

We are currently running an experiment called HumanityDAO, which attempts to maintain a registry of unique humans without a central authority (<https://www.humanitydao.org/>).

The ID verification game is based on web of trust, in which new applicants must be connected to verified humans (on Twitter) to join the registry. New humans earn voting tokens, which can then be used to vote on new applicants.

Interestingly, the payoff matrix for validators looks similar to a prisoner's dilemma game. If the validators cooperate (i.e. vote YES to real applicants and NO to Sybil or duplicate applicants), the registry of unique humans is accurate, which means it can be used in Sybil-resistant smart contract protocols and the token should therefore increase in value. Validators can defect by voting incorrectly either to let Sybils in or just to troll the system. A 1959 paper ([https://en.wikipedia.org/wiki/Prisoner's\\_dilemma](https://en.wikipedia.org/wiki/Prisoner's_dilemma)) shows that rational players can sustain a cooperative outcome in an unknown-round game.

I am curious whether this system can be improved and generalized as a proof-of-stake consensus mechanism. Early results are promising, but the system has yet to be truly battle-tested. I also don't know how it would hold up with the additional complexities of a distributed computing environment.

Apologies for trying to explain this on Twitter, I forgot this forum existed.