This is a writeup of an idea that I introduced at the meetup in Bangkok herehttps://youtu.be/OOJVpL9Nsx8?t=3h24m51s8

Suppose that you have a social media platform on which anyone can theoretically post content; this could be Twitter or Reddit, some blockchain-based decentralized platform, and the internet itself. One highly desirable thing to have is a way of quickly filtering out content that is obviously malicious, such as spam, scams and impersonations. Relying purely on community downvoting for this is not effective because it is not fast enough, and is also vulnerable to sockpuppet manipulation, bridaging and other tactics. Relying on centralized authorities is a common solution in practice, though carries the risk that holders of entrenched power will abuse it, and the simple problem that there is not enough time for the central authorities to inspect every post. Most recently, the cryptocurrency-focused parts of Twitter have effectively been overrun by scammers to the point of becoming unusable.

Let us consider a different kind of relationship between upvoters and downvoters and centralized moderation. Suppose that for every single piece of content that gets created, there is a virtual market, where anyone can upvote or downvote a post by putting down ETH. One simple market design is one where upvotes and downvotes are both offers to bet at 1:1 odds on the verdict that would ultimately be given by some moderating authority (for now, think of it as a guy named Doug), and once Doug makes the verdict the bets would be evenly partially matched against each other (eg. if 40 ETH bet up and 30 bet down, then each upvoter would only risk 0.75 ETH per 1 ETH they originally put at stake), and then executed. The upvoting/downvoting market is effectively a prediction market on what Doug would ultimately end up deciding, and clients could flag or simply not show posts that have more downvotes than upvotes.

Why use the prediction markets at all, instead of just relying on Doug to give the results directly? There are two reasons. First, Doug is slow; he may be asleep, he may have hardware signing keys that take hours to take out, or any number of other issues could prevent him from making fast decisions. Second, Doug does not have the time or resources to adjudicate every piece of content. In fact, Doug may only have the time to inspect less than 0.01% of all posts made. In this scheme, Doug need only adjudicate a small portion of posts, and can wait until a day or two after the content is posted. The small portion could be selected uniformly randomly, or the probability that a post is selected could be proportional to the amount bet on that post. For any post that Doug does not adjudicate, upvoters and downvoters will simply get their money back, but because of the possibility that Doug will adjudicate any post, people are incentivized to participate in the market, and do so quickly[1], for every post.

Note that this scheme is best used for a limited role, of identifying and removing posts that are unambiguously spam, scams or otherwise malicious; it should not be used as a full substitute for traditional upvoting/downvoting in cases that are any more subjective, as in those cases it really is important that the voting system is polling the community's opinion, rather than the community's prediction of some moderation mechanism's opinion.

Note also that in this kind of scheme, Doug being a centralized actor becomes even more dangerous, because Doug has the ability to insider-trade on these betting markets. This is why it's actually very useful for Doug to be something like a DAO: so that the public can be credibly convinced that it is not capable of coordinating to cheat them in the markets, and so that its voting can be more transparent and predictable. The loss in efficiency from a decentralized moderating DAO is not a problem, because it can be made up for by the gain in efficiency from not actually using the DAO most of the time and instead referring to a prediction market.

The scheme could be manipulated by a malicious actor upvoting their own posts, but this has a cost, and inevitably creates arbitrage opportunities for people willing to vote/bet against them, so it should be expected that the portion of times manipulation attempts succeed is very small (and all manipulation attempts, successful or failed, ultimately contribute to the source of revenue that incentivizes everyone to keep upvoting and downvoting). If more incentivization is required, a specialized forum could force everyone who makes a post to put down some small amount of funds (eg. $0.50) on upvoting their own post; that would turn the game into a kind of superset of conditional hashcash. However, the fact that the basic version of the scheme can simply be overlaid onto the existing internet and require no cooperation from any existing institutions in order to operate is a large plus, as it means that it could theoretically be implemented today (with the caveat that transaction fees would need to be much lower, so sharding is likely required).

[1] To encourage rapid participation, the market design I suggested would not work, as it has no incentive to vote earlier rather than later. The alternative is a traditional on-chain market maker, like an LS-LMSR, which does have the incentive to get one's bets in first, but on-chain market makers have the challenge that they require someone to put up capital for each vote. A happy medium could be a system where upvotes and downvotes are used until some small quantity of ETH is bet by both sides, but where the bets on both sides are at less than 1:1 odds (eg. 1.2:1 odds could work), and once the total quantity of upvotes and downvotes reaches some level the system switches into an LS-LMSR, using the implied "fee" siphoned from the existing bets to initially seed the market maker.