaces-campus.html

Stengel, B. S. (2010). The Complex Case of Fear and Safe Space. *Studies in Philosophy and Education*, *29*, 523–540.

The Roestone Collective. (2014). Safe space: Towards a reconceptualization. *Antipode*, *46*(5), 1346–1365.

Whitten, S. (2018). *Why "safe spaces" at universities are a threat to free speech*. The Conversation. https://theconversation.com/why-safe-spaces-at-universities-are-a-threat-to-free-speech-94547

Wilson, J. W. (2018). *I've never had a student ask for a safe space. Here's what they ask for.* Vox. https://www.vox.com/first-person/2018/12/12/18131186/college-campus-safe-spaces-trigger-warnings

Winans, A. E. (2005). Local pedagogies and race: Interrogating white safety in the rural college classroom. *College English*, *67*(3), 253–273.