



Aaron Brown

# Kelly Myths and Heroes

A central concept in risk management, applying the Kelly criterion is in fact more of an art than a science.

**T**he Kelly criterion gives simple—remarkably simple—advice for dealing with any kind of uncertainty. For example, if you are offered an even-money bet with a 60 percent chance of winning, Kelly says to bet 20 percent of your wealth. It doesn't ask about your preferences or risk aversion or future betting opportunities. Therefore, it can't possibly be right all the time for everyone. For example, if your current wealth level provides everything you want, that is, an extra 20 percent wouldn't make you any better off, but any lower level would cause hardship, then you would be foolish to bet anything at all.

Kelly makes assumptions, which you have to be aware of before you apply it to practical situations. The two most important are:

- you know the possible outcomes of any actions you take, as well as their probabilities, and
- there is something (called “wealth” in the usual Kelly formulation but it could be anything) that is both your goal and a constraint on your actions.

When using Kelly for real decisions, these have to be relaxed.

## Why it works

Before discussing how to do this I want to explain why the criterion works and dispose of a few persistent Kelly myths. Your expected profit on an even-money bet with a 60 percent chance of winning is 20 percent of the amount you bet, so obviously the more you bet, the higher your expected profit. However, with average luck your profit is less than 20 percent of the amount you bet. If your wealth is  $w$  and you bet  $b$ , your median profit is:

$$(w + b)^{0.6} (w - b)^{0.4} - w < b \left( 0.2 - 0.48 \frac{b}{w} \right)$$

where the inequality comes from a Taylor expansion.

The difference between the expected profit ( $0.2b$ ) and the median profit is sometimes called volatility drag. As you raise the bet the expected profit goes up but if you increase beyond 20 percent of wealth your median outcome goes down (if you just look at the first two terms of the Taylor expansion above you'd think the maximum point was 20.83 percent, but it's just an approximation). The Kelly argument is that in the long run you get average luck, so it makes sense to maximize the median outcome of all your bets. Betting more than Kelly when you have an edge does result in higher



expected profit, but in the long run all the advantage comes from microscopic probabilities of astronomical (and likely unrealistic) levels of wealth, like one chance in  $10^{100}$  of having more wealth than exists in the world.

To maximize median outcome, we take the derivative of the quantity above:

$$\frac{d \left[ (w + b)^{0.6} (w - b)^{0.4} - w \right]}{db} = w \left[ 0.6 \left( \frac{1 - \frac{w}{b}}{1 + \frac{w}{b}} \right)^{0.4} - 0.4 \left( \frac{1 + \frac{w}{b}}{1 - \frac{w}{b}} \right)^{0.6} \right]$$

And set it to zero, which implies:

$$0.6 \left( 1 - \frac{w}{b} \right)^{0.4} \left( 1 - \frac{w}{b} \right)^{0.6} = 0.4 \left( 1 + \frac{w}{b} \right)^{0.6} \left( 1 + \frac{w}{b} \right)^{0.4}$$

$$0.6 \left( 1 - \frac{w}{b} \right) = 0.4 \left( 1 + \frac{w}{b} \right)$$

$$0.2 = \frac{w}{b}$$

## Kelly myth 1: you need to repeat a bet a large number of times for Kelly to work

One of the great virtues of Kelly is it can be applied on individual bets without factoring in future opportunities. It doesn't matter if all bets are different. The only requirement is that each bet is small in relation to the total uncertainty in your life. So if the stakes of one risky opportunity are high enough, or if you think it's the last

decision with uncertain outcomes you're ever going to make, the Kelly criterion may not be the best way to choose. But for most decisions under uncertainty, maximizing median outcome is sensible. Taking more or less risk consistently is almost certain to lead to a worse overall outcome.

### Kelly myth 2: the advantage of Kelly is that you never go broke

It baffles me why so many people seem to believe this. This is not a feature of Kelly; it's a feature of any criterion that never bets more than total wealth (including never betting at all, or betting any constant fraction of wealth). Moreover, most practical applications of Kelly discard this feature, so there is a non-zero probability of losing everything. Since no real-world strategy can guarantee positive wealth in all possible outcomes, so there's no reason to insist on it in theory.

### Kelly myth 3: the Kelly criterion is about bet sizing

This is a more subtle error, and an easy one to make given how I justified the criterion. But let's suppose we come across someone with \$100 wealth who is betting \$50 on the even-money bet with 60 percent win probability. If she does 100 bets of 50 percent of wealth each time, with average luck she'll win 60 and end up with  $\$100 \times 1.5^{60} \times 0.5^{40} = \$3.34$ , while the Kelly bettor betting 20 percent of wealth has median outcome of  $\$100 \times 1.2^{60} \times 0.8^{40} = \$748.99$ . Betting 50 percent means she needs at least 64 wins to make any profit at all – that only happens 24 percent of the time. The 20 percent bettor only needs 56 wins for a profit – that happens 82 percent of the time. If the bets are repeated more than 100 times, the 50 percent strategy looks worse and worse, while the Kelly strategy looks better and better.

The 50 percent bettor has a higher expected value; however, \$1,378,061 versus \$5,050, but in both cases nearly all the expected value comes from some very low-probability outcomes. If we limit outcomes to a maximum of \$100,000, the expected outcome of the 50 percent bettor is only slightly greater than the expected outcome 20 percent bettor: \$4,323 versus \$4,290.

Suppose we go over these numbers with the \$50 bettor, but cannot persuade her to lower her bet. We can do almost as much good by increasing her capital. Tell her she has \$250 instead of \$100. Now her \$50 bet is betting the Kelly fraction 20 percent, and she thinks we're doing her a favor instead of cramping her style. We haven't changed her bet size, what we've changed is how much she will increase her bet size after wins, and decrease it after losses. With 50 percent bet sizes, her second bet will be \$75 if she wins the first bet and \$25 if she loses. With 20 percent bets on \$250 of capital (\$150 of which is fictitious) her second bet will be \$60 if she wins and \$40 if she loses.

Under this scheme, she has a positive probability of going broke. For example, if she loses her first two bets, she'll be down to \$10 of real capital. She'll want to bet 20 percent of \$160 (her total of real and fictitious capital) which is \$32. We'll just stop her at that point and let her terminate with \$10.

Table 1 compares the outcomes of the \$50 bettor pretending to have \$250 versus the Kelly bettor betting \$20 on \$100 of real capital. The \$50 bettor loses money much more often, 33 percent of the time versus 18 percent, and loses more on average conditional on losing. In some of those cases, she went broke early, and lost out on most of the opportunities to bet with a positive edge. But she has six times the probability of ending up with over \$100,000, significantly higher probability of

**Table 1: Comparing the outcomes of the \$50 bettor pretending to have \$250 versus the Kelly bettor betting \$20 on \$100 of real capital.**

Outcome range	\$50 bettor		\$20 bettor	
	Probability	Average Outcome	Probability	Average Outcome
Under \$100	33 percent	\$21	18 percent	\$54
\$100 – \$200	2 percent	\$147	6 percent	\$148
\$200 – \$500	7 percent	\$377	22 percent	\$359
\$500 – \$1,000	5 percent	\$732	8 percent	\$749
\$1,000 – \$10,000	36 percent	\$4,108	37 percent	\$3,307
\$10,000 – \$100,000	15 percent	\$29,965	9 percent	\$28,224
\$100,000 – \$1,000,000	2.4 percent	\$207,733	0.4 percent	\$233,717
Over \$1,000,000	0.06 percent	\$1,912,699	0.01 percent	\$1,588,058

over \$10,000, and significantly better outcome in the \$1,000 to \$10,000 range. So it is not a crazy strategy to prefer betting \$50 to \$20, as long as it is \$50 of a pretend \$250 (so 20 percent of wealth) rather than \$50 of a real \$100 (betting 50 percent of wealth). The size of the bet isn't the main issue, it's how fast the bettor increases it or decreases it.

Another way to demonstrate that Kelly is not about bet sizing is to consider a bettor who bets 50 percent initially and increases bets 50 percent after a *loss*, while decreasing them 50 percent after a *gain*. This is the same magnitude of bet sizing as the normal 50 percent betting strategy, but it doesn't lead to certain disaster. Instead, 76 percent of the time it turns \$100 into \$140 and the other 24 percent of the time it loses everything. In some circumstances, such as if \$140 is all the money available to win, this can be a sensible strategy.

### Kelly myth 4: Kelly is always the right strategy

Strategies for dealing with uncertainty depend on the situation. The example immediately above might be a good strategy for project management. It's often the case that projects have limited upside, they either succeed or fail, and that there's little value to salvaging budgeted resources. In that case it can make sense to fail fast if you don't succeed. If the project is ahead of schedule and under budget, proceed cautiously and do extra testing; if the project has problems, gamble on new ideas and cut back on testing.

Betting more than Kelly can make sense if only extreme success is worthwhile, and the cost of betting is small. If you try to become a Twitter celebrity, for example, it's no use to have 100 or 1,000 followers. To begin monetizing your status you'll need at least 10,000 – and those better be high value followers who are either desirable to specialized advertisers or good retweeters. One million followers puts you in the big leagues. On the other hand, it costs almost nothing to tweet. So a reasonable strategy would be to tweet for a bit experimenting with different types, and if anything seems to be catching on, to ramp up your effort very quickly. Sure this means you have a low probability of success, but you knew that anyway. Rapid ramping up and down gives you the best chance of really big success. If you don't value moderate success, it can be a good strategy.

As mentioned above, the key is not how big you bet, it's how you change bets after success and failures. Ramping up exposure quickly after gains, and cutting quickly after losses, takes maximum advantage of runs of success and can survive runs of failure. But it pays a high cost in volatility drag. If you increase exposure  $X$  percent after success and cut it  $X$  percent after failure, you pay  $X$  percent<sup>2</sup> if you pair a success and a failure (e.g., you start with \$100 and a \$20 bet, win and increase the bet 20 percent to \$24, and lose, you have \$96, a loss of 4 percent, 20 percent<sup>2</sup>; if you lose the first bet you have \$80, you cut your bet to \$16, so a win brings you up to the same \$96). If you are slower to increase exposures after wins, you cut your chances of really big wins. If you are slower to decrease exposures after losses, you can be hurt a lot by a period of bad luck. But either of those policies will reduce your volatility drag. If you go as far as reducing exposure after successes and increasing exposure after failure, you earn volatility drag, but you cap your upside while amplifying your downside from extended runs. Depending on the situation, any of these strategies can make sense.

### What if you don't know the probability distribution of outcomes?

In most practical decisions under uncertainty, you don't know the exact probabilities and outcomes. You may have no better than rough approximations. A simple mathematical example can illustrate a good technique for dealing with that. Let's continue with the example of the even-money bet, but drop the assumption that you know for sure that your probability of winning each bet is 60 percent.

The conjugate prior for the binomial distribution is the Beta distribution. That just means if our subjective probability distribution for the probability of winning each bet follows the Beta form, it simplifies the mathematics. Since we're interested in distilling principles for dealing with real uncertainty rather than an exact mathematical solution to a textbook problem, there's no point in getting complicated. I'm not even going to discuss what a Beta distribution is, I'm just going to list its nice properties for this problem:

- A Beta distribution has two parameters, which I'll call  $A$  and  $B$  (there are actually two popular ways to parametrize the Beta, I'm using the less common of those two);
- The expected value of a draw from a Beta distribution is  $A / (A + B)$ ;
- All draws from a Beta distribution are in the interval  $(0,1)$ ;
- If my subjective prior probability of winning a bet follows a Beta distribution with parameters  $A$  and  $B$ , and then I observe  $w$  win and  $l$  losses, my posterior distribution of the probability of winning the bet is Beta with parameters  $A + w$  and  $B + l$ ;
- The Kelly bet for a prior Beta  $A, B$  distribution is the same as the Kelly bet if I know for sure that the probability of winning is  $A / (A + B)$ ;
- An easy way to remember this is to pretend that before beginning to bet, I observed  $A$  wins and  $B$  losses; at all future times I estimate my probability of winning as the observed win frequency, counting all the bets I've seen, plus my initial  $A + B$  fictitious bets; then I make the Kelly bet using that probability estimate.

If I do this, starting with parameters  $A$  and  $B$ , then win  $w$  and lose  $l$  bets, my initial wealth will have multiplied by the factor:

$$2^{l+w} \frac{(A + l - 1)! (B + w - 1)!}{(A - 1)! (B - 1)!} \frac{(A + B - 1)!}{(A + B + l + w - 1)!}$$

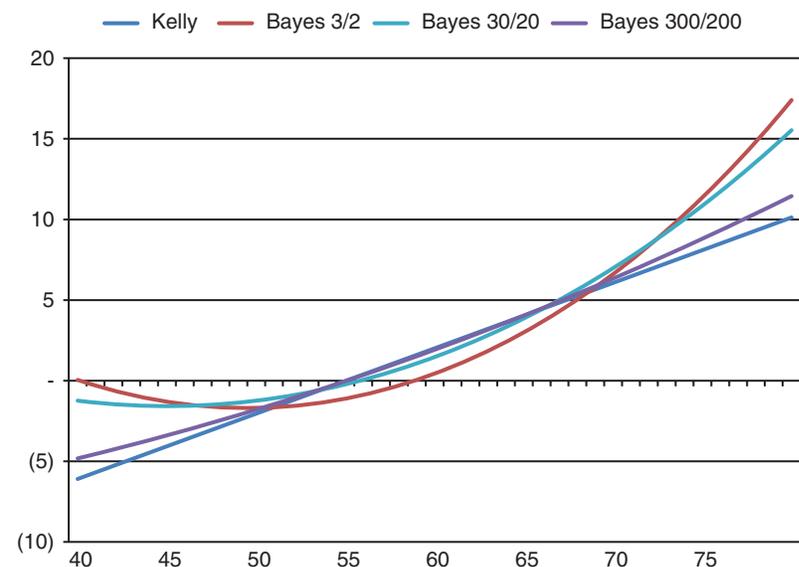
That's a little messier than the formula with fixed probability of winning, but it's not too bad. If you want to evaluate it in a computer program, you're better off using the Stirling approximation to the logarithm of the number, which is even messier but easier to evaluate:

$$\begin{aligned} & (l + w) \ln(2) + (B + w - 1) \ln(B + w - 1) + (A + l - 1) \ln(A + l - 1) \\ & + (B + A - 1) \ln(B + A - 1) - (B - 1) \ln(B - 1) - (A - 1) \ln(A - 1) \\ & - (B + A + l + w - 1) \ln(B + A + l + w - 1) \\ & + 0.5 \ln \left[ \frac{\left(2(B + w - 1) - \frac{1}{3}\right) \left(2(A + l - 1) - \frac{1}{3}\right)}{\left(2(B - 1) - \frac{1}{3}\right) \left(2(A - 1) - \frac{1}{3}\right) \left(2(B + A + l + w - 1) - \frac{1}{3}\right)} \right] \end{aligned}$$

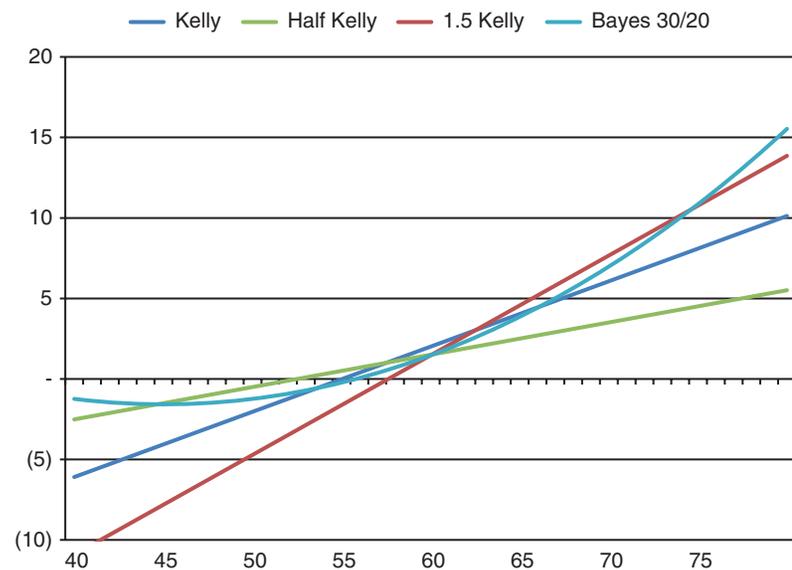
Figure 1 (below) shows the outcome of 100 even-money bets, with the logarithm of wealth ratio on the vertical axis, and the number of wins on the horizontal axis. "Kelly" shows the results if we make the Kelly bet assuming the win probability is 60 percent. Bayes 3/2 means we start with a Beta prior with parameters 3,2. Since  $3 / (3 + 2) = 60$  percent, we start with the same 20 percent bet as when we assumed the win probability was 60 percent for sure. But if we win the first bet, our estimated win probability goes up to  $(3 + 1) / (3 + 2 + 1) = 66.67$  percent, so we bet 33.33 percent of our wealth for the second bet. If we lose the first bet, our estimated win probability declines to  $3 / (3 + 2 + 1) = 50$  percent, so we bet zero.

All the Bayesian versions do worse than Kelly for the most common outcomes (assuming the true win probability is 60 percent) but curve to do better for less common outcomes in either direction. Bayes 300/200 is very close to the Kelly line, because our estimated probability of winning can't vary much from 60 percent. Bayes 30/20 is farther from the Kelly straight line, Bayes 3/2 is farther still.

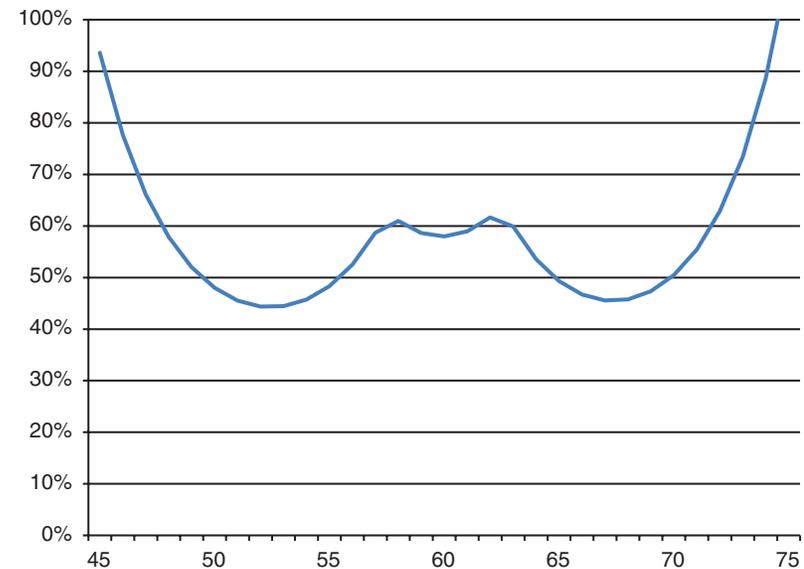
Figure 1: The outcome of 100 even-money bets.



**Figure 2: The Bayes 30/20 outcomes along with full Kelly (bet 20 percent of wealth each time), half Kelly (bet 10 percent of wealth each time), and 1.5 Kelly (bet 30 percent of wealth each time).**



**Figure 3: Ratio of terminal wealth of Bayes Kelly versus the maximum of the three Kelly strategies at each outcome**



## It's easy to see how Bayes Kelly might be adapted for financial trading strategies, but the problem with applying it in general is that the outcomes of prior risks don't necessarily give much information about the probability distribution of future risks

It's important to keep in mind that the horizontal axis is the actual number of wins, not the expected number. So even if we are correct that the true win probability is 60 percent, we might do better using a Bayesian Kelly rule. The way I like to think about the choice is illustrated with Figure 2 (below) which shows the Bayes 30/20 outcomes along with full Kelly (bet 20 percent of wealth each time), half Kelly (bet 10 percent of wealth each time) and 1.5 Kelly (bet 30 percent of wealth each time). For all the likely outcomes (44 to 76 wins), Bayes Kelly does worse than the best of the three Kelly strategies, but it always does better than the worst.

Figure 3 shows the ratio of terminal wealth of Bayes Kelly versus the maximum of the three Kelly strategies at each outcome. If I extended the x axis, the numbers would all be above 100 percent, but I don't take those seriously. On the left, it means a big ratio on a small base, and isn't really worth a lot. Moreover, it comes from reversing the strategy and taking the opposite side of the bet, which is not often possible, and even less often sensible. If you think you have a 60 percent win rate strategy and find you are wrong, it's hard to believe it's a good idea to immediately switch to the

opposite side. You should first figure out why you were wrong. On the right, the extra money is probably unrealistic as you will hit capacity or other issues.

Nevertheless, there is a lot to be said for Bayes Kelly. You can think of it as paying a tax, likely from 45 percent to 60 percent of your terminal wealth, in exchange for being put in the best Kelly strategy for win rates between 55 percent and 65 percent (bets between 10 percent and 30 percent of wealth). I think that's often good insurance, especially compared to more common alternatives like half Kelly.

### The worst case

If you choose to do things this way, your worst case is no longer losing all your wealth. The worst outcome for Bayes Kelly occurs if the outcome makes your optimal bet zero. For Bayes 30/20 that means getting ten more losses than wins, which will cost you about two thirds of the wealth you use for sizing your Kelly bets. Therefore, you could bet 50 percent higher than Kelly (pretending you had \$150 instead of

\$100 and sizing your bets to that) and still not risk losing more than your capital. In that case, you earn likely 70 percent to 90 percent as much as the maximum of the three fixed Kelly strategies.

It's easy to see how Bayes Kelly might be adapted for financial trading strategies, but the problem with applying it in general is that the outcomes of prior risks don't necessarily give much information about the probability distribution of future risks. For example, suppose you are thinking of writing a screenplay for a major motion picture. For your first bet, you decide to allocate three weekends to coming up with a good outline, after which, you'll try to get an agent interested in it. You might be able to guess some probability distribution of success for both of those steps, but it's not clear how information about how the outline turned out affects your estimate of the chances of getting an agent. That's only the Bayes issue. An objection to applying any flavor of Kelly is that you don't gain wealth by succeeding in individual steps. Finally, an objection to applying any sort of quantitative analysis is you don't know much about the probabilities and potential outcomes.

## The second you adopt a Bayesian approach, you lever your real capital, and you buy yourself more latitude in misestimating probabilities and outcomes

Despite these objections, I have come to believe that performing a Bayesian Kelly analysis is an excellent risk management approach to any risky endeavor. I don't believe that it optimizes risk taking, but it does make risk taking more rational and consistent, and most important, means you learn a lot more from failures. In the absence of a formal process, there is a strong possibility that your risk taking will be sabotaged by behavioral biases.

### Venture Bayes Kelly

How might this work for, say, starting a company? You know that the process will entail many risky decisions: how much money and effort to devote to the idea, when to quit your job to work on the company full time, whether to take on partners or seek outside financing, and many others, big and small. You can make only rough guesses about the probabilities and potential outcomes of each.

The first task is to come up with the appropriate definition of wealth to use for decision making. This is crucial because it determines how fast you will ramp up risk after success and how quickly you will cut back after failures. For this purpose,

I recommend focusing on the equity value of your company. Of course, that's extremely difficult to estimate, but coming up with a number will help you make consistent decisions. Any business success, building a working prototype, getting a meeting with a potential customer, getting a good article written, increases the value of the equity of the company; any failure decreases it. This allows you to put rough dollar estimates on outcomes with no direct dollar values. Since Kelly requires lots of risks for the criterion to be reliable, it's important to keep things in one dimension, even at the risk of oversimplification. An oversimple model that gives unambiguous and reasonable answers is better than a perfect model that does not.

The initial equity value of the company is the amount you are willing to lose, counting money, your own time and any resources contributed by others, if the attempt is a complete failure. You don't have to fund this in cash up front, but you should know what the figure is. If you are committed to the idea that value can be quite large even if you are broke at the moment. A large value means you take up risk slowly after success, and cut slowly after failure. If this is just a flyer that you are unwilling to lose to back with significant investment, it means you ramp up quickly after success, and cut back risk sharply after even moderate failure.

At this point, you could try to apply fixed Kelly. For each decision, you could make your best estimate of the likely impacts on the equity value of your company, and the probabilities attached to them. You would pick the decision that maximizes the expected value of the logarithm of equity. While there are obviously huge uncertainties in all those estimates, I believe the discipline of the calculation improves decision making. It has particular value for factoring in low-probability outcomes which your brain is programmed to either exaggerate or ignore, never to consider rationally.

If we want to add a Bayesian overlay to the Kelly process, we have to make an assumption that all the risks you take in the business are affected by some underlying factor or factors. You learn about these factors tracking the results of prior decisions, and they help you refine future decisions. One obvious major factor is whether your business is a good idea in the first place. If so, that should help you in many risky outcomes: projects are more likely to work and people are more likely to sign on as partners, customers, investors, and other capacities. In many cases you will be able to break that down into component factors like does the basic technology work, do people want the product, how good is the competition, or is the business legal?

If you adopt Bayesian principles, it makes sense to scale things so that your equity value goes to zero at the point in which you would give up on the idea anyway. You never want to walk away from an idea you still have value because you ran out of resources, but you also don't want to reduce your chances of success by not risking the full amount of resources you chose to allocate. The second you adopt a Bayesian approach, you lever your real capital, and you buy yourself more latitude in misestimating probabilities and outcomes. You increase both your chance of success, and the amount you learn from failure. In exchange, you pay a tax on any success relative to someone who knew going in everything you know after the fact. I think that's a good trade.

The Kelly criterion is a central principle of risk management. Over 55 years ago, Ed Thorp named it "Fortune's Formula." Simple mathematical examples are useful for understanding it, but applying it in practice is more art than science. Adding a Bayesian dimension expands its usefulness, at the cost of some extra complexity.