Kickstarter for Coordinated Action

Is there an assurance-contract website in work?

I bet I'm not the only one who, after finishing inadequate equilibria, thought in excitement "Ok, so where's the KickStarter for better Nash equlibria?".

I didn't find any existing site that does something like it, but i wonder if maybe someone in this community is working on it.

If you know anything about it, I'd appreciate a hint ;)

If a "Kickstarter for Inadequate Equlibria" was built, do you have a concrete inadequate equilibrium to fix?

[Yoav Ravid asks](): "Is there an assurance-contract website in work?"

i.e. a site where, if there's a locally bad equilibrium that would be better if *everyone* changed strategies at once, but which requires a critical mass of people in order to be worthwhile, you can all say "I'll put the effort if other people put in the effort", and then if X people agree, you all go into work the next day and demand a policy change, or a go to a political rally, or change a social norm, or whatever.

Some attempts have been made at such a system. It's not that technically hard to build. But I think it'd need a couple major "flagship" Coordinated Actions in order to rally people's attention and turn it into a more frequently used tool.

So, *if* a good website existed to coordinate action, do you have a well operationalized action you'd want to coordinate? ("Everyone leaves Facebook at once" doesn't work IMO, because it doesn't say where people are moving *to*, or otherwise replacing FB's tools with.

"Everyone on one platform switches to another platform" seems viable.

"Everyone at my office signs a letter demanding change for a particular policy" seems viable (although in cases like this, where you maybe *don't* want your boss to know you're planning a revolution, and I'm not sure how to best achieve common knowledge without risk)

(For further reading, see "[The Costly Coordination Mechanism of Common Knowledge]()" and "[Inadequate Equilibria]()")

How could "Kickstarter for Inadequate Equilibria" be used for evil or turn out to be net-negative?

Following up to "[If a "Kickstarter for Inadequate Equlibria" was built, do you have a concrete inadequate equilibrium to fix?](#)"

I *think* a kickstarter for coordinated action would be net positive, but it's the sort of general purpose powerful tool that might turn out bad in ways I can't easily predict. It might give too much power to mobs of people who don't know what they're doing, or have weird/bad goals.

How bad might it be if misused? What equilibrias might be we end up in in the world where everyone freely has access to such a tool?

Ideas for an action coordination website

inspired by Inadequate equilibria, and following the ' [*KickStarter for Coordinated Action*](#)*' sequence.*

This is an idea-dump post for a website i thought of after reading Inadequate equilibria.

Today, tools like Facebook and twitter help us coordinate better and faster. but still, that is not enough to solve problems of "high-inadequacy" - where we're stuck in bad Nash equilibrium, and moving away from it demands many things to happen together.

The idea here is to take coordination much further, so we can solve as much of the game theory problems bound in moving to other Nash equilibria.

The goal is to allow 'Exoduses' from bad Nash equilibria, to better a Nash equilibria.

Note - none of this (at least as of writing this) is being worked on, nor are there currently plans to do so. Though, of course this doesn't mean there shouldn't be, or else i wouldn't have wrote the post :)

basic concept

big problems can't be solved by a single person. but sometimes even a large group of people who agree on the problem and the solution, and are even motivated to bring the change - can't do it. That is because [some](#) [problems](#)are more complicated than that.

one such complication is coordination problems - where "everyone is taking some action A, and we'd rather all be taking action B, but it's bad if we don't all move to B at the same time." or each individual wouldn't want to take action A unless he knows the rest do too.

solutions to coordination problems include common knowledge (Where all know the intentions of all actors) and pre-commitment. that's the site's goal.

Coordinated actions

Any user can become a coordinator/initiator by creating a **coordinated action** (CA), in order to solve some problem. Each CA gets its own page (similar to kickStarter). Any user (unless otherwise specified) can obligate to the action - and here's the catch - the commitment is to be realized if and only if a certain number of other users obligated the same. In KickStarter, we commit to pay, here we commit to take action.

Versatile coordination:

Since reality is very dynamic, a rigid structure will work for few situations and be less useful. So to allow a large array of projects, options for CAs should be versatile. I see it as an ongoing project that will develop next to the community's needs. think of how

tesla is implementing features in dialog with their community's wishes. here are some examples of "contracts":

Basic: all of us obligate to some CA if X others do too.

*milestones***:** many actions, listed on the same initiative, which are taken at different amounts of obligations. It is either required to obligate for all actions, or possible to obligate to specific actions only.

*Obligation for obligation***:** a group or individual obligate for something, if a different group or individual obligate for the same or a different action (only one side can be an individual).

Please comment with more types that you can think of, it's very interesting.

further support

And sometimes just committing isn't enough for you, or you support the initiative but done that action (in cases of one time decisions, like going vegan, going zero-waste, getting rid of your car, etc...). So we want to give users a way to support further than just obligating. One example is an option for users would to donate money to the campaign, which will be used to further spread it, through some kind of advertising.

Communities

This is where it gets quite complex, but it has to. people don't want a thousand random humans around the world to something with them. they want people from their country, from their city, from their profession/hobby/Interest-area, from their social circle or organization.

That's where communities come in, the ideal is that for every real-world community you would be able to create a community on the site to resemble it, and that if people from that community are already on the site, they will find out that a new community they belong to has been opened, and will join. the reason is, so it's possible to coordinate action with and within certain communities.

When i first thought about how communities will work in practice it reminded me of set theory, but thinking about it more, it ended up merely resembling it (and probably breaks some of its laws). still, hopefully mentioning it helps to visualize.

the communities structure

There are many communities, it's easy and accessible to create new ones, or join existing ones (Unless said community has some requirements). any community is a sub and/or parent community of other communities. The goal is that the communities on the site will be able to reflect the communities in the real world.

some examples of communities: The earth (The parent of all communities until we colonies mars) and the user base[1], all regions/continents, and all countries, are some that can be added per-launch. examples for user added communities: cities, EA, farmers is Israel, bus drivers in new York, vegans is the US, LessWrong, etc...

"that's a lot communities..", you say? "like, **a ton** of communities", yeah, that's true. but that's how it should be.

it might sound a bit like FB, but except the community grouping aspect, there are two more important differences. here we don't want two groups which are basically the same (Cause that's ineffective, if the real world community is divided between them on the site), and the user doesn't have to be aware of all the communities they're part of[2], which may feel weird in a online platform, but that's how it is in real life.

Whenever a user joins a community he is suggested sub-communities he might fit in, and is automatically added to all parent communities, which he can manually exclude themselves from (I bet the set theoreticians winched). part of the account creation process would be spotting the user's communities (easy ones, for example, are countries and cities)

organizations

not only individual humans are "agents" in this world, but also some communities, like, corporations, non-profits, and any other goal oriented group. organizations will be able to create a community around themselves, but also act as a user on the site. this is important, cause some coordinated actions seek not only the cooperation of single humans, but of the groups they make. if vegans think of taking a coordinated action together, they want to know that the businesses and organizations they support will go with them too. some CAs will be relevant only for organizations, where it doesn't matter how many individuals commit to take some action, they have to have cooperation from their own group.

Motivation and verification

At least until all of our society is nice, educated and rational - We need two mechanisms, one that verifies who really cooperated and who defected, and a mechanism to discourage defecting, and encourage cooperation.

fulfillment verification

Verification is the much harder of the two. how do you verify that someone tried veganism for 28 days? how do you verify that someone has/hasn't posted certain types of posts to FB? Can you verify whether someone really voted third party? whether they went to work that day? Some stuff are easier to verify, but if we stuck only to actions that are easy to verify, this tool won't be very useful. It's a hard-shell to crack, but it needs cracking - ideas?

user verification

In this system it's important that we know user = person, even better if we know user = which person. It can also help with fulfillment verification. There are many ways to do it - Email, phone, face, PayPal(?), ID. it just needs to be. maybe not all users have to be verified, yet still possible for communities and CAs to require verification.

Motivation

given that we solved verification - motivation is simpler. a few options:

cooperator/defector score: users have a publicly displayed score that shows how they acted on their obligations.

Achievement badges: I envision something similar to khan academy's badges, but harder to get so they're more meaningful. an example of one "**good to have:** on average, you referred to each coordinated action you obligated to, at least 10 fulfilled obligations." You can display these badges on your profile to signal how *awesome* you are :)

putting your cash where your mouth is: for each CA, either there's a set amount or the initiator sets it, each user deposits cash against his cooperation - if he defects, he looses that money, if he cooperates, if he wins some extra money. This is a pretty much bound-to-work motivator (unless bill gates starts using this too), But i'd rather incorporate money as a last resort, If we find that we really need this extra motivator, since it makes everything more complicated. It also makes the verification task harder, since if people can use the site as a money-pump, they'll be more likely to look for a backdoor to exploit.

I believe this concept, if it was successfully realized, could bring great benefits to the world.

1. ⌃

   yeah, both are all users, but i see reasons to differentiate, sorting wise - communities might have a counter feature of how many people in the real world community are registered to this online community, in the users it's a 100%, in the world.... you'd target different CAs to the user base and "the whole world". and, Sub community sorting is based on the real world, not the website, so the farmers of Israel are only part of the earth community and not the website's User Base. possible sub-communities are active users, contributors, etc...

2. ⌃

   Hard to estimate exactly how much, but it's at least a few dozens and maybe more than a hundred

Making a Crowdaction platform

In 2017 Eliezer Yudkowsky [wrote](#):

> Coordination isn't as simple as everyone jumping simultaneously every time one person shouts "Jump!" For coordinated action to be successful, you need to trust the institution that says what the action should be, and a *majority* of people have to trust that institution, and they have to *know* that other people trust the institution, so that everyone *expects* the coordinated action to occur at the critical time, so that it makes sense for them to act too.
>
> That's why we have policy prediction markets and… there doesn't seem to be a word in your language for the *timed-collective-action-threshold-conditional-commitment…*hold on, this cultural translator isn't making any sense. "Kickstarter"? You have the key concept, but you use it mainly for making video games?

I thought this was very clever and was happy he broadcasted this idea to a community of capable, well-off computernerds. "I can't wait to see what this community cooks up" I thought. We are currently halfway into 2020 and the dream has not yet been realized. Let's see where we stand.

CollAction

I've searched the web and the closest I could come is [CollAction](#). These people have gone ahead and build a functional website with actual projects on it. The makers of this site call these kinds of websites '[Crowdacting](#) websites'. Since 'inadequate equilibria' didn't name this idea I shall be calling it that from here on out.

The site allows you to upload your own projects (with their approval) and allows you to join other projects. The visually clean lay-out shows how many people have already joined a project and what the desired threshold is. The projects have clear deadlines and goals and once they are reached they are closed forever. I encourage you to check it out for yourself since it's a good way to tickle your imagination about what such a website may look like in the future.

While the site is very pretty, it's not made by this community and as far as I can tell also not used by this community. Not very surprising since it's very barebones. Let's run through a list of features I am missing that I expect to see on a Crowdacting website.

Expected Features of a Crowdaction website

Broad range of projects

The site currently only has projects about crowdacting ecological problems running. Your project also needs to be manually approved by them and it's not entirely clear what they will/won't accept. I'm guessing some level of review is going to be necessary, but there is no reason a crowdaction site should be so narrow-scale.

A Milestone system

[Yoav Ravid](#) suggested we use a milestone system similar to kickstarter. This wouldn't work for all types of projects but could certainly be included. He also suggested badges people can have next to their name if they e.g help push over a milestone or make good on their promise.

Talking to humans

Their site has no real defenses against bots. When you promise to protest something if a thousand other people will show up too, you don't want a thousand bots to sign up and make you the only person to show up to the protest. Something like [it's me](#) should be able to keep out bots and make sure the collaborations is always between actual humans.

Gated communities

When you're trying to organise a strike in Amsterdam you don't want people from New York to interfere with your collaboration project. Those projects should be closed off to the wider world with only community members being able to join in. Some way to signal to an algorithm what kinds of projects you would be interested in joining is also nice and communities help with that. [MakoYass](#) proposed a community system based on [SetTrie](#). I would couple this with community chatrooms so the members can better discuss things amongst themselves.

A registry

For some projects you want to contribute anonymously, for others it's important that people know you support it. So I would add the option to commit anonymously, but the site still vouches for you being an actual person. On the project page should be a registry of people who are committed to this project with the option to upvote and downvote these names. Next to the names people should be able to write a couple words summarizing their influence over this problem. This way CEO's of relevant businesses (or other relevant organization leaders) will be able to get upvoted so that people see what organizations are also joining in.

A voting system

Sometimes people agree that the current equilibria is bad, but they don't agree where they should move to. On the project page should be a [STAR-voting](#) feature, where people can submit their equilibria of preference and vote on the equilibria they want to move to. For example: hospitals have different administration systems and want to use the same one. Most hospitals use administration system 1 so administration system 1 comes ahead in the voting ensuring that the transition inconveniences the least amount of people. This voting mechanism should probably close well before the project itself closes so it's always clear what people sign up for.

Complying with the law

I know this could be a tool to protest stupid laws by e.g joining together to start smoking cannabis in front of the White House. I would urge that the first site does not condone the breaking of laws, since that might kill the reputation of crowdacting and might get those types of sites banned by governments. (this is something CollAction thankfully already does) This doesn't have to be as bad as it's sounds. People will almost certainly find a way to work around this limitation. For example "Let's all

smoke in front of the white house to protest strict cannabis laws" can stay on the site because smoking in itself isn't illegal (they could be talking about smoking tobacco).

Letting different people do different things

This kind of goes together with the registry. If certain community members can't participate in certain ways but can in others, let them. Have different counters running on the project e.g: one counter saying 320 out of 500 have signed up to do X and a second counter saying 100 out of 200 have signed up to do Y. Yes this makes it more complex, but in real life you often need different types of people to do different types of things. This could be incorporated into the registry with flairs like: Mr. Robin Handsome will do Action 1, Action 2 and Action 5. This allows for more complex coordination problems to be solved.

In case of failure, follow-up suggestions

Let's say you want to start a project titled "Change hospital administration systems in the U.S". You start the project and it get's some traction but ultimately it falls short of the threshold. Ideally some algorithm could pick out which sub-communities were motivated and automatically suggest the same project on a smaller scale. If the algorithm noticed that people in California were committed to the project, it could suggest a new project titled "Change hospital administration system in California". This can keep the momentum of a project going. People could also do this manually thanks to the registry, but they might not think of it.

Fulfillment Verification

The hardest problem might be to make sure the project actually goes through and everybody does their part. While this might be impossible to check on an individual basis I think I have thought of something that might broadly solve this in practice. Everyone submits 5 dollars to the project. Some time after the end of the project (depends on the project but for most 48 hours will do) you get the question: "Did this project succeed in it's mission statement?". If you say 'yes' while more than 60% say 'no' you only get part of your money back. The exact amount of money you get back depends on how big the gap was between what you said and what everyone else said. So someone who said 'yes' while 69% said 'no' will get way more money back than someone who said 'yes' while 99% said 'no'. The same is true in reverse if you say 'no'.

This means you try to predict what everyone else will say which will usually correspond with what actually happened. (You also get your money back if you say 'yes' but less than 60% say no just so there's a margin of error). Because this means that overtime you will lose money on the platform you can actually win some money by repeatedly being in projects that then end the way you said it ended. This money will come from people who didn't say it ended like the rest said it ended.

Getting this project off the ground

I'm honestly kinda sad that this site hasn't been build yet. I would if I could, but I simply don't have the expertise in webdesign to pull this off. But if anyone starts working on it I will be more than happy to help in any way I can (graphic design, translation, marketing...) I do however have a suggestion as to how we can funding for this website. Use kickstarter! Use a crowdfunding website to get a crowdaction

website off the ground. I will personally pitch in of course, but I think a lot of people inside and outside this community will be interested too.

I feel like this idea has a lot of potential, but I would like to start working on it sooner rather than later. If anyone has any more suggestions about what a crowdaction platform should have, please put them in the comments and I might add some to the list.

Extracting Value from Inadequate Equilibria

[Much expanded from my comment [here](). Pure speculation, but I'm confident that the bones of this make sense, even if it ends up being unrealistic in practice. Cross-posted from [Grand, Unified, Crazy]().]

A lot of problems are coordination problems. An easy example that comes to mind is scientific publishing: everybody knows that some journal publishers are charging ridiculous prices relative to what they actually provide, but those journals have momentum. It's too costly for any individual scientist or university to buck the trend; what we need is coordinated action.

Eliezer Yudkowsky talks about these problems in his sequence [Inadequate Equilibria](), and proposes off-hand the idea of a [Kickstarter for Coordinated Action](). While Kickstarter is a great metaphor for understanding the basic principle of "timed-collective-action-threshold-conditional-commitment", I think it's ultimately led the discussion of this idea down a less fruitful path because Kickstarter is focused on *individuals*, and most high-value coordination problems happen at the level of *institutions*.

Consider journal publishing again. Certainly a sufficient mass of individual scientists could coordinate to switch publishers all at once. But no matter what individual scientists agree to, this is not a complete or perfect solution:

- Almost no individual scientists are paying directly for these subscriptions – their universities are, often via long-term bulk contracts.
- University hiring decisions involve people in the HR and finance departments of a university who have no interest in a coordinated "stop publishing in predatory journals" action. They only care about the prestige and credentials of the people they hire. Publications in those journals would still be a strong signal for them.
- Tenure decisions involve more peer scientists than hiring, but would suffer at least partly from the same issue as hiring.

What's needed for an action like this isn't a Kickstarter-style website for scientists to sign up on – it's coordinated action between universities at an institutional level. Many of the other examples discussed in *Inadequate Equilibria* fit the same pattern: the problems with healthcare in the U.S. aren't caused by insufficient coordination between individual doctors, they're caused by institutional coordination problems between hospitals, the FDA, and government.

(Speaking of government, there's a whole host of other coordination problems [climate change comes to mind] that would be eminently more solvable if we had a good mechanism for coordinating the various institutions of government between countries. The United Nations is better than nothing, but doesn't have enough trust or verification/enforcement power to be truly effective.)

The problem with the Kickstarter model is that institutions qua institutions are never going to sign up for an impersonal website and pledge $25 over a 60-day campaign to switch publishing models. The time scale is wrong, the monetary scale is wrong, the

commitment level is wrong, the interface is wrong… that's just not how institutions do business. Universities and hospitals prefer to do business via contracts, and lawyers, and board meetings. Luckily, there's still value to be extracted here, which means that it should be possible to make a startup out of this anyway; it just won't look anything like Kickstarter.

Our hypothetical business would employ a small cadre of lawyers, accountants, and domain experts. It would identify opportunities (e.g. journal publishing) and *proactively* approach the relevant institutions through the proper channels. These institutions would sign *crafted, non-trivial contracts* bound to the success of the endeavour. The business would provide fulfillment verification and all of the other necessary components, and would act as a trusted third-party. The existence of proper contracts custom-written by dedicated lawyers would let the existing legal system act as an enforcement mechanism. Since the successful execution of these contracts would provide each institution with significant long-term value, the business can fund itself over the long haul by taking a percentage of these savings off the top, just like Kickstarter.

This idea has a lot of obvious problems as well (the required upfront investment, the business implications of having its income depend on one or two major projects each year, the incentives it would have to manufacture problems, etc) but with a proper long-term-focused investor on board it seems like this could turn into something quite useful to humanity as a whole. Implementing it is well outside of my current skillset, but I would love to see what some well-funded entrepreneur with the right legal chops could make of something like this.

Thoughts?