

# *Expression, Tolerance, and Autonomy*

**Peter Godfrey-Smith**

University of Sydney

This is a revised version of a talk I gave at the University of Sydney and UC Berkeley in 2025. It has a companion paper, "Tolerance and Freedom of Expression," which was given as the Sir Karl Popper Memorial Lecture at the LSE in that year (see the front page of my website, [petergodfreysmith.com](http://petergodfreysmith.com), for links). The two overlap quite a lot, but this one has more detail on the philosophical side and new material in the second half.

## **1. Introduction**

## **2. The Hierarchical Model of Tolerance**

## **3. The Hierarchical Model and Expression**

### **3.1. "Academic Freedom"**

## **4. Justifying Policies**

## **5. Pragmatic Balancing**

## **6. Summary**

## **1. Introduction**

The main topics of this paper are freedom of expression, tolerance, and their links to autonomy, which I take to be a form of personal freedom of choice. I'll defend some arguments that use the idea of autonomy to support tolerance, of some kinds, in the context of expression of ideas.

It will take a while to get to those arguments, as they will be presented within a framework developed, in collaboration with Ben Kerr, to address toleration in general. The framework explicitly distinguishes *levels* of toleration. Kerr and I think this is needed to handle questions about "toleration of the intolerant" – that was our initial motivation, at another time some years ago when questions of toleration were politically pressing.<sup>1</sup>

---

<sup>1</sup> The earlier paper: Godfrey-Smith and Kerr, "Tolerance: A Hierarchical Analysis," *Journal of Political Philosophy*, 2019. <https://doi.org/10.1111/jopp.12178>. We wrote the paper around the

We looked a little at freedom of expression back then, but the cases we discussed in detail were "private" behaviors, such as choice of sexual partner. The framework we developed does not contain arguments for more, or less, toleration. It was intended to be neutral in that respect, though it does help in the formulation of replies to some criticisms of "liberal" attitudes. I'll outline the framework and then embed questions about free expression within it. As before, the embedding does not lead directly to ethical or political conclusions. Then I'll connect this framework to a family of autonomy-based arguments that *do* push in the direction of expansive tolerance in the area of expression. Here I use an old paper by Tim Scanlon (1972), mostly now disowned by him, especially in relation to the ideas I want to use.

This area is one where a powerful role is played – in many people, anyway, including myself – by political instincts, an orientation to practical issues, and perhaps what William James would have described as philosophical temperament. It is hard to discuss these issues without some intrusion from those factors and I don't claim to. I'll occasionally try to indicate where, in the sequence of arguments, these levers attach. Ben Kerr should not be taken to agree (or disagree) with the ideas in the second part of the paper.

## **2. The Hierarchical Model of Tolerance**

Here is the framework developed with Kerr. Our initial motivation was the problem of "toleration of the intolerant." Karl Popper had discussed this in *The Open Society and its Enemies* (1945) as involving a "paradox." If a tolerant society is tolerant even of the intolerant, its tolerance risks destruction at their hands.<sup>2</sup> So we apparently have to compromise our tolerance. John Rawls, in *A Theory of Justice* (1971), took a similar attitude: we can be intolerant of the intolerant when they pose a danger to our society,

---

time of the Charlottesville far-right demonstrations in 2017. We published a less academic version as well: <https://theconversation.com/after-charlottesville-how-we-define-tolerance-becomes-a-key-question-83793>.

<sup>2</sup> *The Open Society and its Enemies*, Vol 1., p. 226. See a companion paper to this one, "Tolerance and Freedom of Expression," for a bit more about Popper. Links can be found near the top of the page here: <https://petergodfrey.com>. That talk is also in podcast form.

though this is a compromise or imperfection in our tolerance.<sup>3</sup> But it's not a compromise or imperfection if we think about tolerance in a more structured way, a hierarchical way.

We can start with a behavior that is largely *private*, like choice of sexual partner. Assume a context where everyone involved is a consenting adult. (This will be assumed in all my discussions of sex.) People make their choices. You may or may not be tolerant of these choices; you'll be tolerant of at least some of them, but perhaps not all. This is a choice about *first-order tolerance* – tolerance or intolerance of those private behaviors.

What is "intolerance" here? Kerr and I understand this as interference. Tolerance is non-interference. I'll come back later to what this amounts to.

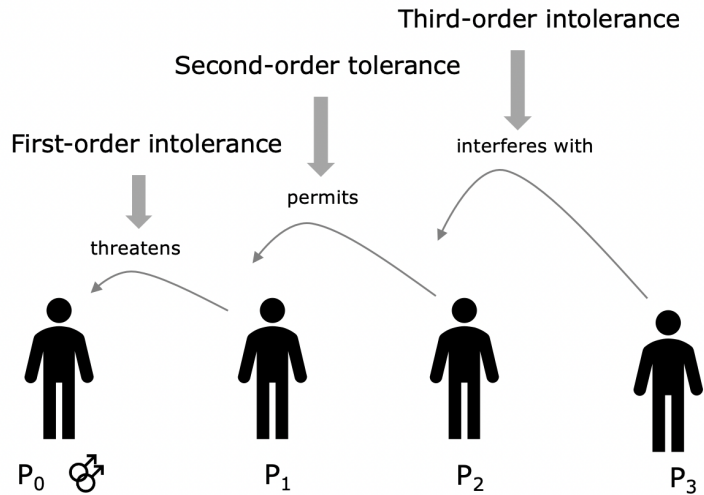
So far we have the idea of first-order tolerance. You can also stand back and look at *those* choices, choices of first-order tolerance and intolerance. Are you tolerant or intolerant of all of these? Some or them? Which ones? Those are choices about second-order tolerance.

There are also third-order choices, and so on up a hierarchy (see the figure on the next page). Third-order choices do have practical importance. Suppose someone you know tolerates intolerance, within their circle, of gay people. Do you "let that go", or not? After that level, things become more rarified.

In the case of sexual orientation, a particular combination of policies has been installed, with much effort, in secular liberal societies. This is a combination of first-order tolerance – adults can enter into whatever consensual sexual relationships they like – and second-order intolerance of some kinds of first-order intolerance. That is part of the force of anti-discrimination laws and the like. We *require* tolerance, in a sense. Specifically, we do not permit interference in people's private choices. If you do seek to interfere in these choices, the state will interfere with *you*. I'll describe policies of this kind, with some second-order enforcement of first-order tolerance, as a "familiar liberal combination".

---

<sup>3</sup> Rawls argues first that an "intolerant sect" cannot complain if it is denied equal liberty, as a person's "right to complain is limited to violations of principles he acknowledges himself". That does not mean that the majority has the right to "suppress" the intolerant minority, as "justice is infringed whenever equal liberty is denied without sufficient reason". But the freedom of an intolerant sect can be restricted "when the tolerant sincerely and with reason believe that their own security and that of the institutions of liberty are in danger."



Person P<sub>0</sub> is not making a decision about tolerance, but about private behavior. Person P<sub>1</sub> is making a first-order decision about tolerance, P<sub>2</sub> is making a second-order decision, and so on.

Whether or not the example involving sex is the kind of behavior that Popper and Rawls were concerned with, we can see how this framework relates to the ideas of "paradox" and "compromise." If we are second-order intolerant of first-order intolerance, this is not a compromise – at least not yet – or paradoxical. We have different choices at different levels, with different justifications. It's not that just *labeling* things first- or second-order takes away the problem. Rather, some relationships between choices at different levels can be normatively justified, and this may yield the familiar liberal combination.

For example, suppose we want to *respect individual autonomy* in situations where no one else is being adversely affected by someone else's choice, such as sex between consenting adults. As long as there's no hidden coercion, we want to respect freely-made choices. We are first-order tolerant. And in addition, we might want to *protect that realm* of free choice. So we are intolerant of those who are not tolerant at the first level. They would like to interfere, and we want to interfere in their interference. That is second-order intolerance which, if we just call it "intolerance," seems to clash with our first-order tolerance. But there is no clash; we are preventing interference in what we think ought to be a realm of personal choice.

Kerr and I think that questions about tolerance need to be handled this way, not by working with a single notion of tolerance that can be applied to anything, including tolerance itself.

I want to get to "free expression." So far, it does not look like that case fits easily into the framework, because speech and other kinds of expression are not usually "private." They are often intended to affect others, to influence what they think and do. Here is a more exact statement of the framework, making it more precise and starting to show how it applies to less simple cases. We start with a *base-level behavioral variable*. One example is who you have sex with. This is a variable, or set of options, not a single behavior. Being "private" is not essential to being base-level. What is essential is that *all* the values of the behavioral variable (all the options) are *non-interfering* in relation to the choices of others.

What is "interference"? The heart of interference is attempted coercion or constraint of some kind. Reasoning with someone – presenting, in a civil manner, an argument that they are making a mistake – is not interference in our sense. The clear cases of interference include physically harming, impeding, or threatening a person, or bringing the forces of law or institutional rules to bear on them. Interference has vague borders. There's a gradient between interference and the respectful offering of reasons, with various kinds of hostile speech and marshalling of economic incentives as borderline cases.

You are first-order tolerant in some domain if you do not interfere with *any* values of the base-level variable when chosen by others. You are intolerant to some extent if you interfere with some of these choices. (Ben and I referred to non-interference directed on a specific behavior as "permission" of that behavior.)

Second-order tolerance is a behavioral variable involving interference or non-interference with first-order tolerance or intolerance. And so on. The hierarchy is defined by the way that behaviors at higher levels are directed on or point at behaviors at lower levels – interference is the "pointer." Each higher-order variable includes a wholly non-interfering option and others that interfere with various behaviors at the next level down.

We assume that you might interfere with any of the behaviors at the level below, though not all of them. The number of options in the higher-level variables then gets very large. The familiar liberal combination, seen in the case of sex and elsewhere, is first-order tolerance and a particular kind of second-order intolerance. We are tolerant of all

base-level choices and intolerant of any interference in base-level choices. Our second-order policy is to only permit first-order tolerance.<sup>4</sup>

This combination reflects both the legalization of formerly problematic kinds of sex (first-order) *and* the legal forbidding of discrimination and harassment based on those choices (second-order). A combination like that is not *supported* by the framework, but it is certainly coherent. Some combinations are problematic in their own terms, though. It is possible to have combinations of policies at different levels that are (what we called) *behaviorally contradictory*. Here, at a higher level you interfere with some lower-level policy that you yourself choose at that lower level. A homophobic person intolerant of homophobes is an example.

We spent a lot of time in the earlier paper talking about combinations of policies at different levels – which ones embody internal tensions, how much freedom of movement there is. We made an assumption we called *teleological*, where the sole point of higher-level choices is to secure lower-level goods, such as base-level autonomy. With that assumption, defensible policies rapidly become very intolerant at higher levels. Protection of the lower-level things you regard as good demands that you interfere with nearly everything at higher levels. You might reject this "teleological" approach, and end up with different relations between levels. The framework does provide some support for the familiar liberal combination by dispelling, in part, the air of paradox that has bothered some people. But it does not dictate choices, beyond identifying some especially problematic combinations.

### **3. The Hierarchical Model and Expression**

One of the most famous expressions of a "familiar liberal" combination of policies is often attributed to Voltaire – though it is from a description of Voltaire's views by Evelyn Beatrice Hall, who wrote about Voltaire in the early 20th century: "I disapprove of what you say, but I will defend to the death your right to say it." That is a combination of first-

---

<sup>4</sup> The second-order policies seen in liberal societies today do not always interfere with *interference*. In Australia and elsewhere we have anti-vilification laws. Is vilification interference? Often it is probably not. (The law in my state, NSW: "It is unlawful for a person, by a public act, to incite hatred towards, serious contempt for, or severe ridicule of, a person or group of persons on the ground of the homosexuality of the person or members of the group.") If you refuse to make the cake for the gay wedding, that is also not interference in our sense. But I will focus on interference.

order tolerance and second-order intolerance. The Hall-Voltaire-figure states a disposition to interfere with those who would be first-order intolerant of their interlocutor's verbal behavior.

This case illustrates the necessity of a hierarchical treatment: you probably don't want to endorse Hall-Voltaire in an unqualified way, as perhaps the person you're offering to defend is presently saying something in an attempt to interfere with others. We can tell what Hall-Voltaire probably wants to achieve. Person A is saying controversial things of a political kind, person B wants to stop them, and Hall-Voltaire will defend the rights of A. But is this distinction between roles something that we're just supposed to pick up in the context? Is it to be based on what we think is *respectable*, even if we don't agree with it? This problem arises from the fact that much speech has nothing like a "private" behavioral role, and the application of the model is made more complicated by the fact that speech or expression is so *much*. Many of the interference behaviors high in the hierarchy are physically comprised of speech or writing.

Here is another case. In 2015, the well-known feminist Germaine Greer was invited to give a talk at Cardiff University in Wales. Because of her views on sex and gender, a petition called for cancellation of the talk. In the end, just a small protest took place on the day.<sup>5</sup> Discussing it afterwards, the journalist and film-maker Louis Theroux expressed concern about the left's new intolerance. Greer said: "But I'm on their side. If they want me to be shut up, then fine. Go for it. I don't think anyone's got a God-given right to speak... Shout away."<sup>6</sup>

This case introduces another theme. Communication is a two-sided phenomenon. There is speaking and hearing, production and reception. Each of these can be sites of

---

<sup>5</sup> See <https://www.theguardian.com/books/2015/nov/18/transgender-activists-protest-germaine-greer-lecture-cardiff-university>. "Germaine Greer defied a fierce campaign to stop her delivering a university lecture on the grounds that she has expressed transphobic views by going ahead with the event, which was conducted under high security.... More than 3,000 people signed a petition arguing that Greer should not be allowed to deliver her lecture on women in political and social life because her opinions on trans people were so offensive."

<https://www.buzzfeed.com/lanesainty/germaine-greer-lecture-goes-aheads-despite-protests>

"In response to the campaign, Cardiff University vice chancellor Colin Riordan said there was 'no case' for withdrawing a 'distinguished and exceptional academic speaker' such as Greer on the basis that she has voiced opinions that are offensive to others."

<sup>6</sup> <https://x.com/louitheroux/status/1764936846071550213>. See also:

<https://www.youtube.com/watch?v=7B8Q6D4a6TM>

interference. "Free expression" is often understood as bearing on the production side, but there are the two sides, and if you interfere in reception enough, you can make production pointless (perhaps even meaningless).<sup>7</sup>

I am not completely sure what combination of views Greer was expressing. She might have meant that she sees others as entitled to interfere with her expression, and doesn't advocate interference with their interference. Why endorse the protestors' interference with her but not someone else's interference with them? Maybe she would be OK with higher-level interference as well – she doesn't advocate it, but would permit it.

Alternatively, she might mean that the protestors are going to interfere with the *reception* of what she says, and she's not advocating interference with that. Perhaps: no one's got a God-given right to be *heard*. There's a difference here. Tim Scanlon, in a 1978 paper that I'll talk about in a moment (not the one he disowned), said: "freedom of expression adequately understood requires affirmative protection for expression, not just the absence of interference." I think those two things (absence of interference, protection) should be treated separately.

Let's embed all this within the hierarchical model. We need to first make sense of the idea of some ground-floor or base-level behaviors. As in the previous section, the idea of interference and non-interference defines the base level. It's not just that particular behaviors can be non-interfering, but there's a domain consisting of non-coercive, non-interfering speech.<sup>8</sup> This is speech that does not harness power in a way aimed at constraining what someone else says.

This base-level collection of speech acts might be mixed in with other acts that are different, but there's still this domain: speech that might be designed to be efficacious, but in a way that leaves receiver choice intact.

---

<sup>7</sup> This will come up in the discussion of pornography below. A podcast discussion: <https://www.abc.net.au/listen/programs/philosopherszone/pornography-and-free-speech/103402352>

<sup>8</sup> In our 2019 paper, "Though [private] behaviors are the clearest cases, what is essential to base-level behaviors is that these behaviors are not themselves cases of interference with others...." But that suggests that any non-interference *behavior* can be base-level. That is not what we mean. Base-level behavioral variables have non-interference as a feature of all their variants. There are some unresolved questions here. Not just any set of behavioral options forms a "behavioral variable" in the present sense.

Not everything that is base-level by this standard will be handled in the same way later in this talk. But let's set special cases aside for the moment, and stay with the main ideas.

First-order tolerance is a matter of whether you interfere with things said at the base level. Second-order tolerance is directed on *those* behaviors – on people's interference (or not) with base-level speech. And so on.<sup>9</sup>

We can re-express the Hall-Voltaire principle: "I disapprove of what you say in your base-level speech, but won't interfere with it, and I'll also defend, as a second-order policy, your right to say it when people try to interfere with your saying it."

We can compare the Hall-Voltaire attitude with another: "I disapprove of what you say in your base-level speech, but I will let you say it. I won't interfere with others who want to stop you from saying it – I will be second-order tolerant of their first-order intolerance." Greer might have been somewhere like this; her tolerance extended to others' attempts to interfere in base-level speech like hers.<sup>10</sup>

We have the multi-level picture in place. Some progress is made just by setting up the structure. As with the sex case, we might say: "If we respect autonomy at the base level, it makes sense to protect it by interfering with people who want to interfere with it." The appearance of paradox or compromise is reduced or gone.

But *justifying* such a policy is more difficult. Given the power of speech, no "privacy" justification can work. If someone says, "At the base level there is no coercion,

---

<sup>9</sup> There's a question of *what* is not being interfered with by speech at the base level. Is it just another person's contribution to base-level discourse, or that person's other behavioral choices as well? It has to be the broader criterion. It would not make a lot of sense to protect speech that didn't interfere with another's speech but did interfere, strongly, with other things they wanted to do, such as private behaviors.

<sup>10</sup> Are we talking about tolerance and protection of *one thing* (communication) which has these two sides? Or are the two sides best considered one at a time, as a pair of activities (production, reception) that might have different treatments? Is this a substantive question or just a matter of set-up? I'll later make a few comments about freedom of artistic expression as a case outside the main line of justification I'll consider for first-order tolerance. In the case of art, the expressive side seems perhaps more important to questions of freedom. In the political cases, a more integrated approach may be needed. The tendency in all the authors I've mentioned so far has been to tie the two sides closely together. This might affect whether some of the combinations I've just discussed are behaviorally contradictory. If you see yourself as a free speech advocate *and* a disruptor, then your "free speech advocacy" might have to be purely expressive, not two-sided, or there is at least a tension there.

so all this is OK," a reply is: coercion is not the only kind of harm. Why should we stand back and tolerate other harms resulting from base-level speech? For now, at least, I want to set aside a simple consequentialist argument – "this will help society overall" – in the style of J.S. Mill. Instead, I'll look at defenses that are less pragmatic and tied to circumstances.

### 3.1 Academic Freedom

Before we get to justification, I'll say something about "academic freedom."

Academic freedom is often seen as closely related to free speech. At my own university (Sydney), policy documents generally use the two phrases alongside each other, suggesting a package. In these settings, "academic freedom" apparently refers to something like a special form of free speech within an academic context. Perhaps special leeway should be given to the expression of unpopular ideas in university settings. But what is at stake in a lot of actual debate about academic freedom is something different. It is protection of the academic domain as a place where *two* sets of behaviors – expressing ideas, on one side, and policing, rewarding, and punishing, on the other – have a self-contained status. Outsiders should not tell academics what to work on, it is said, *and* should not tell academics how to reward, constrain, and interfere with each other. We should be left to make our own decisions about who to hire and who to not hire, who to engage with and who to exclude. Academic freedom in this sense is the idea of the academy as its *own little world*, with respect to a set of base-level behaviors *and* some higher-level policies directed on them. It might more accurately be called academic autonomy.

For clarity, we could distinguish two kinds of academic freedom, AF<sub>1</sub> and AF<sub>2</sub>. AF<sub>1</sub> is a special expressive right of some kind. AF<sub>2</sub> is the autonomy of the academy (within limits) as its own world of discourse-related behavior. When AF<sub>2</sub> holds, outsiders should not interfere in academic expression *or* in internal academic practices of interference and constraint. Plenty of people may want to advocate one kind of "academic freedom" without the other.<sup>11</sup> The question of who is "inside" the academy for purposes

---

<sup>11</sup> All this became vivid to me when I read a defense of academic freedom by someone I know who probably sees a lot of free speech advocacy as "right coded." He was not saying: in the academy, problematic ideas should be protected in a special way. He was saying: the academy should be free to handle its own policing without external interference. This is not a point about

of AF<sub>2</sub> is also becoming less straightforward as professional administrators acquire more power within many universities.<sup>12</sup>

#### 4. Justifying Policies

In 1972, Tim Scanlon published a paper that offered a non-consequentialist argument for broad tolerance in the area of expression.<sup>13</sup> Scanlon abandoned much of this paper within a decade, and I'll discuss the abandonment, too.

The idea is to initially focus not on speech itself, but on *making up one's mind*. Working out what to believe. That is something that often happens in the course of base-level communication. Scanlon said in the 1972 paper that we'd not give up that right to a government. The authorities might be able to tell us what to do, some of the time, but not what to believe. And further, although there might be various reasons the state has to restrict what people say, one thing that would *not* be a good reason for restrictions on expression is the fact that if people are allowed to say certain things, it might lead to bad beliefs being accepted – beliefs that are false or otherwise harmful. If the bad beliefs lead

---

the two-sidedness of speaker-hearer setups (as in the previous section); it is a point about the tolerance hierarchy.

In the UK in 2025, the University of Sussex was fined by the government for "failing to uphold freedom of speech" in relation to academic discussion of policies about trans people. See <https://www.bbc.com/news/articles/cn9vr4vjzqgo>. The government moved against Sussex's AF<sub>2</sub> because they thought the university was not sufficiently respecting AF<sub>1</sub>, especially in relation to the views of the philosopher Kathleen Stock.

<sup>12</sup> In a blog post commenting on the UK debates mentioned in the note above, Liam Kofi Bright says that we should accept that speech, under a free speech regime, will often be "rowdy." In response to an academic paper expressing racist views, for example, people might protest and call for the firing of the author. That's an ordinary exercise of free speech, Bright says, unless there's reason to believe that the university might respond by actually firing the person.

How does this relate to my framework? If calling for the person's firing was done by a person with any likely influence at all, then it's an attempt to interfere (first-order intolerance). What about the case where the speaker is known on all sides to be powerless? Perhaps then it's base-level. That depends on its being truly toothless, an invocation of interference rather than an attempt to interfere. If so, it would deserve its own protection against (genuine) interference. The same would apply to a powerless person who says to one of the protestors: "You should be expelled from the university." In some ways these are base-level contributions, and they are also meta-contributions (comments about the policies that should be in place).

The blog post: <https://sootyempiric.blogspot.com/2025/04/the-free-speech-world-is-rowdy.html>

<sup>13</sup> Thomas Scanlon, "A Theory of Freedom of Expression," *Philosophy and Public Affairs*, 1972

to bad actions, then the state can go after the actions. But the state should respect autonomy in forming beliefs, and that means that the downstream consequences of that autonomy, in action, should not be a basis for restriction on what people say to each other. Justifications for suppression or constraint are illegitimate when they "appeal to the fact that it would be a bad thing if the view communicated by certain acts of expression were to become generally believed."

So speech, thoughts, and non-verbal actions are all in the picture. Scanlon summarizes this with his "Millian principle":

There are certain harms which, although they would not occur but for certain acts of expression, nonetheless cannot be taken as part of a justification for legal restrictions on these acts. These harms are: (a) harms to certain individuals which consist in their coming to have false beliefs as a result of those acts of expression; (b) harmful consequences of acts performed as a result of those acts of expression, where the connection between the acts of expression and the subsequent harmful acts consists merely in the fact that the act of expression led the agents to believe (or increased their tendency to believe) these acts to be worth performing.

Scanlon's argument for this principle is based in the moral philosophy of Immanuel Kant.<sup>14</sup> (Some of the following sentences are near-quotes.) The appropriate powers of a state are limited to those that citizens could recognize while still regarding themselves as autonomous and rational agents. To regard themselves as autonomous in this sense, a person must see themselves as "sovereign" in deciding what to believe and weighing reasons for action. Such a person knows they might get things wrong, and knows they'll never have all the information they need, but they would not give to the state the power to restrict what people say, based on those possibilities for error.

---

<sup>14</sup> Might the idea be expressed more simply in terms of an individual right to make up your own mind? Scanlon does not want to handle it like this. "A right of this kind would certainly support a healthy doctrine of freedom of expression, but it is not required for one. The argument given above was much more limited. Its aim was to establish that the authority of governments to restrict the liberty of citizens in order to prevent certain harms does not include authority to prevent these harms by controlling people's sources of information to insure that they will maintain certain beliefs. It is a long step from this conclusion to a right which is violated whenever someone is deprived of information necessary for him to make an informed decision on some matter that concerns him."

Scanlon allows that the *manner* of expression of ideas can be a legitimate basis for state control. The state can say: you may express those ideas in some settings, but not here and now (blocking traffic, for example).

Scanlon's arguments can be connected to the hierarchical model I outlined earlier. My paper with Kerr supposed at some points (without argument) that preserving autonomy in private matters is part of a reasonable liberalism. Speech is generally not private. But autonomy might re-enter the story in relation to making up one's mind. Tolerance directed on base-level expression might be justified by its relation to individual belief (and desire) formation.

The connection between the Scanlon framework and ours extends to second-order policies, and the familiar liberal combination of first-order tolerance and second-order intolerance. If we value base-level discourse because of its role in reflection and belief-formation, we can reasonably decide to protect this discourse from people who would interfere with it. We can decide to protect a realm of non-coercive discussion. This would justify protecting lectures and talks on controversial topics, by interfering with attempts to make those events non-viable. It can include preventing large-scale interference with the "reception" side of the communicative context, such as attempts to drown out speakers in a way that makes communication impossible within a targeted event.<sup>15</sup> Those

---

<sup>15</sup> In our earlier paper, Ben Kerr and I looked at a way that the familiar modern liberal combination of policies might be rejected:

An argument the other way, as we see it, might proceed by looking more closely at the *justification* that can be given for first-order tolerance. Will this justification be something that picks up other levels beside level 1? If first-order tolerance is justified in terms of the importance of respecting freely made choices, for example, then this justification would seem to apply to any level. That would not mean that second-order intolerance should be abandoned, but it would mean that the familiar modern liberal combination (first-order tolerance and higher-order intolerance) really is a compromise, a falling away from an ideal of wholesale tolerance.

Could the same pressure be applied here? Not in a very direct way. The "importance of respecting freely made choices" above is not the same as: the importance of respecting freely made choices *about what to believe*. The cognitive-autonomy defense of free speech in terms of belief is narrow. It is not about *any* choice. You might value autonomy in a level-free way across the board, including choices to interfere. But there is not specific pressure *from* the valuing of autonomy seen in first-order tolerance. The protestors at controversial talks are not (only) trying to get people to believe something; they are trying to prevent the talk from taking place.

are some things we might think if we knitted the two discussions (Scanlon 1972, and the framework I developed with Kerr) together.

Scanlon abandoned the main ideas of his 1972 paper within a decade. I think his central idea was a good one, though, and an idea that can be developed without the heavy-duty Kantian ideas about agency that Scanlon used, also without Scanlon's restriction to coercion by a state. A version of the view can be developed in a philosophically "lighter" way, and also one that's broader in relation to contexts of evaluation and coercion (laws, institutional rules, social norms). The idea I take from Scanlon is the idea that the protection we should give to many forms of expression comes from the special status of making up one's mind, and this has consequences for the bases on which expression might be restricted. A problem that will then arise is that once we give up on the ambitious philosophical foundations seen in Scanlon's old paper, the link between expression and autonomous belief formation is subject to pressure of this form: "OK, there's something worthwhile here, but it goes into the mix along with a lot of other kinds of benefit and harm." Perhaps it does, and I'll come back to this toward the end. But first I'll try to develop the autonomy-based justification in more detail.

What is "autonomy" in belief formation? Scanlon, Susan Brison, and Joshua Cohen (each of whom I'll discuss below) have all argued that this is an unhelpful idea to appeal to. I treat autonomy as a personal negative freedom that can be valued in social contexts. Perhaps this is enough for the uses I'm putting it to?

Scanlon in his follow-up papers argued that his 1972 view has problems in how it handles some cases. It fails to give us a justification to prohibit deceptive advertising, and is too permissive about libel.<sup>16</sup> In response to the first of these, I would treat commercial contexts as special, and as having at least a special burden in relation to honesty. This exception is not simply *ad hoc*. The usual functions of base-level discourse take a special form in commercial settings. Belief-formation is involved (you are trying to decide if this is the right car to buy), but the network of incentives differs from other contexts of base-level discourse. In US law as I understand it, commercial speech (speech that proposes an economic transaction) now has some free-speech protection but less than non-commercial

---

<sup>16</sup> Scanlon 1978 ("Freedom of Expression and Categories of Expression"): "[T]he Millian Principle has what seem to be implausible consequences in some cases. For example, it is hard to see how laws against deceptive advertising or restrictions such as the ban on cigarette advertising on television could be squared with this principle." The cigarette question need not be one about deception.

speech, especially in relation to honesty. This is to some extent a compromise, but one with some principle behind it.

In the talks I gave at Sydney and Berkeley (2025), I said that I'd extend a "commercial contexts" exception beyond advertising and the like. There are expressive or symbolic *products* in commercial life, as well as expression in the form of advertising. Symbolic products are rightly subject to special constraints involving intellectual property and copyright protection. I thought (still wonder whether) we might treat a category of commercial speech that is broader than the usual one as subject to special constraints.

One motivation for this is that it can address some, though not all, of the problems around pornography. Pornography been defended with appeals to free expression. It has also been attacked through appeals to free expression. Catharine MacKinnon argued that in pornography, "The free speech of men silences the free speech of women."<sup>17</sup> Porn affects how women are perceived in such profound ways that when a woman does try to speak, she cannot get her ideas across, especially in relation to sex and her own preferences. In my terms, this is an argument that pornography is *not base-level*, as it is a form of interference in women's communicative behavior, on the reception side. That is a possible move within my framework. But I agree with critics who emphasize that the argument requires a failure of *comprehension* in listeners, not just lack of esteem, and that is a very high bar (see the West paper for discussion). I add, though, that most pornography is a commercial product (a symbolic commercial product), and hence need not enjoy the ordinary protection from first-order tolerance. We can impose special rules to reduce harm. Harmful non-commercial pornography would not be covered by this, and is a problem case for me.

In discussion at the Berkeley seminar, several people objected that my proposed extension of the category of "commercial" speech that gets special treatment would be much too broad. Academic books and ordinary argumentative non-fiction books are, after all, being sold. In reply: there has to be *some* special treatment in these cases, or copyright would collapse. And I am not envisaging that argumentative commercial books would be open to complete censorship if the authorities decided this would be convenient. As in the case of advertising and the US First Amendment, reduced

---

<sup>17</sup> MacKinnon, *Feminism Unmodified* – for discussion, see Caroline West's "Pornography," in the *Oxford Handbook of Free Speech*.

protection is not the same as an absence of protection. Rules around commercial speech (in my broader sense) might naturally mirror those around non-commercial speech to a large extent, but with special provisions. Still, I see the force of the objection and am not sure what approach to this issue is best.<sup>18</sup>

Other special contexts can be marked out. During a trial or an election, we might want fixed procedures governing the expression of information. Things that can't be said inside those contexts can usually be said outside them.

As well as special contexts, some one-off behaviors can intrude into normal base-level communication in a way that might warrant special treatment. I have in mind defamation, hate speech, threats, and the pair of problems known now as misinformation and disinformation.

Defamation (libel and slander) in the narrow sense – where it is directed at individuals and maliciously makes claims that are demonstrably inaccurate – is a reasonable case for a straight-out exception to protection. Due to its directedness on individual reputation (a special subject-matter), an exception for defamation does not change the picture very much.

Hate speech has been central to recent discussion of freedom of expression. I admire the US approach to this problem, where hate speech is not a legal category and not an exception to the main forms of protection for expression. In Australia and many other countries, it is an exception. In the paradigmatic, narrow sense of "hate speech," one can certainly see the justification. A problem is that the category tends to get extended. This is seen from various sides of politics – we saw a crude attempt at extending the category, as well as mistaking its legal status, from the Trump administration after the Charlie Kirk assassination. But it's not only the crude extensions that are problematic.<sup>19,20</sup>

---

<sup>18</sup> As with special contexts involving disruptive protests and the like, things that can't be said within a special context can usually be said outside it.

Especially at the moment, the distinction between commercial and non-commercial settings is vexed, though, because a lot of what people put online is *proto*-commercial. "Influencers" might not be selling anything now, but they hope to attain enough recognition that they can later monetize their project.

<sup>19</sup> <https://www.nbcnews.com/politics/justice-department/attorney-general-pam-bondi-doj-hate-speech-rcna231633>

<sup>20</sup> Scanlon, in his 1972 paper, battles with a case where a person wants to publish a recipe for home-made nerve gas. Scanlon would restrict this, and the wording of his official principle is set

I'll spend a bit more time on "misinformation" and "disinformation." The proliferation of wild ideas on the internet has prompted some people to defend special measures, including interference by governments with the expression of ideas, especially by leaning on media companies.<sup>21</sup> This is a path I don't support. It is antithetical to the nature of epistemic life. We're all groping forward, trying to work things out. We all – "we" including the authorities – will get many things wrong, through honest error, carelessness, paranoia, selectivity, and bias. Labeling some ideas in a way that sets them off for special treatment, as a discrete pathology, is very much against the spirit of this view.

"Honest error," I said. This brings in the distinction between *misinformation* and *disinformation*, where the latter is said to be malicious and deliberate, and the first simply erroneous. In commercial settings, detecting dishonesty is often feasible and the bodies that make the call can be fairly impartial. That's generally not true in areas of politics and policy (Scanlon makes this point). It is hard to detect insincerity, and the bodies that would make the call often have vested interests.<sup>22</sup>

That is a practical objection. Someone I discuss these problems with said to me: suppose that we could reliably detect insincerity and "bad faith," by some special method. Would I oppose governmental proscription (as well as ordinary base-level rejection) of disinformation in that case?

---

up so that providing technical means for harmful acts, as opposed to arguing that they are reasonable, will be excluded from protection. The fact that he had to engage in some footwork to deal with that case is notable. The general problem of military secrets is related.

<sup>21</sup> "Misinformation and disinformation can no longer be viewed simply as an academic nuisance, but rather they are a societal threat." *The Lancet*, editorial, January 2025. This is in an editorial: "Health in the age of disinformation." Harvard's Kennedy School now has a journal, *The Misinformation Review*: <https://shorensteincenter.org/programs/misinformation/the-misinformation-review/>

<sup>22</sup> Scanlon 1978: "laws against such ['false and deceptive'] advertising seem acceptable in a way that analogous laws against false or deceptive political or religious claims would not be, first because there are reasonably clear and objective criteria of truth in this area, and second, we regard the government as much less partisan in the competition between commercial firms than in the struggle between religious or political views." In addition, dishonesty as opposed to ordinary error is more detectable in commercial settings.

In the US, a defamation trial found that Alex Jones lied in his claims that one of the worst school shootings in the US (in 2012) was faked. This is perhaps a partial exception. Jones was seeking commercial gain, though. <https://www.bbc.com/news/world-63243981>

The thought-experiment is problematic, but perhaps I do have to say "yes." We have looked briefly at several forms of dishonesty: defamation, misleading advertising, perjury (as part of legal proceedings), and disinformation in the current sense. These all have the capacity to undermine ordinary discourse and deliberation, especially by undermining trust. They can undermine the features of base-level discourse that make it worth protecting. If we want to interfere with people who would *interfere* with base-level discourse, why would we not interfere with those who would *undermine* it?<sup>23</sup> A person might offer a libertarian reply – "Let people say what they'd like; others can believe it or not." In this paper I've not taken simple libertarian appeals of that kind to have much force, but more importantly, at least it might be clear in this situation that second-order *protection* of disinformative speech would make little sense. That point, perhaps, is a sort of pivot; it suggests that first-order tolerance also lacks its usual justification in this context.

I accept my interlocutor's argument in a hypothetical context where there is reliable detection of dishonesty. But in practical terms, this is a bad path.

Something I am more strongly opposed to is the idea that *misinformation* (where there is no claim of insincerity) should be handled in a pathologizing, coercive way. Our situation, again, is one where we're groping forward on difficult topics. Much, probably most, of what is actually contested in these areas is hard to know. We are all subject to biases and blindnesses of various kinds, and the claims people end up making have diverse degrees of justification. There's not even a useful thought experiment in this case, in which, hypothetically, we can directly detect falsehood – if we could do that, there'd be little need to discuss factual issues at all.

The fact that misinformation and disinformation tend to be discussed together is problematic, as the clear harm of insincerity accompanies the category of ordinary error or unorthodoxy, and makes the two appear as evil twins deserving of special treatment.

## 5. Pragmatic Balancing

I'll look at one other criticism of the approach I have been applying. The more recent discussions by Scanlon and also a paper by Susan Brison give various direct arguments against any autonomy-based approach to justifying freedom of expression. The best

---

<sup>23</sup> Trust can also be affected by too much constraint – see Nagel's "Personal Rights and Public Space."

argument is a simple one seen in both: why is it that in *this* domain, we are supposed to set aside attempts to balance considerations from each side? There are certainly goods to be gained from open discussion of controversial ideas, but why is this an area where compromise is a special problem? Legally, this might be the situation in the USA because of the First Amendment. But we're not all in the USA, and this is not a discussion of what's legal, anyway – it's more a discussion of what *should* be legal. And then, as Scanlon asks, why are the social costs acknowledged in the Millian principle (quoted on page 11) such that we have to pay *any amount* of these? "Why should we be willing to bear unlimited costs to allow expression to flourish provided that the costs are of these particular kinds?" Brison: "Why do those liberal theorists who defend free speech by invoking autonomy suppose that autonomy requires the absence of interference in the area of speech but not in other areas?"<sup>24</sup> If a person holds Scanlon's Kantian 1972 position, then the answer is: we just have to go along with this. But if we're not like that, why not a balancing?

One possible reply to this argument is *yes*. We could look at things this way. The considerations related to autonomy could go into the mix with other factors, when we think about freedom of expression. That is one way of handling the part of this paper that brings Scanlon's 1972 ideas into contact with the hierarchical framework. The Kantian framework that Scanlon used in the 1972 paper pushes against this, but with that framework set aside, the ideas about autonomy could be handled differently.

Is that the best way forward? At this point, some background meta-ethical assumptions become relevant to my own response to the situation (and others will proceed differently). What kind of activity can we reasonably see ourselves as engaged in here? We are engaged in an attempt at *reasoned valuation*. We're working out what we might reasonably value, and hence use as a basis for policy choices. Those policy choices will take various forms: we can discuss what we think ought to be illegal, what we think would be good social norms, and what rules institutions of various kinds (such as universities) should impose on people within them. That phrase "reasoned valuation" is my own label for a practice that I see as guided by empirical facts of various kinds, also guided by parity and consistency constraints, but not directed on a distinct set of moral facts. My view of ethical matters is a kind of "non-representational functionalism," to use

---

<sup>24</sup> Scanlon 1978, "Categories...", Brison, "The Autonomy Defense of Free Speech," 1998. "*Can we please stop talking about autonomy?*" is a subsection header in another Scanlon paper ("Why Not Base Free Speech on Autonomy or Democracy?" 2011).

a term of Blackburn's. (Blackburn's version of non-representational functionalism is different from mine, but the broad category is useful.<sup>25</sup>) Given this background, am I bound to respect a need for pragmatic balancing of different benefits and harms in situations like this? The short answer is no. One might end up committed to a pattern of valuation that is abstract and inflexible. One might end up resolving to resist pressure from the details of cases and stick to a broad principle (perhaps a principle involving autonomy). That is the short answer. But perhaps there is always some pressure from the specifics of cases, that would not always arise within other backgrounds, once you have a meta-ethical picture of the kind that I have.

Further, although autonomy in belief formation is a rather basic valuation for me, the link between base-level discourse and individual belief formation discussed in this paper is functional or instrumental. The usual features of such instrumental connections will apply – a generally reliable link might be broken in some contexts; a connection might be transformed in some settings in a way that brings other factors into play.<sup>26</sup> The fact that the link between individual-belief formation and discourse has this instrumental character makes it hard to set aside pleas for balancing.

In Scanlon's newer papers, and others, we wind up with a balancing of interests on many sides: the interests of speakers, hearers, and bystanders. People in each of those roles often have legitimate interests in letting discussion take place freely, and in some cases, legitimate interests in restriction. A right to free expression is seen as one possible upshot of a weighing of interests in the circumstances we're in. My instinct is to resist case-specific balancing as much as possible, but that's not the same as inflexibly asserting a general principle. This is one of those intrusions of philosophical and ethical temperament that I mentioned in my introduction. And as things are now, I don't much trust the people who would be in charge of doing the balancing when free expression might be compromised, especially with respect to the relation between immediate and longer-term risks.

Joshua Cohen has a paper that makes the balancing approach seem more viable than it might otherwise appear. For Cohen, the link to belief-formation that I've built the middle of this paper around is just one aspect of the role of discourse and expression. But

---

<sup>25</sup> See his *Ruling Passions* (where this term is used). Blackburn's view is too close to traditional expressivism. In *Living On Earth*, Chapter 7, I go into more detail in this area.

<sup>26</sup> Alex Voorhoeve emphasized this to me at my LSE talk.

if you take into account some standing "background facts," Cohen thinks that a balancing of factors should tend to preserve strong free speech protections in many circumstances.<sup>27</sup>

The question is then whether to embrace, in principle, the need for pragmatic balancing and see the main ideas in this paper as a contribution to that project, or whether to try to retain a more primal role for the link between expression and autonomy in belief formation. As noted earlier, this connection, primal or not, only relates to some of the issues around free expression (freedom of artistic expression is largely distinct, for example), and along the way I've allowed some exceptions to first-order tolerance as well. The side of the free speech debates that I most want to address is the side relating to political debate and advocacy, rather than commercial or artistic issues. With respect to those political questions, I do think a transformed version of the 1972 Scanlon picture is valuable.

## 6. Summary

Here is a summary of what I've tried to do. My first goal was to present the hierarchical framework for understanding tolerance, and embed questions about expression within it. Problems of free expression are made complex both by the hierarchy and by the two-sided nature of communicative contexts (expression and reception). As in other cases, the removal of an appearance of paradox or compromise within familiar liberal combinations of views is achieved by putting things into the hierarchical framework. One might then

---

<sup>27</sup> Joshua Cohen, "Freedom of Expression," *Philosophy & Public Affairs*, 1993. I won't list all Cohen's "Background Facts." He has a lot of them. They include facts like this:

"Under conditions of expressive liberty, people will arrive at conflicting, reasonable evaluative convictions."

"Most people - and those with power in particular - do not like to be criticized or disagreed with and are tempted to use the means at their disposal to avoid criticism or disagreement."

"If sanctions are attached to expression for being false, offensive, insulting, psychologically injurious, etc., then people will be reticent to express themselves (chilled), even if they think their expression is true, inoffensive, not insulting, etc."

"In a society with relatively poor and powerless groups, members of those groups are especially likely to do badly when the regulation of expression proceeds on the basis of vague standards whose implementation depends on the discretion of powerful actors."

Reduced protection for commercial speech arises organically from Cohen's treatment. Cohen is not enthusiastic about the autonomy-based approach: "autonomy has a capaciousness that strikes me as a vice in an account of expressive liberty."

offer an autonomy argument about base-level speech itself (at the base level, people should be allowed to say what they like). I didn't do that. Instead, I brought Scanlon's 1972 position, in a "lighter" form, into combination with the hierarchical framework. The autonomy argument is now made for belief formation, and base-level speech is seen as instrumental to belief formation. I then looked at problem cases, possible exceptions, and objections to the autonomy-based approach. An unresolved question is the extent to which autonomy is best seen as one factor in a treatment that seeks to balance many considerations, as opposed to having a more central and less qualified role.<sup>28</sup>

---

---

<sup>28</sup> I am grateful to audiences at the LSE (the Sir Karl Popper Memorial Lecture), the University of Sydney, and UC Berkeley. Particular thanks to Ben Kerr, Caroline West, Tim Scanlon, Luke Russell, Sam Shpall, Matti Sarkia, Rachael Weiss, Jonathan Birch, Shamik Dasgupta, Richard Bradley, Ryan Cox, Jay Wallace, Kate Lynch, Jane Sheldon, and Rebecca Mann for discussions and correspondence.