Tournament judge management

I do wish everything was simple. I've been proselytizing a lot about MJP, including in an NDCA-sponsored Coaches' Corner article in *Rostrum* (http://www.jimmenick.com/vault/ndca_mjp.pdf). I've also brought up some issues on the NDCA listerver, and regularly on the subject in my Coachean Life blog (coachean.blogspot.com). While I've personally been going all-in on MJP, others have been seriously questioning it, especially in the college debate community. The controversy is unsettled, and often unsettling.

I am in tab rooms literally every week of the season, from national events to local starter tournaments. At the latter, anything goes and everything is random. We do whatever it takes to give students as many rounds as possible. On the other hand, major tournaments, with big fields drawing from wide distances, provide a different situation. Attending these tournaments is not cheap, and the prestige of doing well at them might be necessary for a team's support from its local district. A well-stocked debate trophy case is a nice and often necessary complement to a school's traditional collection of athletic awards: it is frequently what keeps the money coming in. Success at these tournaments can also be instrumental in moving up to even more prestigious events like the TOC or the NDCA tournament. At these tournaments, unlike the local starter tournaments, anything does *not* go, and very little is random.

We do debate because of its educational value, and for us, competition is inextricably linked to achieving that educational goal. Therefore competition has to be scrupulously managed so that it is honest, fair and not an end in itself. Maybe we should look at every aspect of a debate tournament before deciding that MJP (or anything else) is good, bad or indifferent. So let's start with the management of judges per se. We have two overarching options in judge management. We can either pay no attention whatsoever to who is judging whom (aside from blocking conflicts) and let it go at that—in other words, totally random judging,—or we can decide to somehow manage the judging assignments.

While it is easy to pooh-pooh the idea of random judging as a not serious proposal for high-stakes events, I would suggest that it should indeed be seriously studied and considered. Simply put, random judging translates into a belief that any judge is as good as any other judge, insofar as rendering a decision is concerned. Granted, some judges are more experienced than others, some have noted affinities with one style or another, some prefer this and some prefer that, but none of these are undeterminable by the debaters going into a round. As I said above, we use random judging regularly at events for younger students; we do this because we believe the most important thing is for them to get as many rounds as they can to learn how to debate in general, to understand arguments and how they work, and to present themselves and their material in a convincing way. We train students in those early days to recognize and understand the different types of judges they will see. We emphasize the need to adapt to those judges. In any sort of public speaking situation, understanding and addressing your audience is a

primary consideration. However, in high school debate, we often remove that traditional aspect of public speaking and concentrate instead on a relatively homogeneous audience. Policy and LD as they are practiced today at their most competitive level, i.e., the high-stakes national circuit, have evolved into specialized events that can only be understood by a specialized audience. I won't argue whether this is good or bad, but if it is bad, one part of the solution would be to go back to totally random judging. At the point where anybody could be your adjudicator, you have to be able to read and appeal to any kind of adjudicator. Shuffling through judge philosophies on the wiki might indicate some great scope among judges, but that's a false impression. Speed alone, which many judges have no issue with, creates an unbridgeable gap between the every-week "professional" judge and the occasional dabbler. Obtuse research sources, off-case critiques, convoluted theory arguments, etc., do likewise. While one might think that, say, Barrack Obama, Hillary Clinton and Andrew Sullivan would make dream judges (and a really bizarre panel), not one of them would be able to follow the average high-stakes high school Policy or LD round at its normal high speed, and as far as LD is concerned, even if the round were written out and presented to them absent delivery issues, they probably *still* wouldn't be able to make heads or tails out of it. You've got to wonder about this.

Anyhow, the first issue, whether judging should be random or managed, boils down to this: aside from PF (which remains primarily adjudicated by lay judges, with a fairly easy-to-prove link that as a result it remains not too fast and not too esoteric), totally random judging would require/effect a complete change in the nature of debate, especially at the national level. Has the proverbial train left the station on these? Is there perhaps simply a separation of national events from "regular" events (those without bids and with exclusively local attendance)?

Before coming to a conclusion, you need to keep in mind one thing. Any totally random system of judge allocation means that in every round, regardless of the competitive situation, a random judge gets the assignment. If teams are on the bubble, are you willing to give them a judge who has never before seen a debate round? You have to be willing to go that distance if you're really going to support random

Competition – Brackets and Power

I'm going to make a simple assertion: if debate tournaments didn't exist, most students wouldn't debate. Yet from the point of view of education, most of the value derived from debate is in the preparation and not the actual competition. Other competitive activities might make comparable claims, but I'd question them. I absolutely believe in the various benefits of athletics, for instance, but nevertheless I see no great benefit to the practice of pitching baseballs if there isn't going to be a baseball game this weekend, whereas studying philosophy or geopolitics or economics has value regardless of this week's debate calendar. This makes debate competition an unusual animal in that, while the competition per se is all about winning, virtually everything else about the activity is not, and everything else about the activity is what the activity exists for.

Be that as it may, and I could ramble on forever about the inherent values of debate, when it comes to tournaments, it is all about the competition. Once you commit to the idea of having competitions, the competitions have to be real and meaningful. The following is a probably not complete but nonetheless essential list of principles for tournament management:

- Debate tournaments need to scrupulously fair.
- They need to reward the competitors who do the best at that tournament.
- They need to run under rules that are clearly presented and fully understandable.
- They need to be inclusive.
- They need to be open and transparent.
- They need to address the needs and concerns of all who attend—contestants, judges and coaches

The first two principles, that tournaments need to be fair and that competitors who do best at that tournament should be rewarded, are procedurally linked. We want a process that addresses only how you debate at this tournament, so in running tournaments we seldom draw on past performance for placement. One exception to this is the NDCA tournament, which does some power protection in the presets based on the points accumulated over the year, the same points that got people into the tournament in the first place, and that makes sense. We don't want, say, the top four competitors, which was theoretically already determined by their points, hitting one another in the first couple of rounds and, perhaps, eliminating one another before things even heat up. But for most tournaments, everyone is equal before the rounds start. There is no pre-event seeding. However, we do want to protect power, insofar as we want the debaters who are debating the best to make it through to the end of elimination rounds as befits their performance. So what we do is use the first couple of rounds to create a seeding for the particular tournament at hand. That is, in common practice at most tournaments, the first two rounds are random (and usually preset to start things off quickly, starting friction at tournaments often being an issue). Anyone can hit anyone, and the chips fall where they may. Occasionally the presumed top debaters do hit one another in presets, but over time random pairings are just that, and it is pretty hard to imagine the top debaters being eliminated in presets; it just doesn't happen, or at least it happens so rarely that no one sits around worrying about it. The NDCA setup is simply a guarantee that it won't happen.

After the presets, we work from a bracket system, where as much as possible, people with a given win-loss record hit other people with the same record. Within the bracket, we most often pair high-low, i.e., the highest seed hits the lowest seed, usually based on speaker points. So on the one hand, you're hitting people in the same position as you, but at the same time, an effort is made to protect power, once again so that the top debaters don't eliminate one another too early. When the numbers in a bracket don't work, we pull up someone from the bracket below. When we're doing this by hand (which I do relatively often for small divisions at one-day events for younger students), we'll pull up from the middle of the lower bracket into the middle of the bracket we're trying to pair. This seems fair and random, but of course we try to minimize the number of pullups, and

at tournaments with big fields, they are indeed few and far between because the large number in any bracket sorts things out without resorting to breaking the brackets. As for elimination rounds, these are absolutely based on seed, top seed hitting bottom seed, second seed hitting second from bottom, etc., again based on power protection. None of this, by the way, insures that the top seeds always win. Far from it. It is simply the accepted way of handing the need to reward the top debaters at a tournament, challenging them but not putting them into a position of eliminating one another so that lesser debaters outlast them in the competition.

I think that, in terms of the principles of tournament management, what I've just described is fair, and that it does reward the competitors who do the best at that tournament. I've seen variations on the theme and run some of those variations myself. For instance, I've seen geographic barriers set, where in the random rounds at a national tournament, the two schools who happen to be from adjacent school districts back home and who traveled 2000 miles to get here won't hit one another until after the presets. This geography can also prevent them from being judged by their usual locals, going past the presets. This also seems right to me, although the judging issue gets subsumed in our further discussion of MJP, to come. The cost of traveling to a tournament is high enough that the teams involved are probably happy not to spend the event battling with their next door neighbors.

Whatever system one chooses to use for pairing a tournament, it needs to be clear to the competitors. Since most of what I've been saying here is pretty standard, no one questions it much, if at all. But if someone is going to run a variation on the theme, like geography or any sort of interference with the natural one hundred percent randomness of the first two presets, then everyone needs to know about it. If there are 3 presets, or if a round is paired high-high, or if rounds are lagged-paired, everyone needs to know about it. We all have an expectation of how a given tournament works, and we want that expectation to be true. It's not so much that we might object to a certain variation, but just that, whatever it is, we deserve to know about it.

This takes us to the realm of transparency in tab rooms. In my career, I have seen tab rooms go from totally locked black boxes to (one hopes) totally transparent operations open to all. As a tabber, I honestly do find that there are times when having an audience is distracting, but it is never prohibited (except at CFL events I run, which have their own rules on tabbing). At the point where I'm doing something I wouldn't want someone to see, I probably shouldn't be in the tab room. This is one of the reasons why a good tab room is run by more than one person from more than one school, not so much because we don't trust one another, but because we want to present to those not in tab a picture of an operation that will, by its very construction, not be biased. We want to be seen as above suspicion. Of course, anyone who has actually watched my usual tab teammates in action know that, far from trying to cook the books, we never even know who the books are about. We deal so much with data as data that we seldom even notice whose data it is. In the middle of a tournament I'll turn to someone like Sheryl Kaczmarek and remark that one of her students is doing real well in a division that she's tabbing, and she'll be surprised to hear it. In the tab room it is all data as data. Which is the way it ought to be.

So maybe something here is controversial, but I don't think so. But we're getting to the controversial stuff soon enough.

(When this was originally published, the value of bracketing per se was challenged. My answer was that, if the goal of a tournament is, as I proposed, that the best debaters do the best, this can only really happen if the best debaters debate are challenged in all their rounds. Bracketing seems to make for the most challenging rounds, at least on a round by round basis. If we did have a totally random tournament, it would affect both the best and the worst of the field. The best would have no guarantee of serious competition, and not only that, but the worst would have no guarantee of serious competition either. Bracketing doesn't just affect the top. Over the space of a long tournament, bracketing means that the less strong debaters also get to find their level, and to have rounds in which they are challenged, as compared to rounds where they are blown away. In a random system, they could compete all weekend and never seriously have a chance. Chris Palmer added that bracketing also has the advantage of being able to predict where the elim break points will lie, since you get a nice even bell curve out of it; a totally random tournament would be more randomly skewed. "Plus, debaters also get more out of close debates. In debate, each kid's strategy and performance is contingent on the opponent's, unlike in a speech round. A debate between a 5-0 kid and a 1-4 kid is just going to be a bloodbath that will fail to challenge the top kid, while it potentially demoralizes and fails to meaningfully educate the victim much. So it's good for moral and education that everyone is slotted rounds of ever-increasing parity as the tournament continues.")

Assigning Judges

Let's stick with the competitive goal of the best debaters at the tournament doing the best at the tournament.

After deciding how to pair the debaters—and brackets are the accepted best way to do this in line with our competitive goal—you have to decide how to judge them. Our first choice is simple. We can either have random assignment of judges, or we can have some sort of non-random assignment. I talked about the ramifications of this choice above. If you offer a tournament with random judging, that means that you are prioritizing judge adaptation as the most important skill for the competitors, and that the measure of the competitive goal of being the "best debaters at the tournament doing the best" is adeptness at appealing to random judges. A goal like this for a tournament theoretically posits no particular paradigm of what good debate ought to be, aside from that it ought to appeal to a wide body of adjudicators (assuming that your judge pool is fairly heterogeneous). The other choice, non-random assignment of judges, does posit a paradigm of what good debate ought to be, to wit, the ability to appeal to a selected group of adjudicators. The assumption here is that certain adjudicators are "good" or "bad" and that we will somehow screen out the latter in favor of the former. That group of judges we define as good are the people who get to decide what sort of debating is good. This

raises two questions: Who are the "we" who are picking these judges, since this "we" is doing the prioritization of the judges, and then, who, indeed, are these judges.

(I don't think we have to talk more about what happens with the randomly managed tournament. For all practical purposes we already live in that world with PF, and will continue to do so for the foreseeable future. That randomness of judging is inherent to PF, and if it were to go away, PF would become merely a minor variation on a theme rather than a full activity in its own right. Personally, I like the effect random judging has on that activity. Managed judging in PF, aside from offering a handful of strikes, would in my opinion be a very bad idea.)

So let's look at those two questions in judge management: who is picking the judges, and what judges are they picking. This needs to be considered in the light of managing the competition with brackets. As already discussed, the system of random presets leading to brackets seems to create the best chance of the best debaters advancing at a tournament. In this system, at any point after the presets there are debaters on the bubble; that is, debaters who, if they lose this round, will be unable to advance to the elimination rounds. As often as not, this point is when a debater is down two; down-twos can advance, down-threes are out of it. So the focus of judge management, the place where we most want the perceived best judges, is in those down-two bubble rounds. We want the perceived best judges adjudicating the most important rounds. That is consistent with wanting the best debaters at the tournament to do the best at the tournament.

For all of the years I tabbed using the TRPC software, you could rank each judge as an A (best), B or C. If you were ranking judges, you could also click on the situation (down one or down two) where you wanted the software to place the most highly ranked judges. After a pairing was done, everyone on the bubble would be getting an A judge (assuming there were enough to go around).

As a general rule, the determination of the As, Bs and Cs was done by the tab staff in collaboration with the tournament staff. You'd sit down before the tournament and run through the judges, assigning everyone a ranking. On the one hand, this can be seen as having a very experienced group of coaches reporting on their evaluations over time of the individuals in the pool and categorizing them accordingly. Except, of course, most of this experience was from how those judges handled the tab staff's own debaters, and of the tab staff's opinions of what had happened in rounds they themselves weren't in, either through hearsay or glancing at ballots as they came in. Having people from different schools do the ranking may have made for some sense of fairness, but let's face it, this is a biased system and is probably not optimal. If you judge week in and week out, for instance, just the familiarity of your name would probably earn you a higher ranking than if they had no other idea of you as an individual, whereas in fact you might be the biggest clown ever to pop on a red nose. If you gave one of their debaters an unfair loss five years ago when you were first starting out, are they taking into consideration all the learning you've done since, or are they still aggrieved over that unearned loss way in the past? Does the fact that you won TOC make you a good judge? Does the fact that you *didn't* make it to TOC make you a bad judge? You're old, you're young, you're

experienced, you're inexperienced, you're from here, you're from there, etc., etc, etc. Yes, tab would do the rankings, but based on a lot of occasionally mutton-headed thinking. And the thing is, it's not so much that the determinations were based on mutton-headed thinking, as it was who were the mutton-heads. Why should the tab staff get to do this? By no means do they have a totally objective and well-informed view of every judge in every pool. Why should they get to rank those judges? Just because they're in the tab room? That's not good enough.

If tab doesn't determine who the "best" judges are (and I may not always use those quotation marks, but please consider them to be implicit in my every use here of words like good or best), who does? Well, if it isn't the people running the tournament, it's the people *at* the tournament, i.e., the teams themselves. Since the ones who are at the tournament have the most at stake—remember, we demand complete neutrality from the tab staff—why not let them decide?

Let's step back a bit. The bubbles determine who breaks. Either somebody considered a good judge is going to adjudicate the bubble, or some random person is going to adjudicate the bubble. If you put in some random person, so be it, but that means that you have chosen for your tournament to run a very specific way. You can't go in and say, it will be random, except let's not use that judge there for those debaters in that round. That's not random, and in fact, it is specifically and demonstrably biased. It's got to be all or nothing. All random, or not. You can't have it both ways and still run a fair tournament.

One underlying question in this post is the nature of good debate. If, presumably, good debate is winning debate, then whatever wins is good. Winning is determined by the judges, so the judges get to decide what good debate is. The people picking the judges, knowing fairly well what it is that the judges consider good debate to be, determine what kind of debate will win. Following this logic, a system where the tournament staff picks the judges is an oligarchy of sorts where an established elite attempts to set the rules. A system where the debaters pick the judges is a system where the those with the most at stake competitively (and the least at stake educationally) decide what good debate is. Some people think this is fine; some people think this is letting the inmates run the asylum. In any case, short of random judge assignments, someone is going to decide what good debate is through the assignment of the judges.

For a while in the northeast we tried a system of Community Rankings, where a list of judges was sent to all the entrants and the results were compiled and we averaged out the rankings we received, a sort of metadata approach. What we saw with this was that judges were usually ranked the same as the tab staff would have done on their own; this might have been due to the fact that, as this was handled on a by-school basis, the coaches did the ranking, and the coaches not in tab weren't all that different from the coaches who were in tab. While this justified what was already being done by the tab rooms themselves, it had a serious internal flaw: it only worked for members of the local community. We tried it at Yale one year, for instance. Yale draws people from around the country, but most of the LDers were eastern seaboard folks, and they were fairly familiar

with the judges in the pool. The teams from places like California, on the other hand, even if they knew the judges via paradigms, were overwhelmed by the tyranny of the local majority, the existing mindset on the judges that was already in place. As a result, they were disadvantaged from the start, captive to judging presumptions common and well known in a region that was not their own. The bubble judges were not merely familiar to the locals (that would be a hard one ever to overcome), but they had also been favored by the locals. That was a bad combination. It simply wasn't fair (and fairness is one of our important operating premises).

MJP seemed to solve that problem. Each individual debater would have his or her own ranking of the judges. It's a system that would attempt to provide the most mutually preferred judges to all, prioritizing the bubble rounds. In terms of fairness, it seemed complete. Bubbles would get their best possible adjudicators. Everyone else would get their best possible adjudicators as well, after the most important rounds were set. If nothing else, everybody was in the same boat regarding who was in the back of the room.

MJP in Practice

For some people, MJP (or MPJ, as many refer to it) is the hand basket in which debate is merrily making its way to hell. I don't agree, at least at the high school level, after having tabbed a few years' worth of tournaments employing it. But it is also not perfect, and like many things in tournament management, it could use a set of best practices. (As with the rest of this essay, this is entirely my own personal take on things.)

Mutual Judge Preferences, as previously noted, places the determination of which judges adjudicate the rounds into the hands of the competitors. This is done because the two alternatives are less desirable. The first alternative, totally random judging, ignores the competitive rationale of a tournament, i.e., rewarding the debaters who do the best at that tournament. The second alternative, tab room ranking of the judges, is innately biased based on the prejudices of the small group of people in the tab room. This is not to say that MJP won't play into prejudices, but at least they're the prejudices of the people in the trenches. If we're going to play into anyone's prejudices, theirs make the most sense.

The first step in MJP is deciding how to organize the tiers. In TRPC, there are 6 or 9 tiers (although 9 don't work for LD because of flighting, and if I'm not mistaken, they only work in prelims for Policy), although you can zero out a tier if you want fewer. In tabroom.com, you can set as many tiers as you want. Other systems presumably work the same way. The tiers are hierarchal, from 1 (best) to 6 or whatever (strike). At first glance, it might make sense to simply divvy up your pool into 6 (or however many, and I'll stop parenthesizing this and keep it implied) equal tiers, but that's probably ill-advised as far as strikes are concerned. 13% strikes is a lot of strikes, with increasing damage to your ability to tab as the size of the pool shrinks. Which raises the question, what is the point of strikes?

In any team's history, there are some judges who always look upon them favorably, for whatever reason, and some judges, also for whatever reason, who haven't picked them up since the Eisenhower Administration. It could be a style issue, it could be a personality clash, it could be anything. When I was very actively judging, there were some students I dropped over and over again, tournament after tournament, sometimes without even being aware of how poorly they were faring before me. I had nothing against them, since I often didn't recognize them from one tournament to the next (remember, I kept dropping them, so they couldn't have been making that much of an impression), but still I dropped them time and time again. Who knows why? The thing is, if they had had the opportunity to strike me (striking was fairly unusual back then), that's exactly what they should have done. There are just those judges you can't win, or maybe a couple you know are predisposed against you. A handful of strikes allows you to eliminate those people who are virtually guaranteed to drop you.

As I said, strikes used to be rare. Now they're commonplace. You're even seeing them gain acceptance in PF, which otherwise remains firmly based on random judging, probably for the reasons cited above. We let everybody trim a tiny portion of the pool in their own favor to prevent a virtually guaranteed loss. Most people think that's a good idea.

So the first thing we do in MJP is put a small set of strikes into the bottom tier. (These are different from conflicts, a whole separate subject I have discussed this at length elsewhere, with clarifying documentation at http://www.jimmenick.com/vault/conflicts_teams.pdf and http://www.jimmenick.com/vault/conflicts_judges.pdf.) And here I think we need to establish something that most people seem to miss when we're talking about MJP. Most tournaments that we're talking about here—invitationals, college tournaments, circuit events—tend to have roughly 80 to 120 participants in a field. There are bigger and smaller, but that's about the average. And with that size field, you usually have about 40 to 50 judges in the pool. These numbers are way too small to base any serious quantitative analysis. Almost all the complaints I hear about MJP, if they applied at all, would only apply if the numbers were vastly larger than what we're actually dealing with. For argument's sake, and because it's average enough, let's go with a working field of 100 with a pool of 45 judges as the standard for this discussion.

Back to the small set of strikes. Out of a pool of 45, you're talking 4 or 5 stinkers you want to cross out from the start. Not a lot, and not unreasonable. So we put in 10% strikes, i.e., you can strike 10% of the field. You are now left with about 40 judges.

The next decision raises the question of how, exactly, you are viewing MJP and how you wish it to work at the tournament. Every decision you make henceforth will have an impact on how the tournament proceeds, so you need to venture carefully. And always remember this: Mutual Judge Preferences only means that the two opponents agree on the desirability of the judge, and in a tournament, tab will do its best to find the most highly preferred judge for the debaters. It absolutely does not mean that the debaters get to pick their judges, except insofar as the opponents agree on how desirable that judge is. They

didn't just buy a win, in other words, i.e., a favorable judge. First of all, the judge is equally favorable to both sides, and second of all, might not be all that favorable to either. While competitive requirements will work toward debaters getting a high amount of desirability in in-contention rounds, MJP is no guarantee of desirability in those rounds. It only guarantees mutuality. If you're really talking about mutual preferences, a 4-4 is as mutual as a 1-1. On the other hand, a 4-4 is mutual and a 1-2 isn't. In a world of mutuality, a 5-5 is better than a 1-2. The last possible placement of a judge in an MJP tournament, only after every mutual possibility is exhausted, would be a non-mutual assignment. I say this to disabuse people of the idea that mutuality implies high desirability. Desirability and mutuality are two different things. With just about everybody preffing nowadays at the big tournaments I tab, we're seeing plenty of mutual 2s and 3s, and occasional mutual 4s and 5s. If two people don't agree on their 1s, then we find them 2s, or 3s, or 4s, or 5s, in that order. That's the way it ought to work. If for some reason you decide that any unequal pref should be prioritized over an equal pref, then you are not offering MJP at the tournament, and this needs to be clearly advertised.

MJP must honor mutuality at every level. Otherwise it is not mutual, and we should not pretend that it is. There are some who say that they would rather be on the wrong side of a 2-1 pairing than get a 4-4 or a dreaded 5-5. Of course they would. But at the point where we're tossing mutuality out the window so that more desirable judges rather than equally desirable judges will adjudicate, we are abandoning any sense of expecting debaters to pick up diverse ballots. This also tends to isolate a large number of judges who average low in overall preference, who might now never get a ballot. No, you don't want a 4-4. On the other hand, neither does your opponent. Fine. That's what you got, go deal with it. You, and debate in general, will be all the better for it. Strict mutuality is the strongest answer around to most of the perceived ills of MJP. MJP is not favorable to anyone in particular when all teams pref and all judge assignments are mutual. Strict universal MJP will cause many debaters to leave their comfort zone and attempt to pick up judges not predisposed to favor them or their styles. (I've never heard a serious debate coach suggest that debaters in general should not be able to do this, except during tournaments, where many of those same coaches complain every time they don't get a 1.) And keep in mind that, if you're doing any sort of ranking of judges in tab and not running a totally random assignment of judges, then someone's bias somehow will be reflected in each assignment. Isn't it best that the bias be the equal bias of the debaters who have to stand in front of that judge?

Back to the mechanics. At some point, you have to set up the various preference tiers. Ideally, MJP should allow debaters to make meaningful distinctions. But at the same time, there are ramifications to whatever you do in setting up. In our theoretical average pool of 40 judges, going strictly by the math, we could probably do either 4 tiers or 5 tiers, i.e., 4 tiers of 10 each or 5 tiers of 8 each. There are many things to keep in mind in choosing between the two. The more tiers there are, the more granular the debaters' decisions in assigning judges to the tiers: if there's 8 of any category vs. 10 of any category, the assignment of the categories can be perceived as more "real" by the debaters. That is, their 1s are their real 1s, their 2s are their real 2s, etc. Or at least they perceive it this way. I would maintain that there's no perfect percentage, and certainly the

real-ness changes if, let's say, you have 100 judges at a really big tournament. Are there really 20 1s in that pool? So yes, this is ultimately asking debaters to pref at a line of arbitrary distinction, but honestly, any number of judges will force that sort of arbitrary distinction sooner or later. But the smaller number of judges and the corresponding smaller number of tiers would seem to minimize it. It would appear, therefore, that your best bet would be doing 4 tiers of each on our theoretical average pool.

Keep in mind that there are ramifications to the choice on the back end, i.e., in tabbing. The smaller the number of judges in any debater's given tiers, the harder it is to assign judges from the more preferred of those tiers. In other words, if you need to find a judge for two teams and each has 8 1s, it will be harder to find a match than if each has 10 1s. That part of the math is simple. I would suggest that tiers smaller than 8 will become pretty problematic, but I can't honestly say I've had much experience below that. At my own annual tournament, with about that number of 40 judges after strikes, I go with 4 tiers of 10 each, on the belief that it's a good balance between meaningful tiers to the debaters and workable tiers for tabbing.

You can, of course, play fast and loose with this. You don't have to give equal percentages in each tier, for instance. You can make the setup 50% 1s and then make the other 40% (after 10% strikes) 2s through 5s, in equal percentages or even, if you're so inclined, in descending percentages. But at this point there is no question that you've pretty much done two things: first, you've eliminated any meaningful distinction among the judges, rendering the idea of mutuality moot (no pun intended), and you've almost completely ghettoized all the judges not in the top tier because it will be so easy to assign all 1s in tab that no one will ever see a 2, much less a 5. But of course, it won't be anyone's "real" 1s, so the benefit is only in the tab room. As for ghettoization, that's an issue anyhow, marginally referred to once or twice above, which I'll get to shortly; it's only aggravated further by this sort of number juggling.

So let's make the distinctions as real as we can, with equal distribution over 4 or 5 tiers, depending on the size of the pool. We could, of course, add even more distinctions, the 9 for instance that TRPC allows policy prelims, but I think this is just as problematic in the opposite direction from the above unless the pool is remarkably large (100 judges as compared to 40). Personally, I think a standard expectation should be set that a tournament will have 4 or 5 tiers, just so people see some consistency from week to week. Secondly, I believe that the more tiers you offer, the less real they become, according to the proven marketing idea that too much choice is undesirable for most consumers. 9 tiers in our theoretical pool of 45 means 5 of each from 1 to 9; if we were ranking the judges according to an objective criterion like height, fine, but the manifestly vague criteria by which we pref judges can't possibly stand up to this many tiers. I mean, how do you distinguish your 6s from your 7s? I would also maintain that ordinal preferences, simply ranking the judges from 1 to 45, is a similar mug's game of artificial distinctions. If our goal is to create a tournament with the best competitive structure, and we believe that some sort of judge preferencing is in aid of that best structure, then the judge preferencing needs to be as meaningful as we can make it. The more tiers there are, the less meaningful the preferencing. So I'd stick to 4 or 5 tiers, and consider the job

done. (Palmer strongly disagrees with this, claiming that reality is that any pool just has so many 1 judges no matter how you slice it; an ordinal arrangement might be better for him than a tiered arrangement. In any case, this is not material that fits neatly into a cut-and-dried solution, and time will tell how best to handle tiers after we've all tabbed a lot more tournaments.)

Next, we need to talk about that ghettoization of the pool. As I mentioned earlier, we used to use a community ranking system, where all the schools ranked the judges from best to worst (A, B or C). The thing was, at every tournament, the same folks were in the same category. This holds true in MJP, where the most preferred judges are the same from tournament to tournament, as are the least preferred. The why of this doesn't matter; the point is, most people who attend debate tournaments look at the judge pool roughly the same way. What this means in practice with MJP is that there are some judges who can go from tournament to tournament and, in a strictly organized MJP tiering, hardly ever get a ballot. If you're everyone's 4 or 5, unless something is done about it, you're going to be sitting in the judges' lounge with a novel way more often than you're sitting in the back of the room with a ballot.

This is a problem that needs to be addressed, and one that is probably aggravated by MJP, and especially aggravated by pseudo-MJP with uneven tiers. Whereas in the past with simple ranking of A to C, if I saw a weaker judge not getting any rounds I could just toss him or her in to a non-competitive round, doing this sort of random intervention is counter to the whole philosophy of MJP. But something must be done. Regardless of who makes up the less preferred judges (and there's a possibility for much separate discussion of this), any relatively unvarying bloc of less preferred judges will, eventually, be problematic. My recommendation is that, at the point where a team is out of the competition for breaking into elims, we use the non-preferred judges. That is, we throw MJP out the window for teams that can't break. This will help solve the ghettoization problem, although it's not a perfect solution because, obviously, the ghettoized judges will be getting what can be seen from their perspective as the least preferable rounds to judge. Another problem is that, for the debaters, they only get preferred judges some of the time. I have made the argument in the past that every debater has paid the same amount of money to compete at the tournament, and therefore ought to have the same expectation of the best possible judging for the duration. That argument ultimately doesn't hold, however, in the face of other issues facing the tournament. First, there are only so many judges to go around, and if we're upholding competitive viability we would never choose to give a down-3 pairing a 1-1 judge over giving that same 1-1 judge to a down-2 pairing still in the running for elims. Secondly, the ghettoization of some of the pool, for whatever reasons, is intrinsically a bad thing and should be prevented as much as we can. From behind a Rawlsian veil of ignorance, I think most debaters would agree that keeping the best judges for the most important rounds and preserving the integrity of the less-preferred judges overall are both reasonable ends warranting that, if you're out of it, you drop preferences.

(There are on-the-side solutions to the judge ghettoization problem, of course, especially for the judges who are newcomers and therefore less desirable simply because they are

less known. One does have the power in tab to assign them to a different pool, namely PF, if the tournament is offering it, while they are still fulfilling their obligation in the original pool. A perceived weak judge in LD can be a very useful judge in PF. I'm thinking now mostly of parents, smart and ready to judge and sitting in the judges' lounge for two straight days reading the complete works of Dickens when they could be in rounds doing good judging work. Given the opportunity to switch and be useful, most folks will grab it, and it you ask me, the very grabbing is an indication of their great usability. And, of course, it there's a novice LD pool, the same holds. These solutions, however, require preplanning on the part of the tournament. You can't just juggle people around in the middle of things and hope everybody will be happy with it. But offer it as a pre-tournament option, and people might be very happy.)

The final important thing to understand about MJP is how it works in break rounds. Whereas in prelims all we needed was one judge mutually agreed to, now we need three. And the stakes—elimination—are higher. Theoretically, it's all just about the math, but occasionally there's a little more to it that needs to be understood and dealt with.

The ideal three-person panel is 1-1, 1-1, 1-1. This happens plenty of times, but just as often, there are not enough 1-1s to go around. In practice, most break rounds are 1-1, 1-1, 2-2 or 1-1, 2-2, 2-2 after we've evenly distributed the most highly preffed judges. So far, no problem. And sometimes we have no choice but to throw a 3-3 in there. Still no problem, since the mutuality of the pairing forces the debaters to deal with the judges in a similar fashion. The problems arise when we run out of straight-out mutuality (all rankings equal) and go to arithmetic mutuality (all the rankings add up equally). A 1-1, 1-2, 2-1 is an example of the latter, where both sides add up to 5, i.e., it's a total 5-5 pairing. This happens fairly regularly, because the number of mutual judges is finite. This is still mutual, the way I see it. For that matter, so is 1-4, 3-3. 4-1. It's 8-8, and I would scramble to try to fix this because there's only one truly mutual judge out of three; still, in the end, it's one judge you like and one you don't like on both sides, plus one in the middle and it's a mutual situation for both sides, so it is acceptable. But here's one that is unacceptable: 1-2, 1-2, 3-1. It adds up 5-5, but in reality it's 2-1 because one side gets two 1s and one side gets one 1. I've seen 3-1, 3-1, 1-5. It adds up 8-8, but it is even more weighted in one direction than the first example. Assignments like this are simply not acceptable. When they've arisen I've resorted to the old-fashioned idea of assigning 5 judges and allowing each side to strike one. This only happens late in a tournament, when one is really up against the wall resource-wise, but it does happen. The tab room needs to keep an eye open for it.

Best Practices

So let's summarize best practices. The following can be distributed as a short, manageable document. This is far from exhaustive, but it can't hurt as a starting place for tab rooms trying to figure things out fairly.

Tabbing Best Practices

These are our operating principles in offering debate competition:

> Debate tournaments need to be scrupulously fair.
> They need to reward the competitors who do the best at that tournament.
> They need to run under rules that are clearly presented and fully understandable.
> They need to be inclusive.
> They need to be open and transparent.
> They need to address the needs and concerns of all who attend—contestants, judges and coaches.

1. Whatever system one chooses to use for pairing a tournament, it needs to be clear to the competitors. Publish an explanation of how you'll be tabbing on your invitation.

2. To insure neutrality, tab must include more than one person from more than one school.

3. Standard pairing procedure is 2 presets, followed by high-low based on points within brackets.

4. If judging assignments are random, they must be *entirely* random in every round for every pairing. Picking off the top, i.e., going with whatever the computer assigns, or taking off the top of a list if given alternatives, is the only acceptable system. PF rounds will, as a rule, be tabbed in this fashion. Some provision may be allowed to give judges rounds off to even out the burdens. And in random divisions, it is desirable to allow a handful of strikes, for teams to block judges they think may be biased against them.

5. If we do not prefer a random system, the perceived best judges should adjudicate the most important rounds; this is consistent with the principle of wanting the best debaters at the tournament to do the best at the tournament. Since any tab staff has its own biases about who are the "best" judges, the ones who are have the most at stake—the teams debating at the tournament—should decide who those best judges are for their most important rounds. Mutual Judge Preferences allow the teams to do this. MJP means that the two opponents agree on the desirability of the judge; during the tournament, tab will do its best to find the most highly preferred judge for the debaters congruent with the competitive necessities of the round at hand.

6. MJP works best when *all* teams pref. Distribution of this article (http://www.jimmenick.com/vault/ndca_mjp.pdf) or something similar is recommended for tournaments where preffing is not universally understood or accepted.

7. Tournaments should ask registrants to mark their own judges as Circuit, Traditional or New, as a general guide for other teams unfamiliar with the judge pool and/or new to preffing. Tournaments may or may not enforce the publication of paradigms.

8. MJP will offer meaningful tiers for organizing the judges. A suggested range is 8-10 judges per tier, in the number of tiers necessary to contain that number of judges across

the pool, plus a small set of strikes (10%). Each tier, aside from strikes, will be of equal size. (E.g., with 40 judges, 4 strikes + 4 tiers of 9 judges each; 60 judges, 6 strikes + 6 tiers of 9 judges each.)

9. Mutual is mutual: The last possible placement of a judge in an MJP tournament, only after every mutual possibility is exhausted, would be a non-mutual assignment. Placements in order: 1-1, 2-2, 3-3, 4-4, etc., until all tiers are exhausted, then 1-2/2-1, 2-3/3-2, etc., then 1-3/3-1, 2-4/4-2 etc. If for some reason you decide that any unequal pref should be prioritized over an equal pref, you are *not* offering MJP at the tournament.

10. Assignments of the most preferred judges will be into the most important rounds, congruent with maintaining the competitive nature of the event. As a rule that means that first pref goes to Down 2, then Down 1, then Down 0, then Down 3, Down 4, etc.

11. A tournament can stop honoring prefs when teams are out of range of elims, if this will help prevent ghettoization of lower preffed judges. It is also recommended that, if a tournament can, it offers to place judges in advance of the tournament into other pools (PF or novice/JV divisions), where lower preffed judges are probably happier anyhow. Ghettoization of judges for any reason (and going off mutuality with 1-2s rather than 4-4s is a major way to increase underutilization of lower preffed judges) is to be avoided.

12) Care needs to be taken in break rounds for maintaining mutuality, as occasionally the numbers look better than they are. The problems arise when we run out of straight-out mutuality (all rankings equal) and go to arithmetic mutuality (all the rankings add up equally). A 1-1, 1-2, 2-1 is an example of the latter, where both sides add up to 5, i.e., it's a total 5-5 pairing. This happens fairly regularly, and is still mutual. So is 1-4, 3-3. 4-1. It's 8-8, and tab should try to improve, but in the end, it's one judge you like and one you don't like on both sides, plus one in the middle and it's a mutual situation for both sides, so it is acceptable. But here's one that is unacceptable: 3-1, 3-1, 1-5. It adds up 7-7, but it is totally weighted in one direction. In these situations, use the old-fashioned solution of assigning five judges and allowing each side to strike one of them. This usually only happens late in a tournament, when one is really up against the wall resource-wise, but it does happen.

13. At highly competitive circuit tournaments, tab may prefer an ordinal system of preferencing. This is a form of MJP that allows more mathematical flexibility, but is not advised for most events at this time due to its unfamiliarity to the community at large. (It may be the inevitable step after MJP is universal, but not yet.) Use of ordinals should roughly align with the tier system discussed above.

(Full documentation supporting these Best Practices is at http://www.jimmenick.com/vault/TJM_full.pdf.)