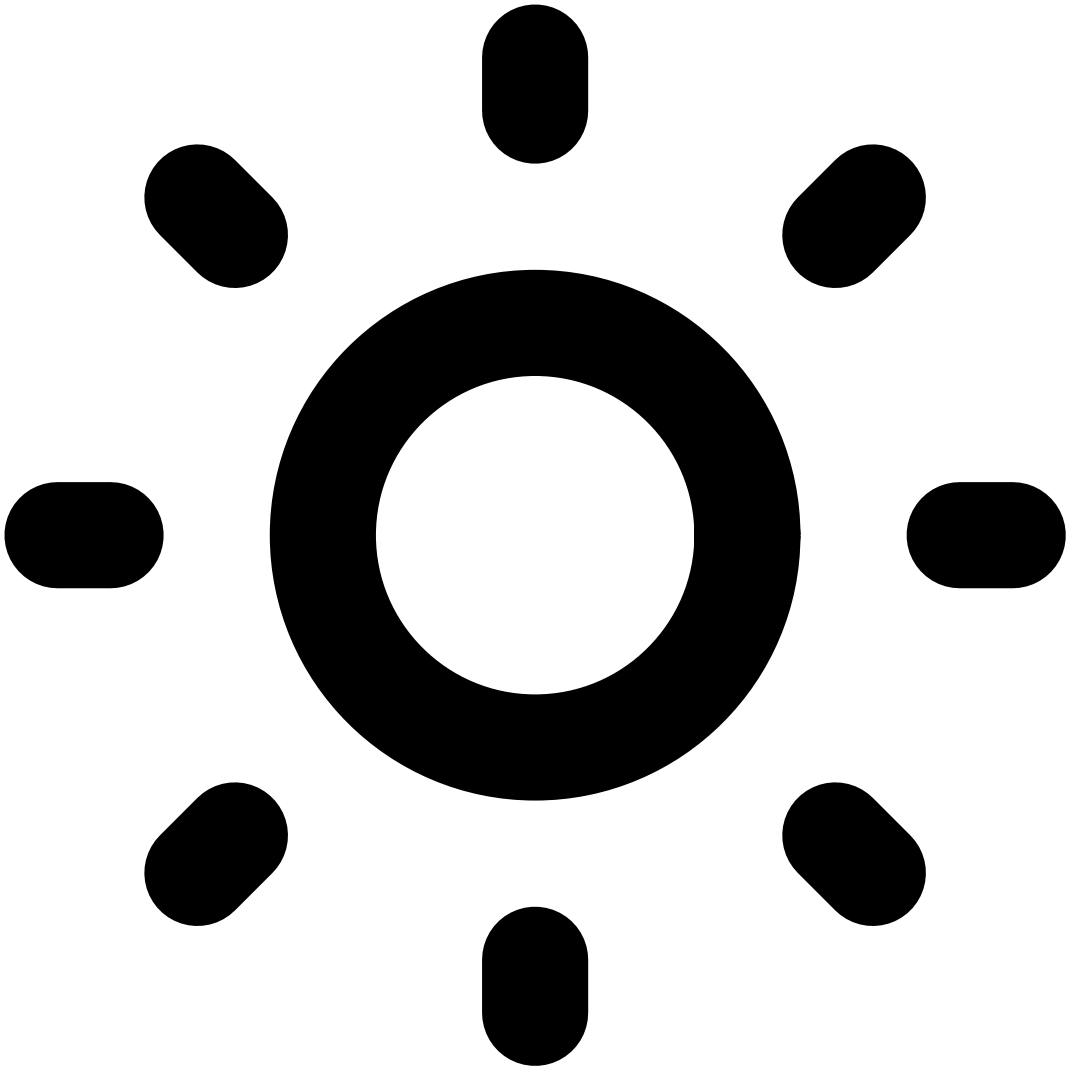
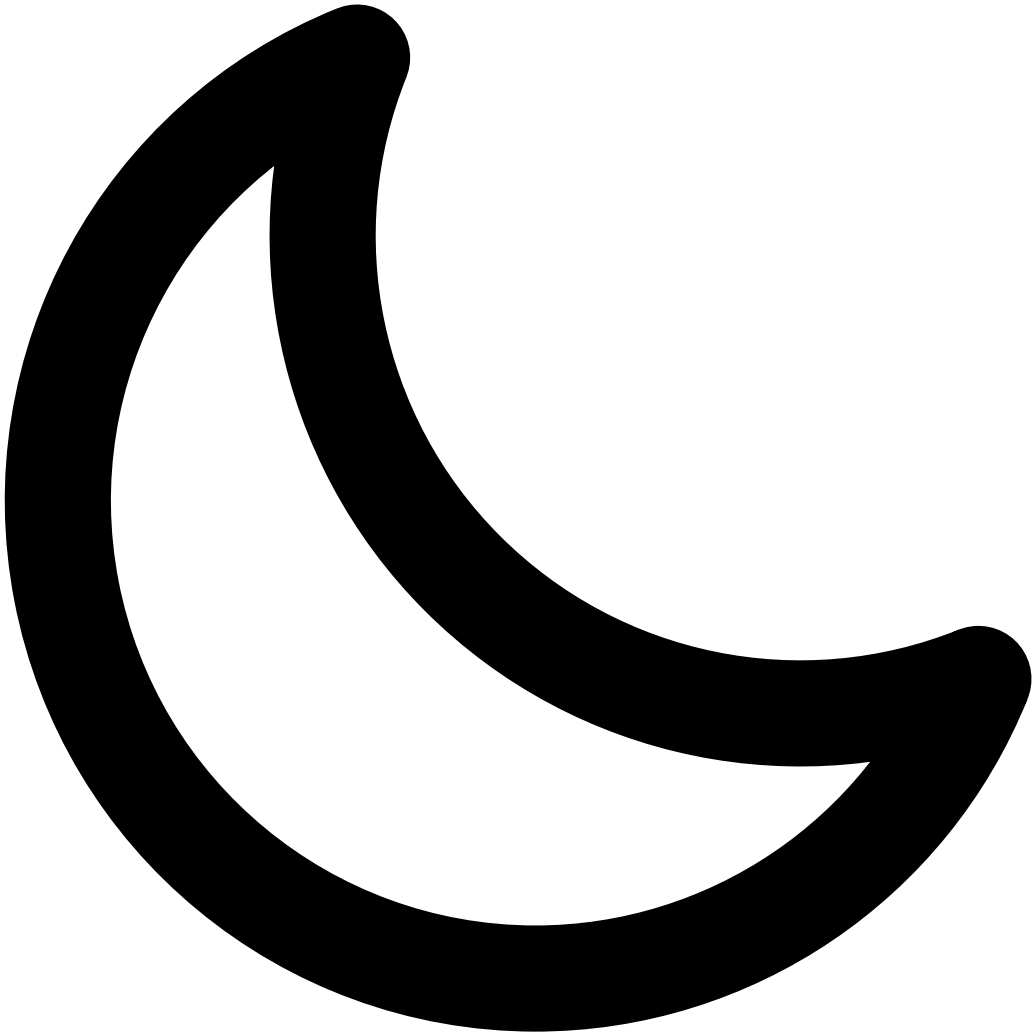


☐ Dark Mode Toggle



Analyzing Token Sale Models

2017 Jun 09

See all posts

Note: I mention the names of various projects below only to compare and contrast their token sale mechanisms; this should NOT be taken as an endorsement or criticism of any specific project as a whole. It's entirely possible for any given project to be total trash as a whole and yet still have an awesome token sale model.

The last few months have seen an increasing amount of innovation in token sale models. Two years ago, the space was simple: there were capped sales, which sold a fixed number of tokens at a fixed price and hence fixed valuation and would often quickly sell out, and there were uncapped sales, which sold as many tokens as people were willing to buy. Now, we have been seeing a surge of interest, both in terms of theoretical investigation and in many cases real-world implementation, of hybrid capped sales, reverse dutch auctions, Vickrey auctions, proportional refunds, and many other mechanisms.

Many of these mechanisms have arisen as responses to perceived failures in previous designs. Nearly every significant sale, including Brave's Basic Attention Tokens, Gnosis, upcoming sales such as Bancor, and older ones such as Maidsafe and even the Ethereum sale itself, has been met with a

substantial amount of criticism, all of which points to a simple fact: so far, we have still not yet discovered a mechanism that has all, or even most, of the properties that we would like.

Let us review a few examples.

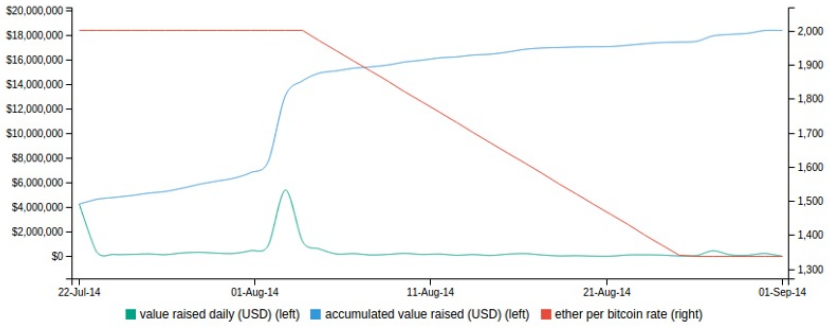
Maidsafe

The decentralized internet platform raised $7m in five hours. However, they made the mistake of accepting payment in two currencies (BTC and MSC), and giving a favorable rate to MSC buyers. This led to a temporary ~2x appreciation in the MSC price, as users rushed in to buy MSC to participate in the sale at the more favorable rate, but then the price saw a similarly steep drop after the sale ended. Many users converted their BTC to MSC to participate in the sale, but then the sale closed too quickly for them, leading to them being stuck with a ~30% loss.

This sale, and several others after it (cough cough WeTrust, TokenCard), showed a lesson that should hopefully by now be uncontroversial: running a sale that accepts multiple currencies at a fixed exchange rate is dangerous and bad. Don't do it.

Ethereum

The Ethereum sale was uncapped, and ran for 42 days. The sale price was 2000 ETH for 1 BTC for the first 14 days, and then started increasing linearly, finishing at 1337 ETH for 1 BTC.



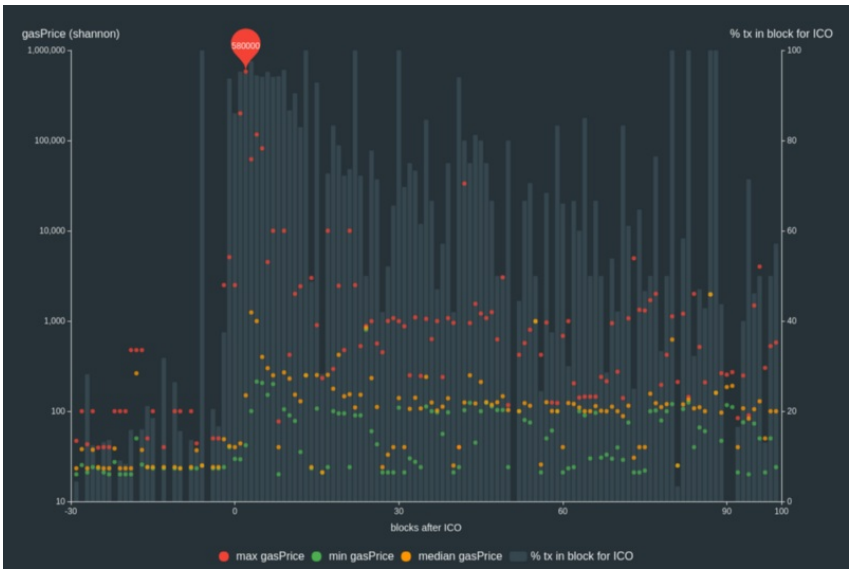
Nearly every uncapped sale is criticized for being "greedy" (a criticism I have significant reservations about, but we'll get back to this later), though there is also another more interesting criticism of these sales: they give participants *high uncertainty about the valuation* that they are buying at. To use a not-yet-started sale as a example, there are likely many people who would be willing to pay $10,000 for a pile of Bancor tokens if they knew for a fact that this pile represented 1% of all Bancor tokens in existence, but many of them would become quite apprehensive if they were buying a pile of, say, 5000 Bancor tokens, and they had no idea whether the total supply would be 50000, 500000 or 500 million.

In the Ethereum sale, buyers who really cared about predictability of valuation generally bought on the 14th day, reasoning that this was the last day of the full discount period and so on this day they had maximum predictability together with the full discount, but the pattern above is hardly

economically optimal behavior; the equilibrium would be something like everyone buying in on the last hour of the 14th day, making a private tradeoff between certainty of valuation and taking the 1.5% hit (or, if certainty was really important, purchases could spill over into the 15th, 16th and later days). Hence, the model certainly has some rather weird economic properties that we would really like to avoid if there is a convenient way to do so.

BAT

Throughout 2016 and early 2017, the capped sale design was most popular. Capped sales have the property that it is very likely that interest is oversubscribed, and so there is a large incentive to getting in first. Initially, sales took a few hours to finish. However, soon the speed began to accelerate. First Blood made a lot of news by finishing their $5.5m sale in two minutes - *while active denial-of-service attacks on the Ethereum blockchain were taking place*. However, the apotheosis of this race-to-the-Nash-equilibrium did not come until the BAT sale last month, when a $35m sale was completed within 30 seconds due to the large amount of interest in the project.



Not only did the sale finish within two blocks, but also:

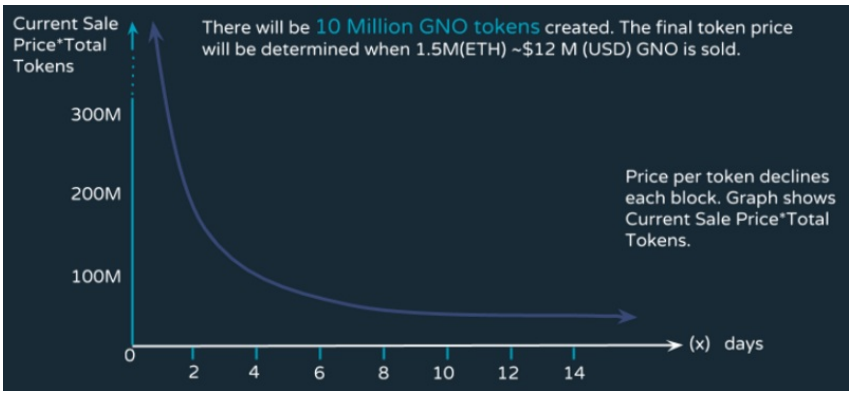
- The total transaction fees paid were 70.15 ETH (>$15,000), with the highest single fee being ~$6,600
- 185 purchases were successful, and over 10,000 failed
- The Ethereum blockchain's capacity was full for 3 hours after the sale started

Thus, we are starting to see capped sales approach their natural equilibrium: people trying to outbid each other's transaction fees, to the point where potentially millions of dollars of surplus would be burned into the hands of miners. And that's before the next stage starts: large mining pools butting into the start of the line and just buying up all of the tokens themselves before anyone else can.

Gnosis

The Gnosis sale attempted to alleviate these issues with a novel mechanism: the reverse dutch auction. The terms, in simplified form, are as follows. There was a capped sale, with a cap of $12.5 million USD. However, the portion of tokens that would actually be given to purchasers depended on how long the sale took to finish. If it finished on the first day, then only ~5% of tokens would be distributed among purchasers, and the rest held by the Gnosis team; if it finished on the second day, it would be ~10%, and so forth.

The purpose of this is to create a scheme where, if you buy at time T, then you are guaranteed to buy in at a valuation which is at most $\frac{1}{T}$.



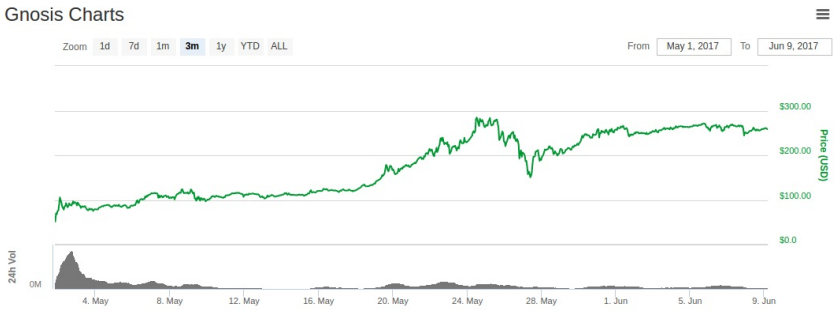
The goal is to create a mechanism where the optimal strategy is simple. First, you personally decide what is the highest valuation you would be willing to buy at (call it V). Then, when the sale starts, you don't buy in immediately; rather, you wait until the valuation drops to below that level, and then send your transaction.

There are two possible outcomes:

1. The sale closes before the valuation drops to below V. Then, you are happy because you stayed out of what you thought is a bad deal.
2. The sale closes after the valuation drops to below V. Then, you sent your transaction, and you are happy because you got into what you thought is a good deal.

However, many people predicted that because of "fear of missing out" (FOMO), many people would just "irrationally" buy in at the first day, without even looking at the valuation. And this is exactly what happened: the sale finished in a few hours, with the result that the sale reached its cap of $12.5 million when it was only selling about 5% of all tokens that would be in existence - an implied valuation of [over $300 million].

All of this would of course be an excellent piece of confirming evidence for the narrative that markets are totally irrational, people don't think clearly before throwing in large quantities of money (and often, as a subtext, that the entire space needs to be somehow suppressed to prevent further exuberance) if it weren't for one inconvenient fact: ***the traders who bought into the sale were right***.



Even in ETH terms, despite the massive ETH price rise, the price of 1 GNO has increased from ~0.6 ETH to ~0.8 ETH.

What happened? A couple of weeks before the sale started, facing public criticism that if they end up holding the majority of the coins they would act like a central bank with the ability to heavily manipulate GNO prices, the Gnosis team agreed to hold 90% of the coins that were not sold for a year. From a trader's point of view, coins that are locked up for a long time are coins that cannot affect the market, and so in a short term analysis, might as well not exist. This is what initially propped up Steem to such a high valuation [last year in July], as well as Zcash in the very early moments when the price of each coin [was over $1,000].

Now, one year is not *that* long a time, and locking up coins for a year is nowhere close to the same thing as locking them up forever. However, the reasoning goes further. Even after the one year holding period expires, you can argue that it is in the Gnosis team's interest to only release the

locked coins if they believe that doing so will make the price go up, and so if you trust the Gnosis team's judgement this means that they are going to do something *which is at least as good for the GNO price as simply locking up the coins forever*. Hence, in reality, the GNO sale was really much more like a capped sale with a cap of $12.5 million and a valuation of $37.5 million. And the traders who participated in the sale reacted exactly as they should have, leaving scores of internet commentators wondering what just happened.

There is certainly a weird bubbliness about crypto-assets, with various no-name assets attaining market caps of $1-100 million (including BitBean as of the time of this writing at $12m, PotCoin at $22m, PepeCash at $13m and SmileyCoin at $14.7m) just because. However, there's a strong case to be made that the participants *at the sale stage* are in many cases doing nothing wrong, at least for themselves; rather, traders who buy in sales are simply (correctly) predicting the existence of an ongoing bubble has been brewing since the start of 2015 (and arguably, since the start of 2010).

More importantly though, bubble behavior aside, there is another legitimate criticism of the Gnosis sale: despite their 1-year no-selling promise, eventually they will have access to the entirety of their coins, and they **will** to a limited extent be able to act like a central bank with the ability to heavily manipulate GNO prices, and traders will have to deal with all of the monetary policy uncertainty that that entails.

Specifying the problem

So what would a **good** token sale mechanism look like? One way that we can start off is by looking through the criticisms of existing sale models that we have seen and coming up with a list of desired properties.

Let's do that. Some natural properties include:

1. **Certainty of valuation** - if you participate in a sale, you should have certainty over at least a ceiling on the valuation (or, in other words, a floor on the percentage of all tokens you are getting).
2. **Certainty of participation** - if you try to participate in a sale, you should be able to generally count on succeeding.
3. **Capping the amount raised** - to avoid being perceived as greedy (or possibly to mitigate risk of regulatory attention), the sale should have a limit on the amount of money it is collecting.
4. **No central banking** - the token sale issuer should not be able to end up with an unexpectedly very large percentage of the tokens that would give them control over the market.
5. **Efficiency** - the sale should not lead to substantial economic inefficiencies or deadweight losses.

Sounds reasonable?

Well, here's the not-so-fun part.

- (1) and (2) cannot be fully satisfied simultaneously.
- At least without resorting to very clever tricks, (3), (4) and (5) cannot be satisfied simultaneously.

These can be cited as "the first token sale dilemma" and "the second token sale trilemma".

The proof for the first dilemma is easy: suppose you have a sale where you provide users with certainty of a $100 million valuation. Now, suppose that users try to throw $101 million into the sale. At least some will fail. The proof for the second trilemma is a simple supply-and-demand argument. If you satisfy (4), then you are selling all, or some fixed large percentage, of the tokens, and so the valuation you are selling at is proportional to the price you are selling at. If you satisfy (3), then you are putting a cap on the price. However, this implies the possibility that the equilibrium price at the quantity you are selling exceeds the price cap that you set, and so you get a shortage, which inevitably leads to either (i) the digital equivalent of standing in line for 4 hours at a very popular restaurant, or (ii) the digital equivalent of ticket scalping - both large deadwight losses, contradicting (5).

The first dilemma cannot be overcome; some valuation uncertainty or participation uncertainty is inescapable, though when the choice exists it seems better to try to choose participation uncertainty rather than valuation uncertainty. The closest that we can come is compromising on *full participation* to *guarantee* partial participation. This can be done with a proportional refund (eg. if $101 million buy in at a $100 million valuation, then everyone gets a 1% refund). We can also think of this mechanism as being an uncapped sale where part of the payment comes in the form of *locking up* capital rather than spending it; from this viewpoint, however, it becomes clear that the requirement to lock up capital is an efficiency loss, and so such a mechanism fails to satisfy (5). If ether holdings

are not well-distributed then it arguably harms fairness by favoring wealthy stakeholders.

The second dilemma is difficult to overcome, and many attempts to overcome it can easily fail or backfire. For example, the Bancor sale is considering limiting the transaction gas price for purchases to 50 shannon (~12x the normal gasprice). However, this now means that the optimal strategy for a buyer is to set up a large number of accounts, and from each of those accounts send a transaction that triggers a contract, which then attempts to buy in (the indirection is there to make it impossible for the buyer to accidentally buy in more than they wanted, and to reduce capital requirements). The more accounts a buyer sets up, the more likely they are to get in. Hence, in equilibrium, this could lead to *even more* clogging of the Ethereum blockchain than a BAT-style sale, where at least the $6600 fees were spent on a single transaction and not an entire denial-of-service attack on the network. Furthermore, any kind of on-chain transaction spam contest severely harms fairness, because the cost of participating in the contest is constant, whereas the reward is proportional to how much money you have, and so the result disproportionately favors wealthy stakeholders.

Moving forward

There are three more clever things that you can do. First, you can do a reverse dutch auction just like Gnosis, but with one change: instead of holding the unsold tokens, put them toward some kind of public good. Simple examples include: (i) airdrop (ie. redistributing to all ETH holders), (ii) donating to the Ethereum Foundation, (iii) donating to Parity, Brainbot, Smartpool or other companies and individuals independently building infrastructure for the Ethereum space, or (iv) some combination of all three, possibly with the ratios somehow being voted on by the token buyers.

Second, you can keep the unsold tokens, but solve the "central banking" problem by committing to a fully automated plan for how they would be spent. The reasoning here is similar to that for why many economists are interested in rules-based monetary policy: even if a centralized entity has a large amount of control over a powerful resource, much of the political uncertainty that results can be mitigated if the entity credibly commits to following a set of programmatic rules for how they apply it. For example, the unsold tokens can be put into a market maker that is tasked with preserving the tokens' price stability.

Third, you can do a capped sale, where you limit the amount that can be bought by each person. Doing this effectively requires a KYC process, but the nice thing is a KYC entity can do this once, whitelisting users' addresses after they verify that the address represents a unique individual, and this can then be reused for every token sale, alongside other applications that can benefit from per-person sybil resistance like Akasha's quadratic voting. There is still deadweight loss (ie. inefficiency) here, because this will lead to individuals with no personal interest in tokens participating in sales because they know they will be able to quickly flip them on the market for a profit. However, this is arguably not that bad: it creates a kind of crypto universal basic income, and if behavioral economics assumptions like the endowment effect are even slightly true it will also succeed at the goal of ensuring widely distributed ownership.

Are single round sales even good?

Let us get back to the topic of "greed". I would claim that not many people are, in principle, opposed to the idea of development teams that are capable of spending $500 million to create a really great project getting $500 million. Rather, what people are opposed to is (i) the idea of completely new and untested development teams getting $50 million all at once, and (ii) even more importantly, the *time mismatch between developers' rewards and token buyers' interests*. In a single-round sale, the developers have only one chance to get money to build the project, and that is near the start of the development process. There is no feedback mechanism where teams are first given a small amount of money to prove themselves, and then given access to more and more capital over time as they prove themselves to be reliable and successful. During the sale, there is comparatively little information to filter between good development teams and bad ones, and once the sale is completed, the incentive to developers to keep working is relatively low compared to traditional companies. The "greed" isn't about getting lots of money, it's about getting lots of money without working hard to show you're capable of spending it wisely.

If we want to strike at the heart of this problem, how would we solve it? I would say the answer is simple: start moving to mechanisms other than single round sales.

I can offer several examples as inspiration:

- **Angelshares** - this project ran a sale in 2014 where it sold off a fixed percentage of all AGS every day for a period of several months. During each day, people could contribute an unlimited amount to the crowdsale, and the AGS allocation for that day would be split among all

contributors. Basically, this is like having a hundred "micro-rounds" of uncapped sales over the course of most of a year; I would claim that the duration of the sales could be stretched even further.
- **Mysterium**, which held a little-noticed micro-sale six months before the big one.
- **Bancor**, which recently agreed to put all funds raised over a cap into a market maker which will maintain price stability along with maintaining a price floor of 0.01 ETH. These funds cannot be removed from the market maker for two years.

It seems hard to see the relationship between Bancor's strategy and solving time mismatch incentives, but an element of a solution is there. To see why, consider two scenarios. As a first case, suppose the sale raises $30 million, the cap is $10 million, but then after one year everyone agrees that the project is a flop. In this case, the price would try to drop below 0.01 ETH, and the market maker would lose all of its money trying to maintain the price floor, and so the team would only have $10 million to work with. As a second case, suppose the sale raises $30 million, the cap is $10 million, and after two years everyone is happy with the project. In this case, the market maker will not have been triggered, and the team would have access to the entire $30 million.

A related proposal is Vlad Zamfir's "safe token sale mechanism". The concept is a very broad one that could be parametrized in many ways, but one way to parametrize it is to sell coins at a price ceiling and then have a price floor slightly below that ceiling, and then allow the two to diverge over time, freeing up capital for development over time if the price maintains itself.

Arguably, none of the above three are sufficient; we want sales that are spread out over an even longer period of time, giving us much more time to see which development teams are the most worthwhile before giving them the bulk of their capital. But nevertheless, this seems like the most productive direction to explore in.

Coming out of the Dilemmas

From the above, it should hopefully be clear that while there is no way to counteract the dilemma and trilemma head on, there are ways to chip away at the edges by thinking outside the box and compromising on variables that are not apparent from a simplistic view of the problem. We can compromise on guarantee of participation slightly, mitigating the impact by using time as a third dimension: if you don't get in during round N, you can just wait until round $N+1$ which will be in a week and where the price probably will not be that different.

We can have a sale which is uncapped as a whole, but which consists of a variable number of periods, where the sale within each period is capped; this way teams would not be asking for very large amounts of money without proving their ability to handle smaller rounds first. We can sell small portions of the token supply at a time, removing the political uncertainty that this entails by putting the remaining supply into a contract that continues to sell it automatically according to a prespecified formula.

Here are a few possible mechanisms that follow some of the spirit of the above ideas:

- Host a Gnosis-style reverse dutch auction with a low cap (say, $1 million). If the auction sells less than 100% of the token supply, automatically put the remaining funds into another auction two months later with a 30% higher cap. Repeat until the entire token supply is sold.
- Sell an unlimited number of tokens at a price of $\$X$ and put 90% of the proceeds into a smart contract that guarantees a price floor of $\$0.9 \cdot X$. Have the price ceiling go up hyperbolically toward infinity, and the price floor go down linearly toward zero, over a five-year period.
- Do the exact same thing AngelShares did, though stretch it out over 5 years instead of a few months.
- Host a Gnosis-style reverse dutch auction. If the auction sells less than 100% of the token supply, put the remaining funds into an automated market maker that attempts to ensure the token's price stability (note that if the price continues going up anyway, then the market maker would be selling tokens, and some of these earnings could be given to the development team).
- Immediately put all tokens into a market maker with parameters+variables X (minimum price), s (fraction of all tokens already sold), t (time since sale started), T (intended duration of sale, say 5 years), that sells tokens at a price of $\dfrac{k}{(\frac{t}{T - s})}$ (this one is weird and may need to be economically studied more).

Note that there are other mechanisms that should be tried to solve other problems with token sales; for example, revenues going into a multisig of curators, which only hand out funds if milestones are being met, is one very interesting idea that should be done more. However, the design space is highly multidimensional, and there are a lot more things that could be tried.