Board Game Ethics

Reconsidering our rush
to turn **Ethics** into **Math**

Last Updated
Jan 25, 2026

Table of Contents

0. What is this booklet?

Hi, I'm Ricky, a philosopher embedded in the Intelligence Amplification Lab at the Johns Hopkins Center for Language and Speech Processing.

In that role, I've been invited to give guest lectures on AI ethics for computer science courses and labs. The number one question I get is, when two ethical principles conflict, how do I decide which to prioritize? The number two question is, where should I start reading about ethics?

I've created this short booklet to answer both questions, and to facilitate more grounded conversations. Through repeated interactions with freshmen, graduate students, and faculty in computer science, I've encountered the widespread assumption that ethics is either an optimization problem, or a set of constraints on one.

So in this booklet, I introduce a form of **precautionary pluralism**, the idea that we shouldn't be too hasty to optimize for any one notion of ethical value. If we start optimizing for X (and treating everything else as merely useful for X), we're very likely to throw the baby out with the bathwater. So instead of thinking about how to pursue ethical *value*, I encourage the reader to begin considering tradeoffs and tensions between different ethical *values*.

Maybe this precaution is ultimately misplaced. It might be the case that **monism** is true—that at bottom, there is only one form of ethical value, and our incomplete, contradictory language leads us astray here. But my precaution is much more practically oriented. We're still not very good at ethics in practice or theory, and our disagreements run *deep*. In the meantime, we shouldn't engineer solutions that breeze past the conflicting historical, cultural, social, and political forces at work within our ethical discourse.

If you have thoughts/feedback on this living document, please let me know!

Ricky Mouser
rmouser1@jh.edu

1. So, what *is* Board Game Ethics?

Real-life ethical decisions can be ***agonizingly difficult***. Usually, *many* parties have conflicting interests at stake.

But we have to do *something*. **So how should we choose?**

Ideally, we'd like to have **fixed, agreed-upon procedures** for coming to and justifying the ethical decisions we make.

If we could just agree on these procedures in advance, we could determine the ***Best*** thing to do according to fixed, uncontroversial standards, even when things get complicated!

We do this all the time when we ask panels of judges to score gymnasts, musicians, or coffees. These experts ***translate*** their **rich assessments:**

"a **wonderful gingerbread** spice with a **light fruit** tone, **almost grapefruity** in flavor and reminiscent of a **spiced lemon cookie**"[1]

into **legal moves** within the procedure:

Color: 9.5/10
Aroma: 9.0/10
…
Overall: 91/100

But no matter how much I like a coffee's aroma, I can't give it an 11/10. We have rules here! And that *rigidity* can be a real cost.

[1] Liberman, *Tasting Coffee*, 268-269.

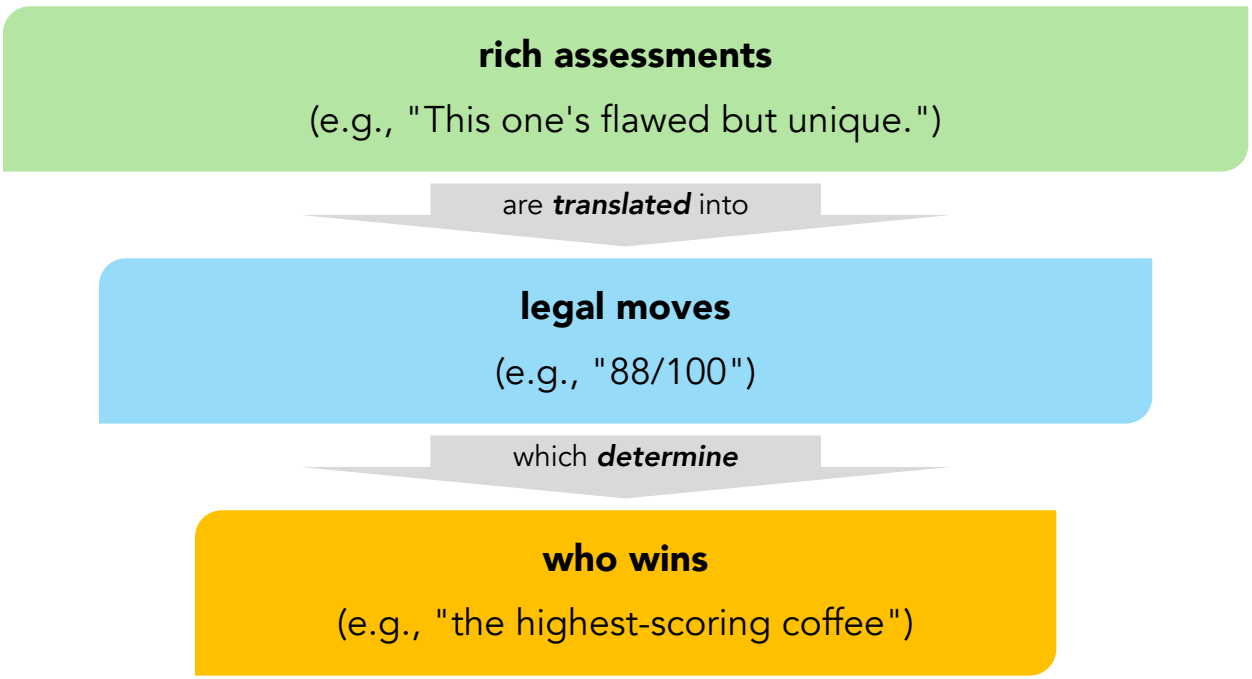
The benefit is that by defining legal moves in advance and sticking to them, we can let the procedure ***determine*** **who wins** for us.

Maybe we eliminate each contestant's highest and lowest scores, average the rest, and crown the **highest scorer** the *Winner*.

Maybe all contestants scoring **90 or higher** are named *Presidential Cups*.

In any event, because these procedures are **formulated clearly** and **agreed upon in advance**, contestants agree to accept the judges' final decision, however the scoring procedures shake out.

Putting it all together:



Good rulebooks help us arbitrate **predictable disagreements** *well enough*.

When we hold competitions or audits or review boards regularly, we can tinker with these procedures to help mold the rules to their local contexts.

But if we could settle on **standardized procedures** for making ethical decisions *more broadly*—when deciding how to raise and spend tax dollars, what sexual education should look like, or how to arbitrate neighborly disputes—that sure could help us bypass a lot of ethical disagreement!

A generalized **ethical rulebook** like this would be a dream come true for bureaucratic decision-makers everywhere. When questioned, they could point to clear decision-making procedures (*policies!*) to explain *what* they've chosen and *why*.

Of course, we know our procedures are never *perfect*. They're negotiated compromises that are hopefully good enough for our local purposes. But they don't always capture what we care about, because our rich assessments **don't always translate well** into the legal moves of fixed procedure.

We run into this problem all the time!

Pretend we have to decide between one of two airport screening policies: Policy S emphasizes **security**, and Policy P emphasizes **privacy**.

Before we even get into any of the details of how the policies differ, notice that **security** and **privacy** seem like **two different *kinds* of goods** raising very different concerns.

Take a peek at what the arguments on each side end up highlighting:

Security prioritizes	Privacy prioritizes
keeping everyone ***safe*** from violent wrongdoers	keeping violent authorities out of ***my business***
avoiding preventable ***deaths*** or ***injuries***	avoiding preventable ***incursions*** into private life
making air travel more ***transparent*** and ***legible***	making air travel less ***surveilled*** and ***recorded***
making air travel more ***consistent*** and ***predictable***	making air travel less ***managed*** and ***controlled***

Talk about rich assessments!
Our values come in tangles, and *lots* are involved here.

Keeping folks safe from physical harm involves creating a more stable, predictable flying environment. Those values are inextricably bound up with producing **security**.

But respecting folks' **privacy** involves conceding oversight and data collection about them. It requires letting go of control and information.

So, what do we do when **security** and **privacy** conflict?
What sort of **fixed procedure** could help us decide what to do?

Well, we can't *really* compare S units of **security** against P units of **privacy**.
What would we do next, add them together?

Policy S	Policy P
Security: 10 points	Security: 9 points
Privacy: 8 points	Privacy: 10 points
Overall: 18 points	Overall: 19 points

Or should **security** be worth *twice* as much as **privacy**?

Policy S	Policy P
Security: 20 points	Security: 18 points
Privacy: 8 points	Privacy: 10 points
Overall: 28 points	Overall: 28 points

Oh no, it's a tie! Now what do we do?
(Hopefully abandon this radically oversimplified approach.)

In real life, these units don't make any sense because that's not how words work. The words we inherit like **security** and **privacy** don't seem to name **quantities**, or *amounts* of things out there in the world.

Our values seem to name **qualities**,
or *kinds* of things we care about,
not quantities,
or *amounts* of things we can score.

Of course, folks end up inventing quantitative metrics all the time, and we end up relying on them quite a bit. When we lack **trust** or **insight**, we can often come up with numbers to help us make *evidence-based decisions*.

That's why governments end up chasing *GDP growth*, and your *credit score* ends up determining where you live and what you drive.

Fixed, agreed-upon procedures and metrics are ***convenient*** **when they work.**

But we regularly find that these artificial practices **fail to capture what we *really* care about**. A three-strikes law putting a jaywalker in prison for life is probably not advancing the cause of **security**, much less ***justice!***

We stand to lose *so much* when translating our rich ethical assessments into legal moves within fixed, agreed-upon procedures. The **violence of translation** doesn't just change our values; it ***flattens*** them.

But maybe universally agreed-upon ethical procedures were always a fantasy.

After all, we *don't* agree on how to arbitrate conflicts between our ethical values in advance. We may not even agree what our ethical values *are!*

But here's one thing we do know:

We don't understand our values *perfectly*.
They are still **works in progress!**

Our conceptions of ***democracy*** or ***equality*** today aren't perfect, but we've come a long way from submitting to the Divine Right of Kings or arguing over whether slavery is natural. Our ongoing disagreements are mostly about how our commitments can continue to be *deepened*, *clarified*, and *improved*.

But fixed procedures assume that we already know our values well enough to **decide in advance** how any possible conflict between them ought to be resolved. By trying to stipulate these standards in advance, fixed procedures end up subtly *terminating* or even *bypassing* ethical inquiry altogether.

Our real values are much subtler than this. We might even be surprised where they end up leading us!

After all, **security** might require keeping me safe from unnecessary interactions with violent authorities, particularly if I'm a person of color. Suddenly, protecting my personal **security** may conflict with advancing the more abstract notion of **security** that Policy S was tracking.

A bit more fancifully, **privacy** might prioritize allowing a *bit* of initial intrusion before we fly so passengers can't bring X-ray glasses on the flight. So, **privacy** may well have to make compromises with **security** to survive!

But these are **qualitative arguments**, not **quantitative scores**, so they're much harder to assess. Here's a ***better question:***

Q: How do we determine where the arguments lead?

A: We have to **philosophize together, not *calculate*** which argument wins!

We can begin by reexamining how fixed procedures pick winners and losers.

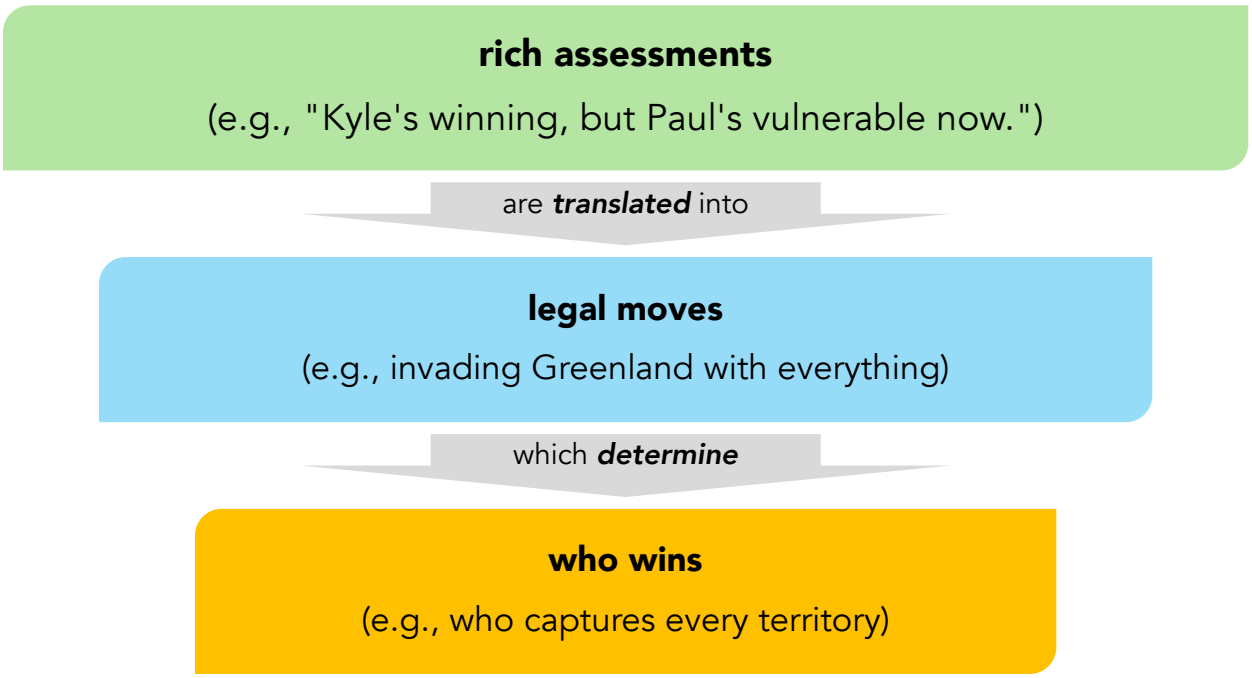
Maybe fixed, agreed-upon procedures could work perfectly—if the underlying practices (gymnastics, jazz, coffeemaking, ethics…) were designed from the ground up, **like a board game**.

After all, a board game *just is* a fixed, agreed-upon procedure for inputting **legal moves** and outputting **winners and losers**.

And the **rulebook** just tells us how those procedures work!

Still not convinced? Take a look:

The board game *Risk* as a fixed procedure for picking winners:

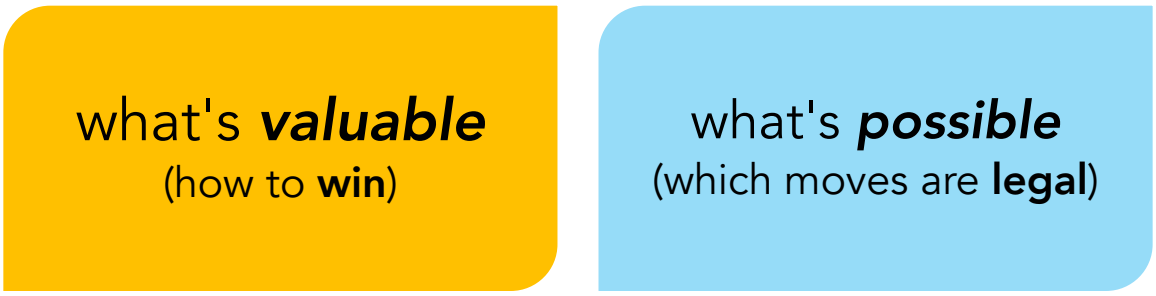


There's just one problem:

Real-life values aren't as simple as **board-game values**, because real life isn't as simple as board games.

After all, consider just how much ***control*** board game designers enjoy.

While creating new game worlds, they get to **stipulate** in advance:



In that context, players can try to **calculate**:

what's ***optimal***
(which moves maximize
winning chances)

But in the real world, we ***can't*** stipulate every possible move in advance. And we ***shouldn't*** assume there's just *one* kind of value we're optimizing for.

So if we rush ahead to calculate what's "ethically optimal," we've already confused our simplified model of the world for the world itself.

And that's the heart of ***Board Game Ethics:***

Board Game Ethics
treats value as a single **quantity** to be **maximized**

(like your *winning chances* in a board game)

On the other hand,

Real-Life Ethics
treats values as different kinds of **qualities** that **coexist**

(like the *many virtues* of a great jazz solo)

Board Game Ethics is based on several subtle confusions about what our **real-life ethical *values* (plural!)** are in the first place. But to fully appreciate what this means, let's keep thinking a bit more about board games.

2. Win, Lose, or Draw

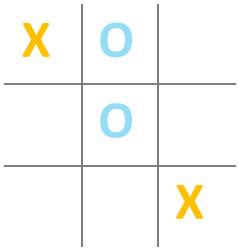
Tic-Tac-Toe is an interesting game, if you're a small child.

But by now, most games you play probably end in **a draw**.

But that's actually quite impressive, because there are **5,478 legal positions** in Tic-Tac-Toe.[2] You're very good at avoiding the ones that are bad for you.

How *do* you manage to do that, exactly?

X to move and avoid losing:

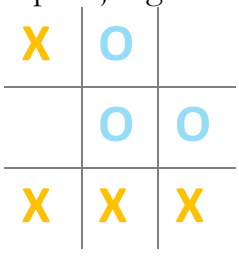


Wow, you did it again!

For starters, the goal in Tic-Tac-Toe is extremely simple: get **3 in a row**. We can state that goal clearly enough to teach it to young kids.

In Tic-Tac-Toe, **exactly *one* thing matters**: getting 3 in a row first.
And it's obvious when that's been achieved: you can just look.

We need an expert judge of Tic-Tac-Toe!



Can you tell us who won?

[2] https://web.archive.org/web/20130628112339/http://www.mathrec.org/old/2002jan/solutions.html.

But *who cares* about getting 3 in a row?

And why *do* we play games, anyway?

As Thi Nguyen notes, we play games for many **purposes**—

to have some ***fun***,
to get some ***exercise***,
to ***de-stress***,
to develop our ***skills***,
to ***vanquish*** our opponents,
to ***achieve*** some difficult task,
or even to ***experience*** the beauty of our own skilled action.[3]

Board games may also encourage us to try many things:

to ***role-play***,
to ***show off***,
to ***experiment*** with new strategies,
to invent ***house rules*** or ***variants***…

But ***mechanically speaking***, board games give players **one ultimate goal**:

to ***win***.

And that's it.

The hope is that *trying* to win at the game will help us achieve our purposes.

[3] Nguyen, *Games: Agency as Art*, 6.

Tic-Tac-Toe probably doesn't have enough going on to satisfy grown-ups. But if I try to win at Connect 4—a *real* board game—I'll probably have a better time, win more often, and maybe even improve my skill level.

The game tells us to care about winning. But since we aren't robots, we still care about our larger purposes, too!

Crucially, we bring many of our own values with us into games:

- Faced with multiple checkmating patterns, the grandmaster may prefer the most ***elegant*** or ***unique*** or ***devastating*** combination.

- Faced with the opportunity to launch a costly assault, the pacifist may prefer an ***alliance*** or ***compromise*** or ***concession*** instead.

- Faced with the chance to win in a truly ruthless manner, the good friend may prefer to ***preserve real-life relationships***.

In these cases, board games aren't telling players that these preferences are valuable; instead, players are **bringing their values** along with them.

But *mechanically* speaking, the outcome of a board game (win, lose, or draw) is **the only thing** that matters:

Board Game	How to Win
Connect 4	get *4 in a row*
Chess	*checkmate* your opponent
Risk	*capture* every territory
Twilight Imperium IV	score *10 Victory Points*

Even in "kingmaking" situations where one player has the power to pick who wins, everyone's still *very* much concerned with the game's **mechanical assessments** of who wins and who loses!

It's worth mentioning that not all games are like this.[4] In **role-playing games**, for example, the mechanics are specifically designed to let players build and explore worlds ***creatively*** and relatively ***open-endedly***.

Role-playing games are less concerned with winning; they're usually much more about ***individual expression*** and ***shared discovery***. Even when players *do* pursue long-term plans or goals, role-playing games don't try to impose a singular concern with winning. (That would interfere with the role-playing!)

But in board games, the mechanics *do* impose a fixed framework where players **compete to win**, either with each other or against the game itself.

What's **VALUABLE?**

Board games ask us to care about just ***one*** thing: **winning**.

This is a board game designer's **first superpower**: They get to stipulate what's valuable in advance, simply by telling us *what winning looks like*.

Players agree to prioritize these mechanical assessments of *Wins*, *Losses*, and (maybe) *Draws*. Those outcomes determine what's valuable within the game.

Every Connect 4 move can be evaluated based on how **useful** it is for helping you get 4 in a row before your opponent. That's what makes a *subtle threat* or a *timely block* **a good move**—it gets you closer to winning, or at least forcing a draw if you're already losing!

[4] Not even all board games, but it makes for a good title and gets folks thinking along the right lines.

And in Risk, all the *armies* and *cards* and *dice rolls* and *continent bonuses* are really just **useful** for helping you capture every territory. That's where all their **mechanical value** comes from, even if players do end up becoming emotionally invested in who controls the tactical backwater of Irkutsk.

Many folks who play a lot of board games like to think in terms of their ***win probability***, which is just what it sounds like: their chances of winning. Good moves increase or preserve win probability, whereas bad moves give it away.

The calculably ***Best*** move is whichever one
maximizes your **win probability**.

But compare that to more complex and sophisticated real-life practices, like gymnastics or jazz or coffeemaking. In these real-life practices, we care about *many kinds* of values. And these values often conflict with one another!

The gymnast sacrifices ***control***
to experience their full ***power*** on a vault.

The musician sacrifices ***clarity***
to experience their trombone's ***texture***.

The coffeemaker sacrifices ***consistency***
to experience ***unique*** fermentations.

In each of these cases, we try to balance genuinely *different* aspects of our experience. Both are valuable, but since they come into conflict, we can't maximize both at the same time.

You might think that means we're facing a *joint optimization problem*, where we should look for the **Pareto frontier** where we can jointly maximize our variables of interest as much as possible. While this way of thinking may be helpful for thinking about simple cases, I want to push you to reconsider how useful this framework is in real life.

In real life, we aren't just dealing with a few variables of interest at a time. We're dealing with *all* of our values! And that complicates things quite a bit:

1. How do we pick between multiple Pareto-optimal points in a principled way? In other words, **how do we manage tradeoffs?**

2. How do we measure each of our values? In other words, **how do we translate our values into metrics?**

3. How do we know how many values we even have? In other words, **how do we account for *all* our relevant values?**

Instead of thinking about conflicts between values as optimization problems, what if we viewed these conflicts as **genuine dilemmas?**

In genuine dilemmas, different *kinds* of values are at stake, and we can't have it all. I can't play trombone with full clarity *and* texture—I have to choose.

In these cases, we have to make **real sacrifices** and decide which truly **uncompensated losses** we're willing to suffer. Even if I get "paid back" for a loss in clarity with a gain in texture, I've sacrificed one value for another.

If only **one** kind of thing mattered,
we really *could* try to maximize it, just like **win probability**.

But even in board games, that's not really true. If you could maximize your win probability in Risk by giving me food poisoning, would you do it?

'Playing' Risk that way would undermine why we're playing together in the first place! Our larger purposes of *spending time together*, or *competing*, or *having fun* would be undermined by your single-mindedly trying to maximize your win probability at the expense of everything (and everyone!) else around you.

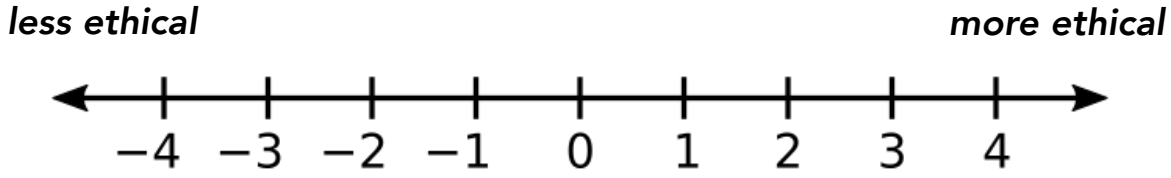
We're always managing all *kinds* of values, even while playing board games. **And if that's true in games, it's all the more true in ethics.**

In real life, we appeal to many different *kinds* of ethical value. We've already seen **security** and **privacy**, but consider ***liberty***, ***justice***, and ***beauty***!

Unless we assume these are all just **useful** for (or parts of) one overarching ***Supervalue***, there won't *be* a single dimension of ethical value.

But…what would it be like if there *were* just **one dimension** of ethical value?

The core hope of Board Game Ethics:[5]



Utilitarians think only **happiness** ultimately matters: more pleasure, less pain. That means all the other values are just *useful* for promoting **happiness**, just like all the armies in Risk are just *useful* for promoting **win probability**.

If that's right, we just have to come up with a scale to measure how much **happiness** each possible action produces, and we can always get a clear answer on what would be best: ***This*** **move maximizes happiness!**

(If calculating *what that move is* seems hard, don't worry! We might not be clever enough, but maybe AGI could run the numbers for us.)

Stepping back, these are very strong, very recent, very English, and yes, very imperialist assumptions for us to make about how ethics works. Utilitarians may still be right! Before we rush off to turn ethics into a **happiness** optimization problem, let's try doing a bit more philosophy first.

As we bring different kinds of ethical value into conversation with one another, we may find ourselves reassessing our commitments to them. For

[5] https://commons.wikimedia.org/wiki/File:Number-line-4.svg.

example, many folks have come to care less about ***chastity*** and more about ***individual expression*** in recent years. Maybe you agree, maybe not.

But if these really are **different kinds** of ethical values, we shouldn't just calculate out the ***Best*** move that optimizes for any one value like **happiness**.

If *many* kinds of things matter,
we shouldn't optimize for just *one!*

After all, optimizing for **happiness** will treat everything else (***Liberty****? **Justice****? **Beauty****?*) as just *useful* for **happiness**. And that means we may lose track of those other values altogether!

Imagine it turns out optimizing for **happiness** requires severely restricting ***liberty***—should we rush to do that, if that's how the math works out?

Are we confident enough in our current ethical judgments to make such enormous sacrifices? If not, where might we inject more ethical humility?

For starters, maybe we shouldn't assume that there's always a ***Best*** ethical course of action. Sometimes there might not be *any* good choices. Other times, there might be *lots* of great, creative solutions to ethical problems.

But the only way to find out is to bring our different values into conversation with one another long enough to see where the philosophical arguments lead. Once again, our ethical values come in tangles. So, shouldn't we examine how they're related and connected to one another before we rush to cut the philosophical knot with **an optimizing sword?**

Board games encourage us to think about outcomes as wins, losses, and draws. But maybe that's too simple and confrontational a framing to carry with us into ethics.

Board Game Ethics
stipulates that only one ***kind*** of thing matters.

(like your *win probability* in a board game)

Real-Life Ethics
recognizes that we seem to value different ***kinds*** of things that we shouldn't rush to "boil down" **prematurely**.

Even when competition judges score musicians numerically, they recognize that their scores **summarize but don't exhaust** how each musician plays.

Do we really want to assume there's **some perfect procedure** out there that could completely account for ***originality***, ***dexterity***, ***intensity***, ***musicality***, ***swing***, ***intonation***, and the countless other values involved in great music?

Do we really think we can boil everything down into **a single dimension** of value without any loss and mechanically rank who's ***Best*** at music?
Or which song is ***Optimal****?*
Or whose performance was closest to ***Ideal****?*

If we can, we should probably try to avoid building these controversial assumptions into how we think about ethics.

After all, our competitive procedures just have to be **good *enough*** for us to hold competitions and hand out prizes. But that doesn't mean fixed, agreed-upon procedures should get the final word on what's valuable!

And there's a lot more to playing music than winning competitions.
Like many of our real-world practices, music is much more ***open-ended***…

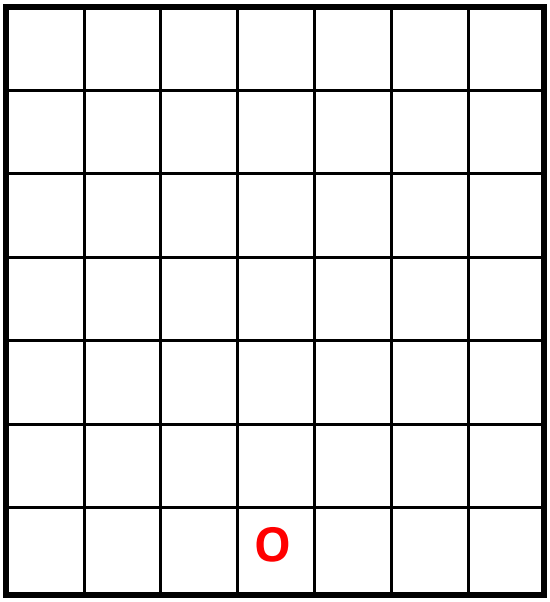
3. You Can't Do That!

Connect 4 is a ***solved game***.

That means we've found the **calculably *Best* move**, the move that maximizes win probability, in *every* position.

That's quite a feat considering that there are over 4.5 trillion legal positions—**4,531,985,219,092** to be exact![6]

So here's a tip: If you're going first, play in the middle.



Congratulations, you're playing perfectly so far!

If both sides keep making calculably ***Best*** moves, Player 1 always wins.

Chess is more complicated, so we haven't solved it yet. Most experts *think* chess is probably a draw with perfect play, but we still aren't completely sure.

After all, in Connect 4, you choose between up to **7** legal moves per turn.

---

[6] https://oeis.org/A212693.

But in chess, you average around **30**. Players often don't even see them all!

We still don't know how many legal positions there are in chess, but it's *about* **482,000,000,000,000,000,000,000,000,000,000,000,000,000**.[7]
(That's around 482 "tredecillion" if you're counting along at home.)

Great! Now here's a weird question:

How many valid **science experiments** are possible?

Huh. I guess…infinite?

In games, once we decide **how to count** (Are symmetrical Tic-Tac-Toe positions the same or not?) we can often **calculate *exactly*** how many legal positions there are under that definition (765 vs. 5,478).

But there aren't exactly ***legal moves*** in science, or at least, not in the same way. There's *better* and *worse* science, for sure. But scientific possibilities are much more **open-ended** than, say, Monopoly possibilities, where each turn you *have* to roll two dice, which *have* to add up to between 2 and 12.

We're running smack into board game designers' **second superpower**:

They don't just get to tell us what's **valuable**.
They get to tell us what's ***possible***.

What's **POSSIBLE?**

Board games tell us we can only try ***some*** things:
legal moves.

Importantly, what's possible in board games is much more **limited** than what's possible in real life.

[7] https://!github.com/tromp/ChessPositionRanking

Think about it…

In Connect 4, you *must* play **1 checker**
in one of exactly **7 columns**.

In Risk, attackers *may* roll up to **3 dice**,
each of which has exactly **6 sides**.

In chess, you *must* move **1 piece**
to one of exactly **64 squares**.

In real life, we can't describe every possibility in advance like this. *Who really knows* what's possible?! But in chess, players know how all the pieces work. You have to make exactly one move, and then it's my turn. Board games **limit** our possible actions quite a bit by telling us what we can and can't do.

Wanna move your rook diagonally, or skip this turn? **You can't do that!**

Of course, there are caveats. Here are just a few:

- Some board games have **dexterity** elements, where my physical ability to toss or balance a small block (for example) limits my ability to *execute* the ***Best*** move. Even so, the rules still constrain *how* I can toss or balance that block. For example, I probably don't get infinite do-overs!

- While I can calculate the **odds** of the next dice roll and maximize my win probability accordingly, there are always limits to my ability to predict what **other players** will do.

Social skills like psychologizing and politicking may seem 'softer' and more slippery than pure calculation. But we can still assess how ***effective*** players are at reading or persuading one another by observing how their choices lead to wins, losses, and draws.

And that assessment is assisted by the fact that board game mechanics are **much simpler to understand** than real life causation!

Why don't we have *board game scientists* puzzling out the Laws of Monopoly?

Obviously, because we have rulebooks. But more subtly, board game rules have to be **simple** and **clear** enough for board game designers to write down and board game players to learn and think about.

So, even complex board games are *way* ***simpler*** than real life!

Earlier, we saw that the board game designer's first superpower lets them stipulate that ***one*** thing is ultimately **valuable**: winning.

That means players don't usually face dilemmas in **translation**. As long as they pretend that only one thing matters (winning), they can try to pursue victory at the expense of everything else. In chess, we don't usually have to wonder *what we should be going for*—it's checkmate, or at least a draw if things are going badly! Players have a strong grip on what's **valuable**, and ***how things should be***. Their "objective function" is unusually well-defined.

Now add the board game designer's second superpower, which lets them stipulate that only ***some*** things are **possible**: legal moves.

That means players don't usually struggle to **describe** what's happening. As long as they follow the rules, they can describe the board state using its own mechanical laws of nature. (Scientists everywhere must be jealous!) In chess, we don't usually have to wonder *who goes first*, or *how that piece moves!* Players have a strong grip on what's **possible**, and ***how things are***. Their "search space" is unusually well-defined.

Since only winning is ***valuable***,
and only legal moves are ***possible***,
board games are much more ***calculable*** than real life.

Because board game mechanics are simple and clear, it makes sense to try to maximize your **win probability** by trying to calculate **the *Best* move**.

And this framework *can* be helpful in real life, when we're trying to solve local optimization problems. (How do I maximize my calories on a fixed budget, or increase my odds of getting promoted?)

But **"How do I win at life?"** is a *profoundly* bad question.
Take a moment to notice how *unlike* ethics this all is.

How many ways can I **apologize** to my friend?

Uh…infinitely?

What's the ***Best*** way to apologize to my friend?

Uh…sincerely?

Board games offer players incredible **clarity** about what's possible and what's valuable. And that can be a lot of fun! We can even *learn* a lot by playing board games.

But we shouldn't try to reduce reality *itself* to a board game.
Ethics is a very different kind of practice than maximizing win probability.

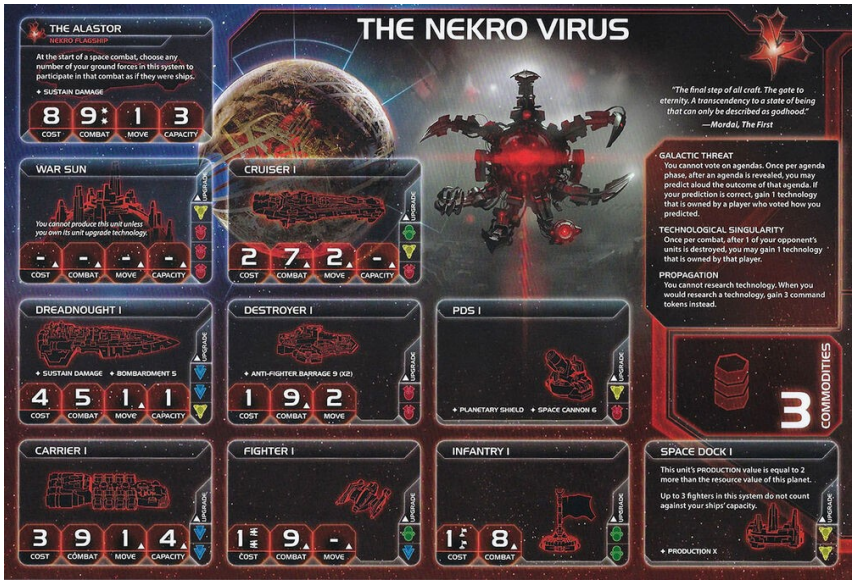
Sometimes board games really *do* confront us with interesting dilemmas. (Should I betray my friend to go for the win?) But even when they do, there's a calculably ***Best*** move, and surprise, surprise, it's the one that maximizes what the game considers valuable—your win probability!

After all, the game's mechanics only ask you to care about winning. And on those terms, "Should I betray my friend?" is just an empirical question about whether betrayal maximizes your win probability or not.

Sure, the procedure can always tell us what to care about.
But should we listen?

If we defer to procedures to resolve conflicts between our values for us, **we cede control or ownership** over what kinds of *people* we want to be, how we want to *live* together, and what our values even *are*. Instead, we let someone's fixed procedure choose for us.

Again, this can be *great* for board game night when we explore what it's like to reason from very different, even *alien* points of view.



Look out everyone, I'm playing as the Nekro Virus and my first faction ability is called…Galactic Threat? I'm sure that's just propaganda.[8]

But we shouldn't rush to farm out our real-life ethical values like this! Instead, we increase our ethical understanding by **doing** ethics together, by **evaluating** the different kinds of ethical values we are committed to, by **discovering** how they are both reliant on and in tension with one another.

You can't do **ethics** without doing **philosophy**!

Okay, fair enough. But if our real-life practices need to serve different *kinds* of values, how do we ever **compare apples and oranges?**

[8] https://twilight-imperium.fandom.com/wiki/The_Nekro_Virus?file=Nekro1.jpg

4. Role-Playing Game Ethics

Suppose the **unemployment rate** is low,
or **GPD growth** is high.
That seems like a *good* thing, right?

But why should we care about these numbers?

Numbers are useful for our purposes.

In a board game, the designer gets to stipulate what matters: Maybe the player who ends with the highest GPD wins. (And you care about winning, right?) As players, we agree to grow our GDPs, because we believe doing so will be **useful for our purposes** as players—to have fun, to compete, etc.

And in real life, metrics like *unemployment rate* and *GDP growth* are **useful for our purposes** of gauging the economy. Of course, these particular metrics fail to capture all kinds of things we care about, from ***quality of life*** to ***economic justice***. They're not perfect, but they're better than nothing!

Going forward, we should continue trying to develop and understand better economic metrics. But it's probably a fool's errand to try to develop the "perfect" economic metric, or even set of metrics.

Within a board game,
everything is ultimately **useful**
for **maximizing *one* quantity**: your *win probability*.

But in real life,
we only care about **quantities** (amounts) if they're **useful**
for tracking **qualities** (*kinds* of things) we care about.

In real life, we have to appeal to **qualities** to explain why **quantities** matter. And crucially, we try to make ***qualitatively* good arguments** to justify and explain the qualities we care about and the quantities we reach for.

Math is useful for telling us whether we've optimized for some **quantity**. But how can we tell whether the numbers are any good?

By seeing whether arguments for and against them are **qualitatively** good. And that means we need to be doing **philosophy**, not just math!

So, **ethics** *isn't* a **numbers game** all the way down!

But board games invite us into **a decision-making fantasy** where everything really *is* an optimization problem in disguise. We really *can* calculate how to ***efficiently*** utilize resources (like bishops or dreadnoughts) to bring about favorable mechanical outcomes (wins over draws over losses).

That isn't *bad!* But it's not so helpful for thinking about real-life ethics. We shouldn't assume that ethics is another optimization problem in disguise.

We turn to ethics to make **better sense of ourselves** and our **ethical experiences** together. But there's not some quantity or amount of *ethical sense-making* we're trying to maximize!

By the way, that's not how *scientific* sense-making works!

Good scientific explanations boast many positive qualities, including ***elegance***, ***coverage***, ***accuracy***, ***insight***, ***unity***, and so on. And as any scientist knows, these different kinds of qualities regularly conflict with one another.

For example, quantum physics is stunningly ***accurate*** at experimental prediction, but infamously bewildering to scientists like Einstein who demanded ***insight*** into why the predictions work and what they mean.

Sorry Einstein, our best available theories don't offer both *accuracy* and *insight!* You're facing a genuine ***scientific dilemma*** where scientific values are in conflict. There's no one thing to maximize, and you can't have it all.

So, you can't ***calculate*** how good a scientific explanation is!

Instead, we have to have (gulp) ***philosophical*** conversations about how different scientific values *relate* and *hang together*, involving questions like:

- How can we make better sense of our own *scientific values?*
- How can we make better sense of our bizarre *scientific experiences* at quantum scales, and why they're unlike our *everyday experiences?*
- How can we make better sense of *how* and *why* theories disagree?
- How could we produce *evidence* bearing on those disagreements?

These sorts of disputes are *always* ongoing within science. There isn't some quantity or amount of **scientific *sense-making*** we could try to maximize!

If even **science** involves many different *kinds* of values, how much truer must this be of **ethics?!**

Disputes are always ongoing within ethics, too. But just like in science, disagreement doesn't mean that ethics is *failing*.

Disagreement means ethics is happening.
That's just how ***inquiry*** works!

Disagreement drives us to ask *better questions*, try *new things*, and continue *making progress*. Disagreement drives us back into doing philosophy to try to make qualitatively better sense of ourselves and our experiences together.

So disagreement doesn't mean ethics is sputtering out; it's how ethics continues to *live* and *breathe* and *grow!* And examples of this are all around us.

"All men are created equal" is an incredibly powerful and revolutionary idea that doesn't exactly pair well with the coldblooded pragmatism of *"Black men count for 3/5ths, now let's get this convention over with."*

These two ideas conflict *deeply*. But which are we more committed to?

Over time, we've expanded our understanding of ***equality*** in ways that we come to understand as **less and less *self-deceptive:***

- First off, slavery's gotta go—that's profoundly unequal.
- We really do have to let women vote—they're adults, too.
- "Separate but Equal" was always bullshit—we have to desegregate.
- Civil unions aren't *actually* "close enough" to allowing gay marriage—that's just doing "Separate but Equal" again!

By increasing the ***coverage*** and ***unity*** of our ethical understanding of ***equality***, most reasonable folks agree we've made **real moral progress** on these issues in living memory. But no one thinks we're done! Our ethical understanding has *not* reached a final, flawless conclusion, and might never.

We keep finding that **we have more to learn**, especially as we continue living together in new contexts. New ethical experiences consistently exceed the categories and limitations imposed by our received ethical understandings. And so, we have to keep going.

So yeah, let's make up a new word like ***genocide***, or ***sexual harassment***, or ***microaggression***, and start trying to use it! Maybe these new words can help us get at something *real* that we didn't notice before we had the words for it.

Or maybe not—these new terms might not all work out. Maybe some of them will go the way of ***chastity*** or ***"Separate But Equal."*** But we should at least do enough philosophy together to try to find out!

This **open-endedness** of our ethical practices is *key*.

In board games, designers get to stipulate *in advance* what's valuable and what's possible, which makes the "right" answers artificially calculable.

But ethics is much more open-ended than board games. If "How do I win at life?" was a bad question, **"How do I win at ethics?"** is even worse!

Ethics doesn't *obviously* have a singular goal of optimizing anything. Don't try to calculate your way through every problem before you get to know it!

Instead, we should try to think of our values as ***sense-making resources***. Our values are **incomplete products**, yes, but they're **tools** developed over long histories of struggling to making sense of and deal with the ethical dilemmas and experiences that folks have encountered in real life.

Here's the thing: We're not very good at ethics yet.

Human history contains a ton of oppression, injustice, willful ignorance, and more, and we aren't *past* those problems. So, we should worry that our current values aren't *final* enough for us to set about optimizing anything!

Even so, our values give us shared ground to begin philosophizing together, to begin seeking greater ***empathy***, ***elegance***, ***coverage***, ***accuracy***, ***insight***, ***unity***, and all the other qualities we value in ethical understanding.

Our values let us tell ***non-deceptively*** **convincing stories**
that help us make sense of ourselves
and our ethical experiences living together.

By appealing to the kinds of things people *claim* to care about (like ***equality***) we can at least start holding them to their own standards, and figuring out what those standards really look like once we start thinking them through.

So the role of real-life values *isn't* to tell us what's valuable in advance.
It's to get new ethical conversations and experiments in living started!

Everyone agrees we've made at least *some* ethical progress by doing this. We've gotten ***better*** at ethics, but 'better' involves many open-ended goods, which we can't exhaustively spell out in advance. And *that* means ethical inquiry looks more like **a role-playing game** than a board game.

Instead of *board game ethics*, a better notion would be…
role-playing ethics.

After all, role-playing games give us a chance to ***practice*** **living together in different ways** and see what we can take away from these experiments!

You can *undeniably* role-play better and worse. But great role-playing, like most other real-world practices, involves tons of different values—***individual expression***, ***shared discovery***, ***commitment***, ***imaginativeness***, ***thoughtfulness***, countless ***social and emotional skills***, and all the other **qualities** that make a great role-player or gamemaster.

Come to think of it, those are all qualities we'd want in ethical inquiry, too!

Even if we abandon the idea of a core ethical value to optimize, that doesn't mean *anything goes!* We can see this by taking a page from great role-players and gamemasters, who are engaged in **ongoing, shared experiments** into what it would be like to live together in *this* or *that* particular way.

Board game ethicists could learn a thing or two from role-playing games, where we get together to **create** and **explore** different kinds of roles we might play in each other's lives. Ethical inquiry requires a more **open-ended** approach that looks beyond the winners and losers of fixed procedures.

Instead of rushing to **calculate** our way through **ethics**,
maybe we need to **philosophize** together first,
and practice taking on different **perspectives**.

We could call that ***role-playing game ethics***.