Tolerance: A Hierarchical Analysis[*]

PETER GODFREY-SMITH

History and Philosophy of Science, University of Sydney

and

BENJAMIN KERR

Biology, University of Washington

I. INTRODUCTION

Tolerance has become a central and contentious topic in recent years, especially in diverse societies. To what extent should an individual, a majority culture, or a particular group in a society abstain from interference in behaviors and life choices that differ markedly from their own? An especially difficult aspect of this question concerns tolerance of people and belief systems that do not exhibit much tolerance themselves.

Karl Popper wrote of this as a "paradox of tolerance." If a tolerant society tolerates intolerance, then tolerance itself will be broken down.[1] Popper introduced only a simple form of the problem; tolerance was treated as a unified matter, with tolerance of ordinary behaviors and of tolerance itself collected together. His question was largely practical: how do we prevent tolerance (in this unified sense) being undermined by toleration of the activities of those who are intolerant? We think the discussion needs to be framed in a more complex way, using a hierarchical structure. Tolerance (and intolerance) may be directed at a set of "private" or base-level behaviors—this is first-order tolerance. Then there is the question of tolerance or intolerance of *these* behaviors,

---

[1]See K. R. Popper, *The Open Society and Its Enemies, Volume 1: The Spell of Plato* (London: Routledge, 1945), p. 226.

the ones that express tolerance or intolerance of base-level behaviors—that is second-order tolerance. Then there is tolerance of *those* behaviors, and so on, through further levels of meta-tolerance and its absence. In this article we work through this hierarchy.

One aim is to handle situations like the following. Suppose you say to someone: "You are being intolerant. Stop your interference in another's choices." The person you have criticized replies that *you* are being intolerant—no less so. They say that you should tolerate this behavior of theirs, not intrude and try to prevent it. Borrowing Glen Newey's terminology, there appears to be a "symmetry" present in many situations in which appeals to toleration are made; each side to a dispute can describe the other as intolerant.[2] Does tolerance as a moral or political goal collapse in the light of this problem? What relations (logical, normative, empirical) are there between different levels in the tolerance hierarchy?

Our aim in describing this hierarchy is to identify a structural feature that will play a role in questions of tolerance of several kinds (political versus personal, for example), and that may be handled in different ways by proponents of different normative theories. Our goal is not to give a complete normative account of tolerance, but to describe this structural feature.[3]

Tolerance is seen by some as associated with additional tensions. One that is perhaps more fundamental, if real, has been discussed by Mendus, Williams, and others. Toleration seems to require that a person disapprove of something and yet protect or at

---

[2]See Glen Newey, "Is democratic toleration a rubber duck?", *Res Publica,* **7** (2001), 315–36; Glen Newey, "Political toleration: a reply to Jones," *British Journal of Political Science*, 41 (2010), 223–7. Newey doubted that toleration can be a "a coherent political value"; "Is democratic toleration a rubber duck?", p. 315.

[3]In some discussions, "toleration" is used to refer to a political requirement or ideal, while "tolerance" refers to a personal virtue. Here we do not follow usages of that kind. Our main points are intended to apply to both political and personal contexts, and we use the terms "toleration" and "tolerance" roughly interchangeably, though "toleration" tends to refer to an actualized pattern and "tolerance" to a disposition. (Because "toleration" does not have a widely used one-word English opposite—"intoleration" is rare—"intolerance" often has to do double duty in discussions that make more of the toleration/tolerance distinction.) When the distinction between political and personal contexts is relevant, we mark this distinction explicitly.

least accommodate its existence.[4] We won't address this question, though it will sometimes intersect with our topic. Arguments from the value of autonomy and freedom are often seen as sufficient in some form to resolve that problem.[5]

With autonomy on the table, it might also be argued that the right approach to our issue, too, is one that treats autonomy as the primary consideration, and handles tolerance in a derivative way. It is a familiar idea that the realm of an individual's autonomy should be limited by the effects of one's choices on the autonomy of others. This principle might be used to handle at least some questions about higher-order tolerance. Here, though, we address tolerance in its own right, and in a way that does not make initial assumptions about the value of autonomy.

The next section introduces the hierarchical framework. Then we consider cases and applications—particular combinations of lower- and higher-order tolerance and intolerance. Are only some combinations coherent? To what extent do commitments at one level constrain commitments at others? A formal section proves some propositions about the relations between levels of tolerance. We then compare our treatment to others, and examine what can be said within our framework about the relations between lower-order and higher-order policies that tend to be seen in modern liberal democracies.

II. A HIERARCHY OF TOLERANCE

We begin with some comments about tolerance in general—tolerance of anything at all, of any X, which could be a belief, practice, institution, or activity. Our treatment is simple. A person is tolerant with respect to X if they are *disposed not to interfere* with others in ways that would tend to curtail or constrain their choices concerning X.[6] This

---

[4]See Susan Mendus, *Toleration and the Limits of Liberalism* (Atlantic Highlands: Humanities Press, 1989); Bernard Williams, "Toleration: an impossible virtue?", David Heyd (ed.), *Toleration: An Elusive Virtue* (Princeton: Princeton University Press, 1996), pp. 18–27; and (on the other side) Jeremy Waldron, "Toleration: is there a paradox?", New York University Public Law and Legal Theory Working Papers, Paper 376.

[5]Waldron regards the "paradox" as illusory, not requiring a substantive reply of this sort; "Toleration".

[6]An interference criterion of this kind is common. It has been endorsed, for example, by Peter

conception of tolerance is thinner than usual. What if the person who refrains from interference is just lazy? They disapprove, but don't bother to act. Or perhaps they don't care at all. Most treatments of tolerance want to rule out cases like this, as insufficient for genuine tolerance.[7] One might say instead that a person is tolerant with respect to *X* if they are disposed not to interfere in others' choices concerning *X because* they are motivated by the view that non-interference in individual choice is appropriate whenever possible.

However, we say that if someone is lazy or indifferent, they are de facto tolerant. A person can also be tolerant through adherence to a principle. For us, any of these is sufficient. A society can be tolerant in a genuine but weak sense, because its people are not strongly motivated to interfere in certain practices. The society is tolerant in another sense if its people do care about the issue, but opt to protect minority behaviors because of beliefs about the importance of individual choice. Tolerance, in our framework, is just the absence of interference.[8]

Above we said that tolerance is being disposed not to interfere with others in ways that would tend to curtail or constrain those others' choices. Which behaviors count as *interference*, in the relevant sense? Reasoning with someone—presenting, in a civil

---

Jones, "Making sense of political toleration," *British Journal of Political Science*, 37 (2007), 383–402. Newey objected that some intolerance does not involve interference at all: "For example, the ritualistic abuse of each other by Glasgow Rangers and Celtic football fans seems the epitome of intolerance. But the supporters do not thereby curtail each others' *agency*"; Newey, "Political toleration," p. 224. We don't agree that cases like this must be handled by a theory of toleration (and neither does Jones; see "Political toleration: a reply to Newey," *British Journal of Political Science*, 41 (2010), 445–7.

[7]See, for example, Andrew J. Cohen, "What toleration is," *Ethics*, 115 (2004), 68–95, at p. 71: "Toleration is not indifference. If I see someone playing baseball (in which I have no interest) and I walk past without interfering, we would not say I tolerate the behavior … Some negative response is necessary for our lack of interference to count as toleration. Put another way, we must care." See also Tim Scanlon, "The difficulty of tolerance," reprinted in his *The Difficulty of Tolerance: Essays in Political Philosophy* (Cambridge and New York: Cambridge University Press, 2003), p. 187: "Tolerance thus involves an attitude that is intermediate between wholehearted acceptance and unrestrained opposition."

[8]One might ask: what if people do not interfere because they *endorse* what is being done? As discussed below, in our framework tolerance is directed on variables, not specific behaviors. An agent will endorse one or more variants and may or may not interfere with the choices of others.

manner, an argument that they are making a mistake—is not usually seen as incompatible with tolerance, though occasionally it is. The clear cases of interference include physically harming, impeding, or threatening a person, or bringing the forces of law to bear on them. There is a gradient between those behaviors and the respectful offering of reasons, with various kinds of hostile speech, marshalling of economic incentives (and so on) as borderline cases. In this article we seek to keep this side of the discussion simple, as our focus is the hierarchical structure. Much of our discussion assumes a distinction between cases of behaviors that do, and do not, interfere with a person's autonomy. We accept that there is considerable complexity here, and add brief discussions of less clear cases later. Our aim is to clarify questions about both legal–institutional forms of tolerance and more informal, non-institutional practices, and we use examples from both sides of this divide below. Various behaviors might count as interference in an everyday sense, but not a sense that invites legal sanction. (Threats of social exclusion are an example.)

We now describe the hierarchy.[9] At the bottom level we have behaviors that were called in Section I "base-level." In the clearest cases, these behaviors are *private* in a familiar, intuitive sense. Other people, or others outside a small immediate context, are not significantly affected by them. Examples include sexual behaviors in a context where

---

[9]We are not aware of earlier published discussions of the relations between tolerance and "meta-tolerance" in our sense. The closest we know of is Andrew Sable, "'Virtuous to himself': pluralistic democracy and the toleration of tolerations," *Nomos*, 48 (2008), 220–40. Sable says: "Toleration in a pluralistic society is therefore *toleration of tolerations* or meta-toleration. We put up not just with other people's false and potentially dangerous world-views, but with other people's false and potentially dangerous ways of thinking about why they should put up with ours"; p. 228. The following informal remark on an internet discussion board converges closely with our ideas, however, and makes use of the same terminology and a central example: "What say you about a sort of first-order tolerance, second-order intolerance? (Which is nothing more than having real confidence in one's convictions of the importance of (a certain kind of) tolerance.) E.g.: 'Gay marriage will not be disallowed *cock the shotty* [shotgun] *or else!*' hehe"; <http://www.unfogged.com/archives/comments_9142.html#903764>, listed as posted Aug. 21 2008. As Mark Colyvan (in discussion) pointed out to us, the treatment of a "paradox" of tolerance by establishing a hierarchy of this kind is also reminiscent of approaches to the liar paradox that establish a hierarchy of languages. For a review, see J. C. Beal, Michael Glanzberg, and David Ripley, "Liar paradox," Edward N. Zalta (ed.), *Stanford Encyclopedia of Philosophy* (fall 2017), <https://plato.stanford.edu/archives/fall2017/entries/liar-paradox/>.

"consenting adults" are the only agents present, and no question of coercion or deception arises.

Though behaviors like these are the clearest cases, what is essential to base-level behaviors is that these behaviors are not themselves cases of interference with others— not forms of coercion or incentive-giving that are aimed at constraining others' choices. Later we will consider behaviors such as public speech that are not "private" in an intuitive sense, but need not be cases of interference either.

Beginning with private behaviors, we now take a step back from any particular behavior, to consider a behavioral *variable*, *B*. This might be the practice of having, or not having, sex of various kinds, in a context in which all the behaviors in question are private. The behavioral variable *B* has *variants*—engaging in sex with an opposite-sex spouse or partner, engaging in sex with a same-sex or non-binary spouse or partner, engaging in casual sex (of some particular kind), and so on. Those are all variants of the variable *B*. What we are describing is the variable/value distinction as seen in mathematics. A *variable* (time of day) takes on different *values* (for example, noon). We say *variants* rather than *values* of a variable, as the term "values" is confusing in this context, given its other meanings.[10]

Now consider behaviors that are responses to what people might choose with respect to *B*. One family of behaviors comprises forms of *interference* in others' choices. These include prevention by physical or legal means, threats of such action, and milder responses involving other sanctions intended to constrain choice. As noted above, what counts as interference can be a difficult question and may vary across contexts. But however interference is marked out, *tolerance* in general is non-interference in a behavioral variable; that is, non-interference across all variants of that variable. A person might also decline to interfere with choices only among some variants of a variable (might be tolerant of another's choices of sexual partner with respect to gender, but not with respect to ethnicity). Here we use "tolerance" to apply to non-interference across all

---

[10]A referee pointed out that this structure might also be set up using sets and their elements rather than variables and their values. Our ideas could be expressed equivalently in this way, and below we'll occasionally speak of the set of values associated with a variable.

variants of a variable; other cases could be handled by constructing a new behavioral variable with a smaller set of variants.

We use the term "tolerate" to refer only to responses to variables. In a colloquial sense, a specific behavior (a variant) can also be "tolerated." We use the word "permit" to refer to non-interference in a particular behavior. A person who is tolerant of private sexual choices is someone who permits lots of different behavioral choices in sexual matters.

More specifically, though, if B is something like private sexual partner choice, then what we have on the table so far is *first-order tolerance* and its alternative, *first-order intolerance*.

First-order tolerance with respect to B is non-interference in all variants of B.

First-order intolerance with respect to B is interference in some variants of B.[11]

Many Western democratic societies have moved closer to first-order tolerance in this sense. They have become tolerant about a range of behavioral variables that involve private choices, including many related to sex.

Tolerance we said is behavioral; people interfere or do not. This can be seen as another behavioral variable, which we will call variable $T(B)$. In the simple case, $T(B)$ might have just two variants: *tolerate all variants of B* and *interfere in some variants of B*. Generally, though, differences between various forms of intolerance will be important, and it will make sense to track which base-level behavior(s) a person tends to interfere with. If every kind of intolerance is distinguished, there will be more variants of $T(B)$. If there are r private behaviors to consider, and it's assumed that a person can't be disposed to interfere with all of them, then the number of variants of $T(B)$, the number of choices possible with respect to first-order tolerance, will be $2^r - 1$.

So far we have been discussing $T(B)$ as a choice someone might be confronted with: do I interfere with some private behaviors or not? But $T(B)$, as a behavioral variable, is also something that people can be tolerant or intolerant about in turn. Whether

---

[11]More exactly, it is interference with a strict non-empty subset of the set of variants.

or not you are tolerant of private sexual behavior, for example, how do you feel about others who differ from you on this point? Would you look for sanctions that might make *them* behave differently?

A person shows s*econd-order tolerance with respect to B* if they permit (do not interfere with) all the different variants with respect to first-order tolerance of *B*. Such a person does not interfere with first-order tolerance, and does not interfere with first-order intolerance of any kind. They do not interfere in choices between variants of $T(B)$. A person is s*econd-order intolerant with respect to B* if they are not permissive with respect to some variants of $T(B)$. That person seeks to interfere either with first-order tolerance or with first-order intolerance (of all kinds, or of some particular kind). That is, they seek to interfere in response to choices of some variants of $T(B)$.

The choice between second-order tolerance and various forms of second-order intolerance gives us another behavioral variable, $T(T(B))$. Anti-discrimination laws that protect sexual minorities are second-order intolerant. They promote interference in all forms of first-order intolerance; they thus enforce first-order tolerance.

A formula often attributed to Voltaire (actually formulated as a description of Voltaire's views by Evelyn Beatrice Hall) is: "I disapprove of what you say, but I will defend to the death your right to say it."[12] That is also a combination of first-order tolerance and second-order intolerance. Despite disapproval of what someone says, the Voltaire-figure believes that that person should be able to say what they think. The Voltaire-figure states a disposition to interfere with those who would be first-order intolerant of their interlocutor's verbal behavior.

Speech is often not a "private" behavior of the sort we initially assumed above, and this case also brings us to the heart of questions about the nature of interference. The Voltaire-figure will interfere with those who would interfere with their interlocutor. But if those who would interfere would do so by speaking (perhaps with threats or insults), the Voltaire-figure has to be willing to interfere in some kinds of speech. For this combination to make sense, there must be a distinction between speech of the sort

---

[12]See Evelyn Beatrice Hall, *The Friends of Voltaire* (London: Smith, Elder and Co., 1906) [originally published under the pseudonym "S. G. Tallentyre"].

Voltaire wants to defend—perhaps speech that does not interfere with the autonomy of others, even though it might be aimed at altering others' behaviors—and interfering forms of speech which Voltaire is prepared to fight to prevent.

Compare the attitude of Hall's Voltaire with another: "I disapprove of what you say, but I will let you say it. I won't interfere with others who want to stop you from saying it." Here we have first- and second-order tolerance. The speaker is permissive of those who display a certain kind of first-order intolerance regarding B. (If the speaker is truly second-order tolerant, they should also not interfere with other first-order choices.)

We said that the choice between second-order tolerance and second-order intolerance gives us another behavioral variable, $T(T(B))$. So far, $T(T(B))$ is a variable that describes a choice someone is making. But those choices might also be targets of interference, or non-interference. There can be tolerant and intolerant attitudes to second-order tolerance. How do I feel about people who are second-order tolerant, or second-order intolerant, about sexual choices? Do I look for sanctions of some sort, to prevent people from exhibiting one or the other? A person is *third-order tolerant with respect to B* if they permit all different variants with respect to second-order tolerance of B. A person is third-order *intolerant with respect to B* if they interfere in some choices with respect to second-order tolerance of B—if they seek to interfere in some variants of $T(T(B))$.

Suppose someone abuses and threatens a gay person. Another person watching would never engage in such behaviors, but does not interfere with the abuser because of an attitude of second-order tolerance. They tolerate first-order intolerance. That presents a new choice about toleration. More fully: person P_0 is gay; person P_1 abuses and threatens them from an attitude of first-order intolerance; person P_2 does not endorse the abuse but does nothing, being second-order tolerant of P_1's first-order intolerance. Now person P_3 says to P_2: "If you just stand by and let that happen, I'll make sure you regret it." That person takes P_2's non-interference as a target, and seeks to interfere in it. P_3 is third-order intolerant of P_2's second-order toleration.

More elaborate stories can be told to illustrate still higher-order toleration and its absence. The structure is a bit like a Gricean hierarchy of mental states of the kind seen in

some kinds of communication (I recognize that you recognize that I intend you to think …) and is more closely related to hierarchies of enforcement that have been discussed in work on the evolution of cooperation in social groups: I will enforce this behavioral norm, and will act against those who do not enforce the norm, and will also act against those who do not act against those who do not enforce … . That is a kind of hierarchy of intolerance, where the first-order behavior is one taken to have evolutionarily relevant consequences.[13] It can be difficult to intuitively process cases that extend beyond third-order tolerance, but we think there is no limit to how far the hierarchy might go—no limit to the kinds of tolerance and intolerance that might be the targets of higher-order behaviors of this kind.

III. RELATIONS BETWEEN LEVELS

In this section we first work through the entire range of first-order and second-order choices, and begin to discuss the relations between lower and higher levels of tolerance in a more general way. Below, the term "policy" is used to refer to the acceptance of a particular variant at some level as a guide to behavior at that level. Sometimes we will also need a term for the acceptance of a combination of such choices at several levels; we use the phrase "general policy" for those multi-level choices.

One combination of first- and second-order policies is tolerance at both levels. You do not interfere in private behaviors, and also do not interfere with choices about first-order tolerance. This seems prima facie coherent, reflecting a general policy of non-interference.

A second "unmixed" case is intolerance at both levels. You interfere in some private behaviors, and also interfere with those who would permit them. Unlike the first case above, which could be described simply as tolerance at all levels, multi-level intolerance takes more specification, as there are several kinds of intolerance and it is possible for interference policies at each level to be misaligned. Suppose b_1 and b_2 are the

---

[13]See Samuel Bowles and Herbert Gintis, *A Cooperative Species: Human Reciprocity and Its Evolution* (Princeton: Princeton University Press, 2011).

available private behaviors. A person might be first-order intolerant as they tend to interfere with b_1, and also second-order intolerant in a way directed on those who permit b_1—that would make sense. It would not make sense, however, for that person to be first-order intolerant by interfering with b_1, while being second-order intolerant by interfering with those who interfere with b_1. In effect, that person would be intolerant of themself; they would be intolerant of those with their own first-order policies (a homophobic person intolerant of homophobes).

We will refer to combinations like this as *behaviorally contradictory*. This is not the only kind of tension that can arise between first- and second-order commitments, just the clearest. Others will be discussed below.

Coherent combinations of first- and second-order intolerance might be seen in very conservative societies in which those who turn a blind eye to private behaviors deemed unacceptable might themselves be shunned or otherwise sanctioned.

A third combination of first and second orders was discussed above, the familiar liberal combination where first-order tolerance is enforced by anti-discrimination legislation, and the like. This is first-order tolerance combined with a particular kind of second-order intolerance. In such a combination, a range of private behaviors is permitted and individuals who interfere with any of these private behaviors are subject to some form of interference themselves.

Once a person is first-order tolerant, there is again a form of second-order intolerance that creates a behaviorally contradictory combination: first-order tolerance and second-order intolerance directed on the first-order tolerant.

The fourth family of combinations is the least familiar and intuitive: first-order intolerance and second-order tolerance. Though these are less familiar, they generate no behaviorally contradictory combinations, because the second-order policy is tolerance; it's then not possible to be intolerant in a way directed on oneself. Perhaps this combination might be recognizable as the upshot of some kinds of cultural relativism: a person is intolerant in their interference with a particular private behavior, but accepts that others might permit it. While not permitting the behavior itself, they permit non-interference in this behavior by others, as reflective of their perspective or norms. The

combination is awkward—why respect the choices of those who are tolerant of those who choose a particular behavior, but not respect the choices of those who engage in the behavior?[14]

We now discuss constraints between levels in a general way. To what extent are the choices made at each level independent? Is it possible in principle for someone to have a very patchy combination of choices—for example, tolerance at level n, intolerance of some kind at level $n+1$, and tolerance again at $n+2$?

A first set of constraints was encountered above. There we saw some cases where a person was intolerant at the second level in a way that implied interference with their own first-order policies. If these combinations are always irrational, then we have an inter-level constraint that involves a kind of internal consistency.

A second kind of constraint has a more teleological character. Having decided to value a particular arrangement at one level, you may decide to protect that arrangement at higher levels, by interfering with those who would interfere with it. This is very much a further decision; nothing like an inconsistency results if someone decides not to follow such a path. One might value a low-level arrangement, but decide not to interfere with those who would interfere with it. The value of protection of a desirable lower-level arrangement, for you, might not be sufficient to overcome what you see as a problematic denial of autonomy at the higher level. Any project of protecting at higher levels the policies or arrangements valued at lower ones relies on bridging principles of some kind.

Here we will follow one particular line of argument that does yield strong constraints between levels. We do not suggest that the assumptions used in this model are the only reasonable ones, or the most reasonable ones. In this section we summarize some results; the next section gives a formal treatment with proofs.

The initial assumption made is that the value of higher-order policies lies in its protection and facilitation of desirable lower-level policies, so constraint moves upwards in the hierarchy. The proof given is general, in that no assumptions are made about where

---

[14]If this person is truly second-order tolerant, then they also permit the first-order variant that interferes with everything they permitted and permits everything they interfered with. They allow (in a second-order sense) their first-order polar opposite.

and how this chain of justificatory relations bottoms out. One possibility is that non-interference in base-level choices is deemed valuable, and higher-level policies are then determined by that first-order evaluative grounding. Another possibility is that base-level autonomy is rejected because of the desire to foster a particular base-level behavior and exclude others. Then an evaluation at the base level constrains the higher levels—the judgment that homosexuality is an abomination that must be opposed in every possible way may determine higher-level policies.

In order to present these results, it is necessary to track every combination of permission and interference at each level. At each level there is a single option of permission of everything at the next level down, and a range of different kinds of intolerance. The range of policies available at each level grows quickly. For example, suppose there are only two base-level behaviors available, and suppose also that at each level, interference with *everything* at the next level down is not an option (you can permit everything, but you can't permit nothing). Then at the third level there are 127 available policies, one that is permissive (tolerant) across the board, and 126 that are intolerant in different ways (disposed to interfere with different second-order policies). If interference with everything at the next level down *is* included as an option, then the third level contains over 65,000 policies directed on those two base-level behaviors. In our proofs below we include this fuller list of options, without loss of generality, for reasons of accounting simplicity.[15]

We define an *evaluation* at level *n* as a determination for each policy at that level of whether it is *desirable* or *undesirable*. Desirability has two origins. At some level in the hierarchy, an *initial evaluation* is made (perhaps on the basis of a moral principle, perhaps not in a way based on principle). At higher levels than that one, the evaluation is made according to the relations between policies at that higher level and the evaluations below. Here, the only desirability we consider for higher-level policies is a kind of

---

[15]If interference with everything at the next level down from a given level is not an option, the number of third-order policies is $2^{\left[2^{(2^2-1)}-1\right]}-1 = 127$. If every possible policy is included, the number is $2^{2^{2^2}} = 65536$.

perfection. We will say that a *conditionally perfect* policy at level n is one that does not interfere with any desirable policy at level n–1, and does not permit any undesirable policy at level n–1.

Once an evaluation at some level is given, the value of policies at all higher levels is determined. Think of the initial evaluation as a list of ones and zeros attached to each of the variants at that level. A variant at that level that receives a 1 is desirable and a variant that receives a 0 is undesirable. This assignment then determines the value of all policies at the next higher level, and only one policy at that next higher level can be conditionally perfect. The crux of the matter is the fact that at the next level up, there is *only one way* of permitting everything deemed desirable at the lower level and interfering with everything undesirable at that level, so there is a single conditionally perfect policy at the higher level. All the other policies at that level get at least one action directed on an immediate lower-order variant wrong.

For example, if tolerance is the only desirable policy at level 1, then there is just one level-2 policy that does not interfere with first-order tolerance and does interfere with all the intolerant level-1 policies. That is a conditionally perfect policy at level 2. Then there will be a single level 3 policy that has the analogous relation to this conditionally perfect level 2 policy—permitting that conditionally perfect level-2 policy, and hence all the lower-level policies it permits, and nothing else. We have rapidly reached policies that are not only intolerant, but are *maximally* intolerant—they permit only one policy at the level below (here we set aside the possibility of permitting nothing at the lower level). There is a special situation if the initial value specification (perhaps at level 1, perhaps elsewhere) assigns desirability to more than one policy. Then there is a single conditionally perfect policy at the next level, but not a maximally intolerant one. However, at the next level up after *that*, maximum intolerance is established and remains in place at higher levels.

What are the messages of this part of our discussion? First, any conclusions must be qualified by the fact that we chose to look only at "conditionally perfect" higher-level policies—those that permit everything desirable and nothing undesirable at lower levels. This is both a strict and entirely teleological handling of the higher levels (which can

include levels 1 and 2). The *only* thing that justifies a policy at a higher level is how it handles the protection of policies deemed valuable at lower levels; there is no independent justification for tolerance (or anything else) at higher levels. It is possible to weaken this treatment, within a teleological treatment of higher levels, so that a desirable higher-level policy might (for example) permit desirable lower-level policies, but not necessarily interfere with undesirable ones. (This we plan to follow up in further work.) The rather stark case modeled here does make clear some features of the situation, however. It shows that tolerance at level 1, together with further assumptions about the value of protection of that tolerance, generates intolerance at higher levels in a strong way; we rapidly reach not only intolerant but maximally intolerant policies, as at each level only one policy protects everything desirable (and interferes with everything undesirable) at the level down.

At least within assumptions of the kind made here, first-order and higher-order tolerance are seen to be very different things. Discussing a situation simply in terms of "tolerance" will not be adequate—or it will not be adequate unless there is a commitment to tolerance at each level being comparably valuable, and valuable for the same reasons, regardless of how higher-level policies affect lower-level ones. If, on the other hand, tolerance is not only endorsed at a low level, but also deemed worthy of protection, this asymmetry between levels leads to desirable higher-level policies rapidly becoming intolerant.

IV. A FORMAL TREATMENT

The previous section began to describe, in an informal way, some general relationships between levels of tolerance. This section, which can be skipped if desired, gives a more formal treatment of the main ideas in Section III.

First we introduce a symbolism for representing behavioral variants at different levels. The ith variant at level n is denoted v_i^n. Suppose there are two private (level-0)

behaviors: v_1^0 and v_2^0. There will be four level-1 variants: v_1^1, which permits both v_1^0 and

v_2^0; v_2^1, which interferes with v_1^0 and permits v_2^0; v_3^1, which permits v_1^0 and interferes with

v_2^0; and v_4^1, which interferes with both v_1^0 and v_2^0. There will be $2^4=16$ level-2 variants,

$2^{16}=65,536$ level-3 variants, and so on.

We define an *evaluation* at level n as a determination for each variant at that

level of whether it is *desirable* or *undesirable*, or more formally:

> *Definition*: We say that level n is *evaluated* when every variant at that level is
>
> deemed either desirable or undesirable. Evaluation at level n implies that
>
> there exists a single *desirability vector*: $d_n = \langle p_{v_1^n}, p_{v_2^n}, p_{v_3^n}, \dots p_{v_N^n} \rangle$, where
>
> $p_{v_i^n} = 1$ if the ith variant at level n (v_i^n) is deemed desirable and $p_{v_i^n} = 0$ if
>
> the ith variant at level n is deemed undesirable. (We let N denote the number
>
> of variants at level n.)

At some level in the hierarchy, an *evaluation* determines which variants are

desirable. Here we outline one way in which higher-level desirability might be

specified, which depends on a higher-level feature that we label *conditional*

perfection, defined as follows:

> *Definition*: A *conditionally perfect* variant at level m interferes with all
>
> undesirable variants at level m-1 and does not interfere with any desirable
>
> variant at level m-1. (A *conditionally imperfect* variant at level m is any
>
> variant at that level that is not conditionally perfect.)

We now define a connection between conditional perfection and desirability:

Definition: A *perfectionist general policy* is one that, after an initial evaluation of desirability at level n, designates variants at level m that are conditionally perfect as desirable; and designates variants at level m that are conditionally imperfect as undesirable, for all $m > n$.

We can now formulate the following propositions:

Proposition 1: Given an evaluation of desirable variants at level n, a perfectionist general policy will always have a single desirable intolerant variant at level m, for all $m > n$.

Further, if we define a *maximally intolerant* variant at level m to be behavior that interferes with all variants except one at level m-1, we can express a second proposition:

Proposition 2: Given an evaluation of desirable variants at level n, a perfectionist general policy will always have a single desirable variant that is maximally intolerant at level m, for all $m > n + 1$.

The proofs are as follows. Consider an initial evaluation associated with a particular vector d_n defining desirability for each of the N variants at level n. Of the 2^N variants at level $n + 1$, only one interferes with all undesirable lower-level variants and does not interfere with all desirable lower-level variants. Specifically, there is a vector:

$$P_{n+1} = \langle p_{v_1^{n+1}}, p_{v_2^{n+1}}, p_{v_3^{n+1}}, \ldots p_{v_{(2^N)}^{n+1}} \rangle$$

where $p_{v_x^{n+1}} = 1$ for v_x^{n+1} that is defined by the xth variant at level $n + 1$ such that

interference occurs for every $p_{v_i^n} = 0$ and permission occurs for every $p_{v_j^n} = 1$ (here the

index i represents all undesirable variants and index j represents all desirable variants at

level n) and $p_{v_y^{n+1}} = 0$ for all variants at level $n + 1$ where $y \neq x$. Thus, there is a single

perfect variant at level $n + 1$ (and all other variants at this level are imperfect, as they

permit undesirable variants at level n, interfere with desirable variants at level n, or

both). Given that we have a perfectionist general policy, this single perfect variant at

level $n + 1$ is desirable and all other imperfect variants at this level are undesirable.

Thus, the vector P_{n+1} can be relabeled as a new desirability vector d_{n+1}, which

plays a similar role at level $n + 1$ to that which the vector d_n played at level n. We use

exactly the same argument to show that there exists a vector P_{n+2} that defines a single

variant at level $n + 2$ that is perfect (which interferes with the vast majority of

undesirable variants at level $n + 1$ and permits the single desirable variant at level $n + 1$,

v_x^{n+1}). All other variants at level $n + 2$ are imperfect (because they either permit one or

more undesirable variants at level $n + 1$ or interfere with the desirable variant at level

$n + 1$, or both).

The argument continues to show that whenever there is a single perfect m^{th}-order variant, there must be a single perfect $(m + 1)^{\text{th}}$-order variant. Thus, for an initial evaluation at level n (that is, given by d_n), there is a series of single desirable intolerant variants at each level above level n. Above level $n + 1$, we are guaranteed to have this desirable variant permit only a single (desirable) lower-level variant and interfere with all other (undesirable) lower-level variants, which proves the propositions.
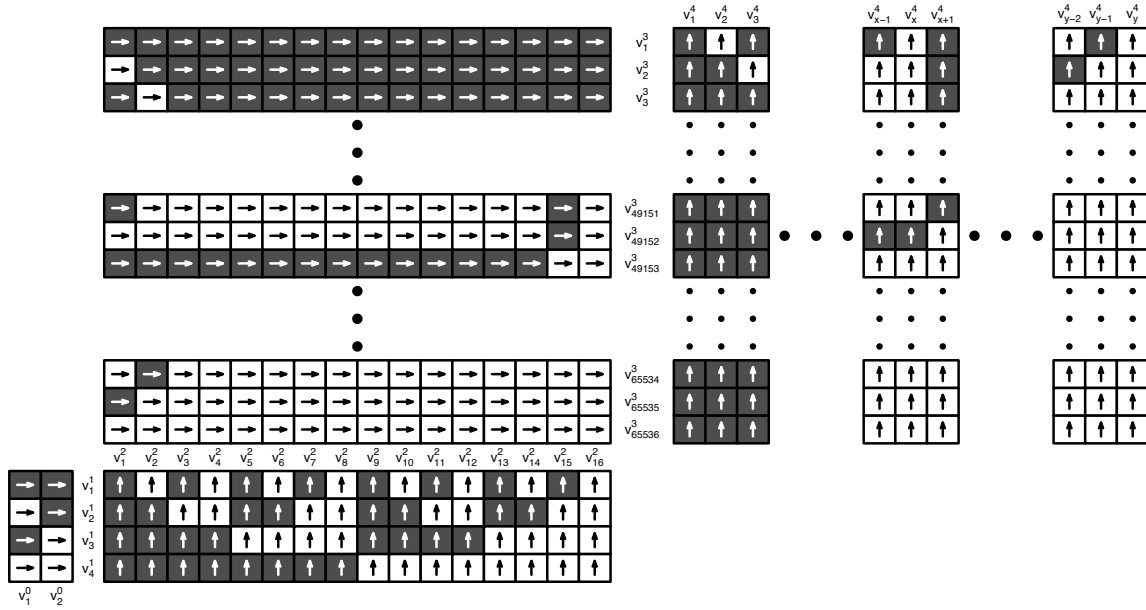

Figure 1. A visual representation of the hierarchy of variants

Supplementing our treatment in this section, Figure 1 gives a graphic representation of the hierarchy of behavioral variants. We start at the lower left of the figure with two variants of the base-level behavior (v_1^0 and v_2^0). There are $2^2 = 4$ first-order variants, which

involve either interfering with or permitting each of the base-level variants. These four variants are visualized by the rows of the 4×2 matrix of squares above the base-level variant labels, where dark gray corresponds to the base-level variant below being permitted and white corresponds to it being interfered with. The rightward pointing arrows indicate the location of the first-order variant label to which the relevant square belongs. For instance, v_3^1 is the intolerant first-order variant that permits v_1^0 and interferes with v_2^0. The columns of the 4×16 matrix of squares to the right of the first matrix illustrates the $2^{2^2} = 16$ second-order variants. Here each square indicates whether the first-order variant that heads its row is permitted (dark gray) or interfered with (white). The upward-pointing arrows again give the location of the second-order variant label to which each square belongs. For instance, v_7^2 is the intolerant second-order variant that permits first-order variants v_1^1 and v_4^1, but interferes with first-order variants v_2^1 and v_3^1. The rows of the 65536×16 matrix above the second matrix illustrates the $2^{2^{2^2}} = 65536$ third-order variants. The large dots refer to third-order variants (rows in this matrix) that are not shown. Again, the arrows point to the relevant third-order variant label. Finally, the 2^{65536} fourth-order variants are illustrated in the columns of the 65536×2^{65536} matrix to the right of the third matrix. The large dots refer to fourth-order variants (columns in this matrix) that are not shown, while the small dots refer to squares that are not shown in a particular fourth-order variant.

We can use this figure to illustrate our propositions. Imagine that we declare the tolerant first-order variant v_1^1 as desirable and the intolerant first-order variants v_2^1, v_3^1, and

v_4^1 as undesirable. Then, the only second-order variant that is perfect (and thus desirable) is the intolerant variant v_{15}^2, which permits v_1^1 and interferes with v_2^1, v_3^1, and v_4^1. All other second-order variants are imperfect (and thus undesirable), as they either interfere with v_1^1 or permit one or more of v_2^1, v_3^1, or v_4^1; or both. The only third-order variant that is perfect is v_{49152}^3, which permits the desirable second-order variant v_{15}^2, but interferes with all the other (undesirable) second-order variants. Finally, the single perfect fourth-order variant is v_x^4 (where $x = 2^{65536}$-$2^{49152} + 1$), which permits v_{49152}^3 and interferes with the other 65535 (undesirable) third-order variants. All other y-1 (where $y = 2^{65536}$) fourth-order variants are imperfect. Thus, we see only a single perfect (desirable) variant at each level above the first level.

We now offer some further comments about these relationships. Similar perfectionist policies at higher levels can arise for different reasons and for initial evaluations at different levels. Consider two base-level (level 0) variants of a private behavior. In society A, suppose that each variant (v_1^0 and v_2^0) is deemed desirable, leading to the vector $d_0 = \langle 1,1 \rangle$. Society B, in contrast, refrains from evaluating desirability at level 0, but does evaluate the first-order variant that permits both base-level variants as desirable, and views the three other first-order variants as undesirable. Thus, for society B, $d_1 = \langle 1,0,0,0 \rangle$ (assuming the tolerant variant is listed first, followed by the intolerant variants and ending with the unconditionally interfering variant).

How would each society evaluate the higher levels? For society A, the first-level variant, v_1^1, that permitted both v_1^0 and v_2^0 would be conditionally perfect, whereas all other variants at this level (v_2^1, v_3^1, and v_4^1) would be conditionally imperfect. If society A practices a general perfectionist policy, then its desirability vector at the first level is $d_1 = \langle 1,0,0,0 \rangle$. Society B has the same desirability vector at the first level, because this tolerant policy reflects its initial evaluation. If society B is also perfectionist, there is complete agreement across these two societies for all levels above the base level. Societies A and B may have very different reasons for their initial evaluations. Individuals in society A may want to promote diversity in base-level behavior, while those in society B may want to promote first-order tolerance, whether or not they have a preference for one base-level variant. Nonetheless, these two societies can end up endorsing similar variants at higher levels.

V. COMMENTS ON OTHER TREATMENTS

We believe that earlier discussions of tolerance would have benefited from explicit consideration of levels of tolerance. We illustrate this claim by looking first at the discussion of tolerance in John Rawls's classic *A Theory of Justice*, and then at some more recent discussions.[16]

Rawls imagines a society governed by the principle that each person is to have an equal right to the most extensive set of basic liberties that is compatible with the same liberties being accorded to others, and also the principle that social and economic inequalities are justifiable only insofar as they are associated with positions that are both open to everyone and beneficial in absolute terms to the least well off. Suppose intolerant

---

[16]John Rawls, *A Theory of Justice* (Cambridge, MA: Belknap Press, 1971).

groups arise within such a society. Should they be tolerated?

Rawls argues first that an "intolerant sect" cannot itself complain, if it is denied equal liberty, as a person's "right to complain is limited to violations of principles he acknowledges himself". But that does not mean that the majority has the right to "suppress" the intolerant minority; this may still be unjust, as "justice is infringed whenever equal liberty is denied without sufficient reason". However, in some circumstances the majority can intervene on the basis of self-protection. 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."[17]

Rawls's own discussion is "flat" in the sense of this article—no distinction is made between first-order and higher-order toleration. We think that our hierarchy provides a Rawlsian with useful resources. Suppose, in the most obvious disambiguation of Rawls's scenario, that the "intolerant sect" is both first-order and second-order intolerant. For Rawls, the question is whether to restrict its liberty, and he answers *no* as long as the society's institutions are secure," but *yes* if they are endangered. Within a hierarchical framework, one can instead advocate first-order tolerance towards the intolerant sect—non-interference with their base-level behaviors—but second-order intolerance. We may interfere with their attempts to interfere with others. This is not a decision made and revised according to whether institutions are in danger; it applies all the time. We do not permit their first-order intolerance when all is stable, or interfere with their base-level behaviors when there is trouble.

We think this combination is justifiable within a Rawlsian framework. Rational agents in Rawls's "original position" would want to live in a society that is first-order tolerant (that is much of the content of Rawls's first principle), but also shows a certain kind of second-order intolerance.[18] Otherwise those agents would have their society give up its ability to protect free choice in private behaviors; a person who found themselves in an unpopular minority, with respect to their private behaviors, would have no recourse

---

[17]Ibid., pp. 190–2.

[18]The "original position" is a fictional context of choice, in which one chooses the basic rules for a society without knowing what sort of position one will have in that society.

if others chose to interfere.

We turn to some more recent discussions. Glen Newey, as noted earlier, argued that a problematic "symmetry" affects concrete debates about toleration; both sides to a dispute over appropriate behavior can describe the other as intolerant. Newey regards this problem as undermining toleration as a political goal. Peter Jones, Rainer Forst, and others agree that there is a prima facie problem here, but attempt to solve it without giving up on the ideal of toleration.[19] As before, our aim is to make a structural point, not to give a complete resolution of these problems, and many of the insights in this literature can be re-expressed in our terms. But an explicitly hierarchical recasting also makes a difference.

Forst argues that in many cases where two disputing parties each claim intolerance in the other, two meanings of "intolerance" are on the table. On one side there is "the intolerance of those who lie beyond the limits of toleration because they deny toleration as a norm in the first place," and on the other, "the intolerance of those who do not want to tolerate a denial of that norm."[20] This brings Forst close to a hierarchical treatment, but that is not the direction he takes. For Forst, there is such a thing as *toleration* (not specific to a level in a hierarchy) that has a normative rationale, one based in the possibility of reciprocal justification in a context of discourse. This normative rationale determines the proper limits of toleration. Intolerance of those who reject the norm is not equivalent to intolerance as it appears within the norm: "To call both of those viewpoints equally "intolerant" presupposes that there is no non-arbitrary, impartial way of drawing the limits of toleration in the light of higher-order normative considerations."[21]

---

[19]See Jones, "Making sense of political toleration"; Rainer Forst, "The limits of toleration," *Constellations*, 11 (2004), 312–25; Rainer Forst, *Toleration in Conflict: Past and Present*, trans. Ciaran Cronin (Cambridge: Cambridge University Press, 2013). Here is Forst's expression: "There simply is no toleration, one could say, for any concrete understanding of that concept leads to intolerance towards those who are arbitrarily called "intolerant"—which means that toleration is always merely a more or less effectively veiled form of intolerance"; "The limits of toleration," p. 313.

[20]Forst, "The limits of toleration," p. 314.

[21]Ibid.

Our response is that it may be true that a general normative theory can mark out the proper limits of toleration in this way, but a normative account is not the only way to make sense of the distinction Forst is discussing. The distinction can be marked in structural terms. A distinction between first-order and higher-order toleration can be recognized by people who disagree on many normative matters. The justifiability of tolerance at each level can then be addressed in its own right.

VI. FURTHER ISSUES AND CONCLUDING REMARKS

In this final section we discuss two additional points and offer closing remarks. First, the simplest way to present the tolerance hierarchy is to set out a base-level behavioral variable that is private, in an intuitive sense, as in the example of private sexual choices. But we also discussed freedom of speech, and saw speech as an example of a behavior that might be protected with a combination of first-order tolerance and higher-order interference. Public speech is not a "private" behavior in an ordinary sense, but it can be a "base-level" behavior in the sense we introduced above. It can be a behavior that is not itself a form of coercion or interference with another's choices. Most likely there is no sharp line between speech that interferes, in our sense, and speech that merely offers reasons to someone, intended to help change their mind. Many sentences might constitute interference (in a sense relevant either to legal sanction or to informal norms) when uttered in some contexts but not others, and might have very different effects on different members of a single audience. Much of the complexity of contemporary problems of tolerance resides here, in the diverse roles that speech can have in recommending, motivating, prescribing, and compelling action. "Offensive" speech that does not interfere with the choices of others would count as a base-level behavior in our terms. However, as noted below, considerations of tolerance need not be the only ones relevant to assessing and perhaps restricting speech in a liberal society.

Second, we will make some more general comments about what might be said for or against a combination of first-order tolerance together with higher-order intolerance aimed at protecting that first-order policy, via anti-discrimination laws and the like. This

25

we said earlier is a "familiar liberal" combination of attitudes, but that does not mean it is uncontested.

The first thing to say is that this is a coherent combination. It is not behaviorally contradictory in the way some combinations are. Part of the point of this article is that if "tolerance" is described in a non-hierarchical way, there seems to be more tension in this combination than there really is. The components are different—first-order tolerance, plus a specific form of second-order intolerance—and compatible with each other. A more difficult question is whether this should be seen as a compromise of some sort—a falling away from a reasonable ideal. Here we will offer some points on each side of this question.

The previous section emphasized that first-order and higher-order tolerance are different sorts of things, which may have different justifications. Higher-order intolerance can function to secure first-order tolerance, which in turn can be defended in terms of autonomy. Similarly, political equality might be valuable in itself, while being defended by means of institutions that give some people (police, judges) powers that others do not have. Once first-order tolerance has been recognized as a good, it's natural to expect that it requires protection with other measures. We sketched a Rawlsian argument for this combination, and other work in the liberal tradition can be seen as making a case for this combination.[22] When this combination is in place, and a response to first-order intolerance becomes necessary, that response need not be violent, and should not target behaviors by the intolerant that would otherwise receive protection under first-order tolerance. A response of this kind is not an expression of a "tit-for-tat" policy. The aim instead will be to protect, using the most restrained means available, a regime of first-order tolerance.

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

---

[22]See, for example, John Stuart Mill, *On Liberty* (London: John Parker and Son, 1859).

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. Both Rawls and Popper wrote in a way that suggests this view, even though their discussions did not distinguish levels of tolerance. Both saw intolerance of the intolerant as sometimes necessary to preserve a society and its general norms of tolerance. They did seem to see intolerance of the intolerant as regrettable, and (at least in the Rawls case) not to be pursued if the society is in no danger.

If the moral and/or political justification for non-interference is one that is level-independent in this way, then that is certainly a possible attitude. There may be general reasons why autonomy regarding private behavior should be protected more strongly than the autonomy of behavior that specifically affects other individuals. If so, when it comes to weighing the value of the autonomy of a first-order intolerant individual against the value of the autonomy of an individual expressing a controversial base-level choice, the latter may have more weight. However, if the roles and justifications for first-order and higher-order tolerance are seen as quite different, then there is no need to see a combination of tolerance at level 1 and intolerance at higher levels as a compromise of any kind. It does not represent a trade-off between competing values, but may represent a whole-hearted endorsement and defense of base-level autonomy—of autonomy restricted only by the effects of one's choices on others.

We accept, however, that that the distinction between interference and non-interfering behaviors is graded and complex, and also accept the possibility of subtle trade-offs and weightings between the endorsement of intolerant higher-level policies that protect lower-level goods, and policies that value autonomy at many levels. Lastly, we do not claim that *all* questions about the appropriate regulation of behavior in a liberal society need be handled within a framework based on levels of tolerance. Other arguments based more directly on the avoidance of harm may play a role. These arguments might, in principle, justify some restrictions on speech and related behaviors. We leave open how many questions the framework sketched here does cover.