

# *Tolerance: A Hierarchical Analysis*

Peter Godfrey-Smith\*

History and Philosophy of Science  
University of Sydney

Ben Kerr

Department of Biology  
University of Washington

[Under review. September 2017.]

We present a framework that helps clarify moral and political goals of tolerance. A central aim is progress on questions of "toleration of the intolerant," discussed by Popper in the 1940s and made urgent by a range of recent events. We argue that tolerance is best discussed in a hierarchical framework, distinguishing first-order tolerance (and intolerance) that is directed at a set of base-level behaviors, second-order tolerance (and intolerance) directed at first-order policies, and so on through successive levels of meta-tolerance. We argue that some problems in earlier discussions of tolerance stem from their use of a "flat," non-hierarchical conception of tolerance. Rawls' treatment in *A Theory of Justice* (1971) is discussed as an example. Some propositions describing relations between different levels of tolerance are proved in an Appendix.

## *1. 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" – one of several now

---

\* Corresponding author. School of History and Philosophy of Science, University of Sydney, NSW 2006, Australia. Email: pgodfreysmith@gmail.com.

discussed. If a tolerant society tolerates intolerance, then tolerance itself will be broken down.<sup>1</sup> 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, 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 paper 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. Our focus here is not so much on the practical question, as in Popper – how do we prevent tolerance being broken down? Instead, we ask: how coherent is the *ideal* of tolerance, in the face of this problem? Does tolerance as a moral or political goal collapse, once expressed in a systematic way? What relations (logical, normative,...) are there between different levels in the tolerance hierarchy?

Tolerance is seen by some as associated with other tensions. One that is perhaps more fundamental, if real, has been discussed by Mendus, Williams, and others.

---

<sup>1</sup> See Popper, K.R., *The Open Society and its Enemies. Volume 1: The Spell of Plato* (London: Routledge, 1945): "Less well known is the *paradox of tolerance*: Unlimited tolerance must lead to the disappearance of tolerance. If we extend unlimited tolerance even to those who are intolerant, if we are not prepared to defend a tolerant society against the onslaught of the intolerant, then the tolerant will be destroyed, and tolerance with them. In this formulation, I do not imply, for instance, that we should always suppress the utterance of intolerant philosophies; as long as we can counter them by rational argument and keep them in check by public opinion, suppression would certainly be most unwise. But we should claim the *right* even to suppress them, for it may easily turn out that they are not prepared to meet us on the level of rational argument, but begin by denouncing all argument.... We should therefore claim, in the name of tolerance, the right not to tolerate the intolerant" (p. 226).

Tolerance seems to require that a person disapprove of something and yet protect or at least accommodate its existence.<sup>2</sup> (This is often called the paradox of toleration, not tolerance.) 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.<sup>3</sup>

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 meta-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? Are some but not others justifiable, either in a ground-up way or through appeal to general moral and political principles? To what extent do commitments at one level constrain commitments at others? What can be said for or against the combinations that tend to characterize modern liberal democracies? The main text of the paper is informal, though stating the ideas requires introducing some terminology. An Appendix gives a more formal treatment of some of the ideas.

## *2. 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*. This

---

<sup>2</sup> See Susan Mendus *Toleration and the Limits of Liberalism* (Atlantic Highlands: Humanities Press, 1989); Bernard Williams, "Toleration: An Impossible Virtue?," in 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.

<sup>3</sup> Waldron ("Toleration") regards the "paradox" as illusory, not requiring a substantive reply of this sort.

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.<sup>4</sup> 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.<sup>5</sup>

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 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 (etc.) as borderline cases. In this paper we start with a distinction between clear cases of

---

<sup>4</sup> See, for example, Andrew J. Cohen, "What Toleration Is." *Ethics* 115 (2004): 68–95: "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" (p. 71). See also Tim Scanlon, "The Difficulty of Tolerance," reprinted in *The Difficulty of Tolerance: Essays in Political Philosophy*. (Cambridge and New York: Cambridge University Press, 2003): "Tolerance thus involves an attitude that is intermediate between wholehearted acceptance and unrestrained opposition" (p. 187).

<sup>5</sup> 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.

behaviors that do, and do not, interfere with a person's autonomy, but we accept that there is considerable complexity here, and we discuss less clear cases later. Our discussion here is broad in that we aim to help clarify questions about both legal-institutional forms of tolerance, and more informal, non-institutional practices – our framework is supposed to apply to questions of tolerance both as legal requirement and as personal ideal.<sup>6</sup> We use examples from both sides of this distinction 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.<sup>7</sup> At the bottom level we have behaviors that were called in the Introduction "base-level." In the clearest cases, and the easiest ones to use when presenting our framework, these behaviors are *private* in a familiar, intuitive sense. Other people, or others outside a small immediate context, are not significantly affected by these behaviors. Examples include sexual behaviors in a context where "consenting adults" are the only agents present, and no question of coercion or deception arises.

---

<sup>6</sup> Sometimes now the political requirement is referred to as *toleration* and the personal virtue as *tolerance* (see Waldron, "Toleration"), but we use those terms roughly interchangeably, though sometimes with toleration as an actualized pattern and tolerance as a disposition. When the distinction between legal and private issues is relevant, we mark this distinction more explicitly.

<sup>7</sup> We are not aware of earlier published discussions of the relations between tolerance and "meta-tolerance" in our sense. The closest we know of in published work is Andrew Sable, "'Virtuous to Himself': Pluralistic Democracy and the Toleration of Toleration," *Nomos* 48 (2008): 220-240. 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](http://www.unfogged.com/archives/comments_9142.html#903764) (accessed 9.13.17, listed as posted on 08-21-08).

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 JC Beal, Michael Glanzberg, and David Ripley, "Liar Paradox", *The Stanford Encyclopedia of Philosophy* (Fall 2017 Edition), Edward N. Zalta (ed.), URL = <<https://plato.stanford.edu/archives/fall2017/entries/liar-paradox/>>.

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 a 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 spouse or partner, engaging in casual sex (of some particular kind), etc. Those are all variants of the behavioral variable *B*. What we are describing is the variable/value distinction as seen in mathematics. A *variable* (time of day) takes on different *values* (eg., noon). We say *variants* rather than *values* of a variable, as the term "values" is confusing in this context, given its other meanings.

Now consider behaviors that are responses to what people might choose with respect to *B*. One family of behaviors comprise 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 tolerant person is not disposed to interfere. A person might also decline to interfere with choices only among some variants of a variable (might be tolerant of anothers' choices of sexual partner with respect to gender, but not do so with respect to ethnicity). Here we use "tolerance" to apply to non-interference across all 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  $B$ .<sup>8</sup>

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 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 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 *second-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 *second-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

---

<sup>8</sup> More exactly, it is interference with a strict non-empty subset of the set of variants.

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 as follows: "I disapprove of what you say, but I will defend to the death your right to say it."<sup>9</sup> 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 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

---

<sup>9</sup> See Evelyn Beatrice Hall, *The Friends of Voltaire* (London: Smith, Elder and Co., 1906) originally published under the pseudonym "S. G. Tallentyre").



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 a gay person. Another person watching would not abuse the gay person, but does not interfere with the abuse because of an attitude of second-order tolerance. They tolerate the first-order intolerance of the abuser. That presents a new choice about toleration. More fully: person  $P_0$  is gay; person  $P_1$  abuses 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.<sup>10</sup> 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. Long chains of authority, for example, (chair, dean, provost...) can generate decisions involving high degrees of meta-tolerance.

---

<sup>10</sup> See Samuel Bowles and Herbert Gintis, *A Cooperative Species: Human Reciprocity and Its Evolution* (Princeton: Princeton University Press, 2011).

### 3. Relations Between Levels

In this section we first work through the entire range of first-order and second-order policies, and then discuss the relations between lower and higher levels of tolerance in a more general way.

A first 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 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 themselves; they would be intolerant of those with their own first-order policies (a homophobic person intolerant of homophobes).

We will refer to policies like this as "quasi-contradictory" – not quite contradictory, but not far off. 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 policy, a range of private behaviors is permitted and individuals that 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 quasi-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 quasi-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 recognisable 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?<sup>11</sup>

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

---

<sup>11</sup> 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.

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'll now summarize a proof discussed in detail in the Appendix. We do not suggest that the assumptions used in this proof are the only reasonable ones, or the most reasonable ones.

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 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 the proof, 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 56,000 policies directed on those two base level behaviors. In the Appendix we include this fuller list of options, without loss of generality, for reasons of accounting simplicity.)

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 level and the evaluations below. Here, the only desirability we consider for higher-level policies is a kind of perfection. We define a *conditionally perfect* policy at level  $n$  as one that does not interfere with any desirable policy at any lower level, and does not permit any undesirable policy at any lower level.

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.

Here is a summary of results proved in the Appendix:

- Given any specification of desirable  $n^{\text{th}}$  order policies, there will always be a single conditionally perfect (and hence desirable)  $m^{\text{th}}$  order intolerant policy for  $m > n$ .

- Given any specification of desirable  $n^{\text{th}}$  order policies, there will always be a single conditionally perfect (and hence desirable)  $m^{\text{th}}$  order maximally intolerant policy for  $m > n+1$ . (Multiple lower-level policies may be permitted by a conditionally perfect policy at level  $n+1$ .)

What are the messages of this formal 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 in terms of "tolerance" in general 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.

#### 4. Rawls on Tolerance

We believe that earlier discussions of tolerance would have been benefitted from explicit consideration of levels of tolerance. Here we look at just one example, the discussion of tolerance in John Rawls' *A Theory of Justice*.<sup>12</sup>

Rawls describes two principles that should govern societies. The first or *equal treatment* principle holds that each person is to have an equal right to the most extensive set of basic liberties that is compatible with the same liberties be accorded to all others. The second or *difference principle* holds that social and economic inequalities are justifiable insofar as they are beneficial in absolute terms to the least well off, and associated with positions equally open to everyone. Rawls discusses whether a society following his principles should be tolerant of intolerant groups within it. He uses the example of religious commitment, but intends it to apply generally with minor changes.

Several questions should be distinguished. First, there is the question whether an intolerant sect has any title to complain if it is not tolerated; second, under what conditions tolerant sects have a right not to tolerate those which are intolerant; and last, when they have the right not to tolerate them, for what ends it should be exercised. (p. 190)

Rawls' discussion is "flat" in the sense relevant here – there is no distinction between tolerance and various levels of meta-tolerance. Let's work through his treatment with a hierarchical re-framing. Assume that there is diversity in behaviors of private religious observance, behaviors that do not themselves interfere with others' religious choices, and diversity also with respect to behaviors that constitute interference. Then we can work through his questions. Rawls says his first question is "whether an intolerant sect has any title to complain if it is not tolerated." Both of these intolerances can differ with respect to order. Assume first that the "intolerant sect," assumed to be a minority, is first-order intolerant. Then there are two sub-questions within Rawls' first question. There is (1a) a question of whether the first-order intolerant sect can complain if others do not tolerate (permit) its own private behaviors, and a separate question (1b) of whether the intolerant sect can complain if its intolerance is not second-order tolerated.

Rawls says that the answer to his first question is no:

---

<sup>12</sup> John Rawls, *A Theory of Justice* (Cambridge MA: Belnap Press, 1971).

[I]t seems that an intolerant sect has no title to complain when it is denied an equal liberty. At least this follows if it is assumed that one has no title to object to the conduct of others that is in accordance with principles one would use in similar circumstances to justify one's actions toward them. A person's right to complain is limited to violations of principles he acknowledges himself. (p. 190)

Suppose we accept, with Rawls, that "a person's right to complain is limited to violations of principles he acknowledges himself." Then regarding question (1a), the first-order intolerant sect cannot complain of first-order intolerance directed at its own private behaviors. Regarding (1b), we don't know yet. We have not yet said whether the first-order intolerant sect is second-order intolerant. Suppose it is – suppose it is intolerant in a way directed on first-order tolerance (and on first-order intolerance directed at its own behaviors). Then the sect cannot complain to the majority. If, on the other hand, the sect is second-order tolerant, then it might (for all that has been said so far) complain if others do not tolerate its first-order intolerance.

Rawls' second question is: "under what conditions tolerant sects have a right not to tolerate those which are intolerant." Again there are several possibilities regarding order. One disambiguation asks (2a): under what conditions do generally first-order tolerant groups have a right not to first-order tolerate the private behaviors of sects which are first-order intolerant? Another is (2b): under what conditions do first-order tolerant sects have a right not to be second-order tolerant of a sect's first-order intolerance? These are closely related to Rawls' "first" questions, but approached now from the side of a first-order tolerant majority, not a first-order intolerant minority.

For example, suppose first-order intolerant sect *U* protests at the place of worship of the generally first-order tolerant sect *V*, preventing *V*'s members from entering their place of worship. If *V*'s members responded by protesting at *U*'s place of worship in the same sort of way, that would make *V* first-order intolerant in this case – a tit-for-tat with respect to first-order intolerance. But it would also be possible for *V*'s members to counter-protest *U*'s protest at *V*'s place of worship, preventing the prevention of their own members entering, while not interfering with *U*'s other activities. This would retain



*V*'s first-order tolerant character by restricting its response to second-order intolerance.<sup>13</sup>

Rawls does not distinguish these two responses to a first-order intolerant sect. What he is instead concerned to argue is that the answer to his "second" question does not follow from what he said about the right to complain:

Let us suppose, then, that an intolerant sect has no title to complain of intolerance. We still cannot say that tolerant sects have the right to suppress them. For one thing, others may have a right to complain. They may have this right not as a right to complain on behalf of the intolerant, but simply as a right to object whenever a principle of justice is violated. For justice is infringed whenever equal liberty is denied without sufficient reason. (p. 191)

As we interpret this, Rawls is probably talking about a first-order intolerant response (the tit-for-tat case above), as well as a less controversial second-order response. For Rawls, the majority cannot "suppress" an intolerant minority simply because the minority cannot complain; it still might be unjust to do this. However, in some circumstances the tolerant part of the society can intervene. (This is Rawls' answer to the third in the list of questions he raises.) Intervention can be justified by self-protection:

Justice does not require that men must stand idly by while others destroy the basis of their existence.

The conclusion, then, is that while an intolerant sect does not itself have title to complain of intolerance, its freedom should be restricted only when the tolerant sincerely and with reason believe that their own security and that of the institutions of liberty are in danger.

[J]ust citizens should strive to preserve the constitution with all its equal liberties as long as liberty itself and their own freedom are not in danger. They can properly force the intolerant to respect the liberty of others, since a person can be required to respect the rights established by principles that he would acknowledge in the original position. But when the constitution itself is secure, there is no reason to deny freedom to the intolerant. (p. 192)

A first problem here is that the "since" clause in the third quotes passage would seem to justify more than Rawls wants to allow in this context. The claim made is that because

---

<sup>13</sup> This would be consistent with a conditionally perfect (in our sense) higher-level move in defence of first-order tolerance, if *V*'s members counter-protested at all sects that *U*'s members protested at.

people in the original position would opt for a system that includes respect for the liberty of others, people can be forced to respect the liberty of others. (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.) This principle does not only apply when the constitution is in danger, however. It seems to generally justify the enforcement of tolerance of (at least) private behaviors; it seems to justify enforcement of first-order tolerance.

Might people in the original position instead choose a scheme that is explicitly second-order tolerant? If so, they would have their society give up the 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 if others chose to interfere. Rawlsian agents seem more likely to choose a scheme that is first-order tolerant but second-order intolerant. Indeed, this seems to be a reasonable interpretation of what Rawlsian agents are really choosing, when they opt for his first principle of justice.

### *5. Further Issues and Concluding Remarks*

In this final section we discuss two additional points and offer some closing remarks. First, the simplest way to present the tolerance hierarchy is to set out a base-level behavioral variable which is private, in an intuitive sense. We used the examples of private sexual choices and (at one point in our discussion of Rawls) private religious observance. 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. However, 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 the 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 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 contradictory, or quasi-contradictory in the way some combinations are. Part of the point of this paper 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, in some broad sense, 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. Above 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.<sup>14</sup> 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-

---

<sup>14</sup> See, for example, John Stuart Mill, *On Liberty* (London: John Parker and Son, 1859).

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 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. 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.

## **Appendix to "Tolerance: A Hierarchical Analysis"**

This Appendix gives a more formal treatment of the relations between policies at different levels in the tolerance hierarchy. It includes proofs of two propositions discussed in section 3 of the main text.

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  $i^{\text{th}}$  variant at level  $n$  ( $v_i^n$ ) is deemed desirable and  $p_{v_i^n} = 0$  if the  $i^{\text{th}}$  variant at level  $n$  is deemed undesirable. (Note  $N$  is the number of variants at level  $n$ ).

At some level in the hierarchy, an *initial evaluation* determines which variants are desirable at that level. Here we do not elaborate on what guides this initial evaluation, and do not specify the level at which it occurs (although we mention a few possibilities in the main text and below). However, here we do outline one way in which higher-level desirability might be specified. This 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.

If level  $n$  is evaluated, then we can determine the conditionally perfect and imperfect strategies at level  $m = n + 1$ . But how do we determine what is desirable at level  $m$ ? One way of doing so requires another definition:

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

We are now ready for our first proposition, where we refer to variants at each level as “policies”:

**Proposition 1:** Given an initial evaluation of desirable policies at level  $n$ , a perfectionist group will always have a single desirable intolerant policy at level  $m$ , for all  $m > n$ .

We add a second proposition after one last definition:

**Definition:** A *maximally intolerant* policy at level  $m$  interferes with all variants except one at level  $m - 1$ .

In this definition, we set aside the possibility of interfering with *everything* at the level below. The second proposition is now framed as follows:

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

**Proof of both propositions:** 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:

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

where  $p_{v_x^{n+1}} = 1$  for  $v_x^{n+1}$  that is defined by the  $x^{\text{th}}$  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 conditionally perfect variant at level  $n + 1$  (and all other variants at this level are conditionally imperfect, as they permit undesirable variants at level  $n$ , interfere with desirable variants at level  $n$ , or both). Given that we have a perfectionist group, this single conditionally perfect variant at level  $n + 1$  is desirable and all other conditionally 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  $n + 1$  that 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 conditionally 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 conditionally 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 conditionally perfect  $m^{\text{th}}$  order variant, there must be a single conditionally perfect  $(m + 1)^{\text{th}}$  order variant. Thus, for an initial evaluation at level  $n$  (i.e., 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. ■

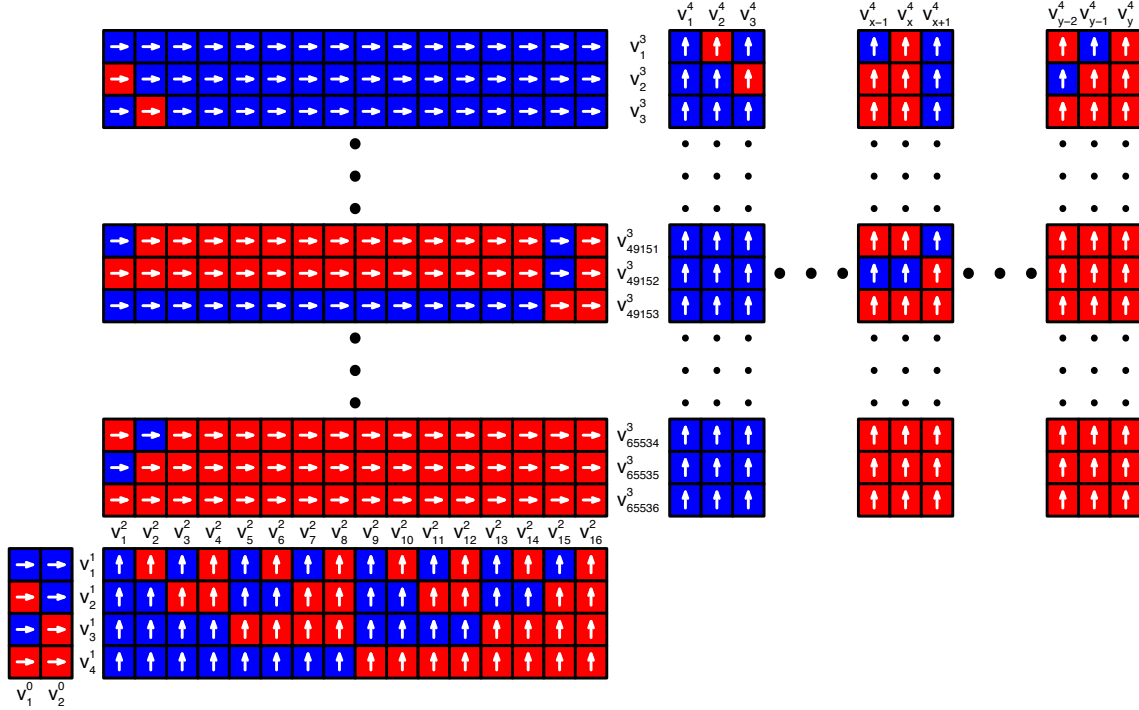
*Note:*

There can be similar perfectionist groups for different reasons and for initial evaluations at different levels. Let's consider two base-level (level 0) variants of a private behavior. For group A, suppose that each variant ( $v_1^0$  and  $v_2^0$ ) was deemed desirable, which leads to the vector  $d_0 = \langle 1, 1 \rangle$ . Now, consider society B, which refrained from evaluating desirability at level 0, but did evaluate as desirable the first-order variant that permitted both base-level variants to be desirable, and evaluated the three other first-order variants to be undesirable. Thus, for group B,  $d_1 = \langle 1, 0, 0, 0 \rangle$  (assuming the tolerant variant was listed first,

followed by the intolerant variants and ending with the unconditionally interfering variant). Now, given these specifications, how would each society view the higher policy levels? Well, using the language of the proof of our theorem, in society A,  $P_1 = \langle 1,0,0,0 \rangle = d_1$ . This means that for all levels above the base level, there is complete agreement on all desirable variants in these two groups. More generally, when the desirability vector for the initial evaluation at some level  $n$  for one perfectionist group leads (through conditionally perfect variants) to the desirability vector for the initial evaluation at some level  $m$  ( $m > n$ ) for another perfectionist group, then the groups will have identical desirable variants at level  $l$  (for all  $l \geq m$ ).

In our groups A and B, there could be very different reasons for the initial evaluations. For instance, individuals in group A may want to promote diversity in the base level behavior. However, individuals in group B may want to promote first-order tolerance even though they may have a preference for one base level variant. These two groups end up desiring similar variants at higher levels if they are both perfectionist.





**Figure:** A visual representation of the hierarchy of variants. We start at the lower left of this 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 or permitting each of the base-level variants. These four variants are visualized by the rows of the  $4 \times 2$  matrix of squares above the base-level variant labels, where blue corresponds to the base-level variant below being permitted and red 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  $4 \times 16$  matrix of squares to 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 (blue) or interfered with (red). 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  $65536 \times 16$  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  $65536 \times 2^{65536}$  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 conditionally 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 conditionally 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 conditionally 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 conditionally 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 conditionally perfect (desirable) variant at each level above the first level.