Part II - On Trust

In Part I - My Identity, we came to the conclusion that your identity is usefully split between yourself and your community. Which leads us to ask, who is your community? To some extent, that can be answered by saying, *your community is who you trust to hold your identity* - almost a circular definition but with that important link, trust.

To understand Identity, then, we must first understand trust. This section looks at *trust* - what it is, how it comes about, and more importantly how can we tame it for the purposes of assisting Alice with her life? We come back to *community* in the next part.

II.1 Micro-trust - Alice trusts Bob

> *"What's in a name? that which we call a rose*
> *By any other name would smell as sweet."*
> William Shakespeare, *Romeo and Juliet*

The Essence of Relationship

Let's assume that trust is something that humans do with each other[1]. Let's start with a basic scenario of two people, Alice and Bob[2], who are approximately equal as human beings. There is something that draws them together:

> *Alice trusts Bob...*

Although nice, such statements use romance to avoid a lack of precision. When a person trusts another, she makes a decision over a particular question of some current interest. When we say the above, we mean something like

> *Alice trusts Bob to make dinner.*

being that Alice makes that decision over that action, and not over another. There is some risk involved, and these two can be both true

> *Alice doesn't trust Bob to mind her kids.*

Or not, or either/or. Context is important! And therefore, trust isn't some universal thing, it's a set of situation-result pairs. Indeed, because we know that there are so many situations - *does Alice trust Bob to go shopping?* - we can also suggest that there must be some element of purposeful decision by Alice.

Go to the movies with Bob? Introduce Bob to her parents? Ask Bob to recommend an app or a dress or a garden gnome - Alice knows the answer in most cases pretty instantly.

Which tells us something else:

> *Alice knows something about Bob.*

---

[1] Which is to say, we are for now ignoring other idiomatic uses of the term trust: you trust your car starts, you trust an institution, you place your loyalty and trust in the party, etc.
[2] In this cycle we use the cryptographic convention of Alice & Bob as first and second persons respectively. In English, this also allows us to use third person gender as an efficient signal, but that doesn't work in Spanish.

Alice has a base of experience with Bob that supports her decision.

Then: *trust* is from one person to another, concerning a particular question, in which the decision to trust is made quickly based on experience and information already known about the person(s) and the situation.

So far so good. Yet there is more to say: Let's ask

> *what does Alice think about Bob's gender?*

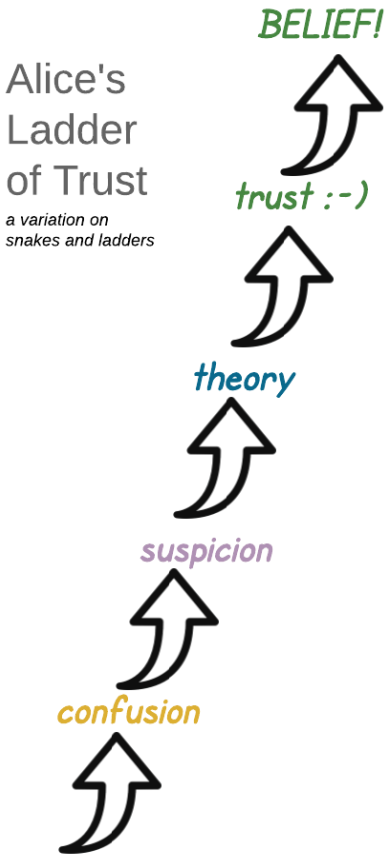
This is a particularly interesting question because in classical thinking, gender is both unchangeable and knowable. Although in modern times we challenge the stereotype, and gender is a delicious plot element in comedies, the basic claim is remarkably reliable:

> *Alice trusts Bob is male.*

For Bob's sake, we're glad that is sorted out!

On Knowing

> *"Alice: Would you tell me, please, which way I ought to go from here?*
> *The Cheshire Cat: That depends a good deal on where you want to get to.*
> *Alice: I don't much care where.*
> *The Cheshire Cat: Then it doesn't much matter which way you go.*
> *Alice: ...So long as I get somewhere.*
> *The Cheshire Cat: Oh, you're sure to do that, if only you walk long enough."*
> Lewis Carrol, *Alice in Wonderland*



But actually, that's not quite right! If we ask Alice what she thinks of this question, she will not disagree but would prefer to say:

> *Alice knows that Bob is male.*

Leaving aside how Alice knows this, what is interesting here is that knowledge is a stronger thing than trust. Alice *trusts* Bob

to go shopping, but he might forget - and may incur Alice's wrath. On the other hand, Alice *knows* Bob is male, and there is no question about this; he will not return from the shopping mall with a sudden shift in gender.

Which speaks to certainty - if we have complete certainty, then knowledge is used, but if we have a doubt, then a trust decision is called for.

The boundary between trust and knowledge is a fascinating place. A child for example knows the gender of parents - but this was not always the case. A baby of 0 days old knows no such thing, indeed, knows nothing, not even what a male or female is, nor what a person is. As the child grows, first there is mother, then there is father, who perversely is primarily useful only for providing evidence for mother, and later on there is a metamorphosis into the general classes of males and females.

And so it is with all things: Our state of mind over some vague question transitions from nothing to a formative idea or suspicion to a theory to trust to certainty. Read the following table from the bottom to the top:

Name or Level	**Description**	**Process**	**Source**
Knowledge		Belief, Truth, Facts - we are beyond process	Life experience, authority, religion, family, dress
Trust	Analyse by cues at a distance, confirm as more information arrives	Observation of cues	Many prior sources that have generally proven reliable
Theory	gender can be guessed from clothing, shape, voice, etc	Testing, listening	child tests and asks each new person - what gender are you?
Suspicion	observations mostly point in a direction, some point wildly elsewhere.	Observation, curiosity, questioning	senses something different and discordant between choices
Confusion	sense of ambiguous and contradictory information	Angst, discordance	sees differences but cannot isolate
Nothing	absence of any thought on the subject	ignorance	Light, mystery, amusement, love

From trust to certainty, what happens? We no longer make conscious decisions, *we just know,* or in the lingo of the psychologist, we internalise the knowledge. Which is to say, in the transition from trust to knowledge, we forget how we got here because it isn't worth the mental energy anymore.

> *"To believe is to know you believe, and to know you believe is not to believe."*
> Jean-Paul Sartre

What happens between a theory and trust? A theory is for fun experimentation - we try things out and watch what happens. Gambling is maybe a theory with rewards and losses; Bob did a mean steak in the past, let's try him out on a pizza?

The differentiator is doubt. When there is substantial doubt, we must use other strategies, but when there is substantially less doubt, we can trust in our decision making and run with the consequences.

Then, Alice trusts in Bob when she has enough information to take a risk - with the reliability of knowing her losses are both acceptable and less than her wins.

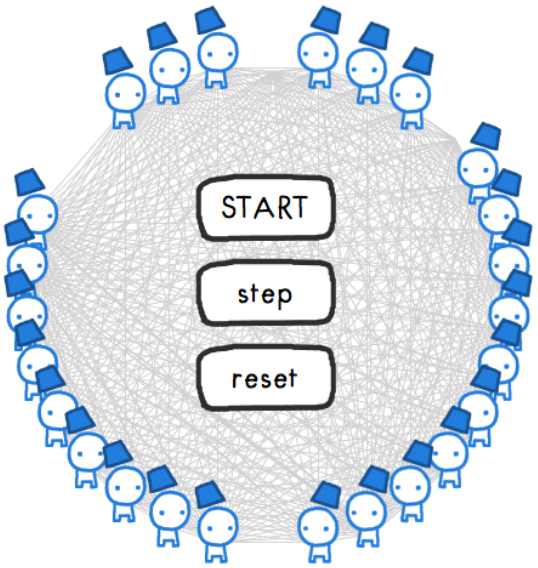
So trust is inherent with risk - the loss from a wrong decision is balanced by the reward of a good decision. If there is zero risk, it would be knowledge or irrelevant. If there isn't enough probability of net wins for Alice to go with the risk, then it is not trust, but hope, gambling, experimenting, investment in future knowledge, and present knowledge that she has to deal directly, immediately with failure.

Trust is therefore that space in human observation where the probability of success is high enough to take a risk, and accept the consequences, but not high enough to internalise as knowledge.

Skin in the Game

> *"Trusted" means "Someone who can screw you over by acting in bad faith."*
> Ray "Bear" Dillinger, *Cryptography list,* 20170115

In this game, on making a *decision*, Alice takes on *risk*. Bob's delivery of his obligations to Alice include some non-trivial uncertainty. E.g., If Alice does risk leaving her kids with Bob, he could drop a pot of boiling water on them, lose them in the park, generally fail to feed them or lose them playing on the freeway.



He probably won't, but he might.

Alice accepts the risk in exchange for some reward. That reward might be intangible and incalculable, but it is generally present in some sense or other - and therefore Alice has to have *skin in the game*! It doesn't make a lot of sense to expose herself to some risk without a commensurate benefit, even if the payoff is small, perceived, manipulated.

Having received the payoff, Alice can analyse the consequences of her decision. Once the deal is done, her analysis is not constrained in time or event - it can happen immediately, slowly, or even wait for a future proposition.

These propositions come along in a continuous series, and are subject to continuous analysis - unconsciously or consciously - until she finds a proposal that offers her a profit. Or not. Therefore we can say, in the sense of game theory, that

> trust cannot exist in a one round game.

The only way that trust can exist - build up - is within a repeated round game which is unbounded.

If you are unfamiliar with this notion of game theory, and how it impacts trust, spend 30 mins or so going through "The Evolution of Trust." It's a fun interactive demo of game theory.

The Cycle of Trust - RADR

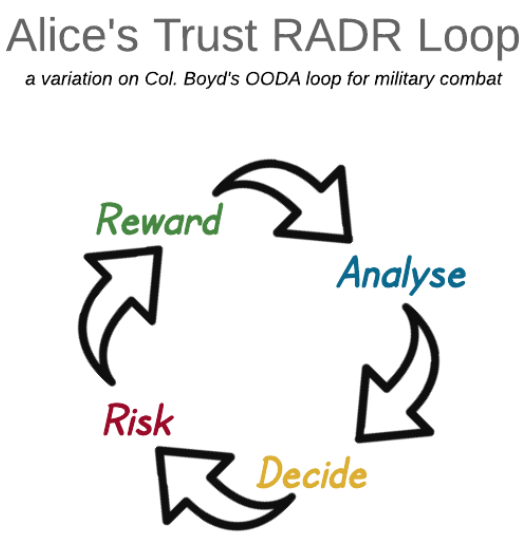
> *"If there's one big takeaway from all of game theory, it is this:*
>
> ***What the game is, defines what the players do.*** *Our problem today isn't just that people are losing trust, it's that our environment acts against the evolution of trust.*
>
> *That may seem cynical or naive - that we're "merely" products of our environment - but as game theory reminds us, we are each others' environment.* ***In the short run, the game defines the players. But in the long run, it's us players who define the game."***
> Nicky Case, "The Evolution of Trust."

Against any proposition she receives, Alice's analysis needs to promise her that the expected reward will be in excess of her costs. When she finds that, she decides - again!

But her analysis can only be well informed by previous experiences. Therefore, Alice's trust machine is a loop (or spiral), one per each person that she knows. Starting at any point, it continues with each person, going around and around. It has to be a loop because the analysis

of any new proposition can only be based on facts which relate to the proposition - and the best proposition is one that looks like the old facts, in which a decision was taken, a risk held against a reward delivered, allowing analysis for a future round.

Trust is by definition circular - once a trust cycle has been run a few times, Alice can develop an expectation of future returns.



This expectation of the future allows for favours - Alice can ask for a favour, which in some intangible sense she is expected to return at a later date, or she can deliver favours, and store them in the 'trust bank' for a future call. In such a way, a decision can be built from many such decisions, allowing for amortization of the costs of each decision over collective events - trust builds over time. (If you've spent some fun time on "The Evolution of Trust" you will perhaps notice now how Alice's trust machine is so much more capable than say Copycat or Copykitten, which have almost no memory, or very limited memory.)

Alice's trust of Bob always involves a decision based on some base of information. Hopefully, that information is a collection of past interactions. Even better is if these interactions include a dose of prior trust decisions, in which case her trust increases with each succeeding successful event.

If not, in the absence of previous experience, her decision can rest on proxies such as recommendations, personality, metrics, likeness, tests, environment, safety, customs, etc, but in that case, her risk goes up so Alice would typically reduce the value at risk. Alice might start with something very small - something she can afford to lose.

Limiting Trust

"Don't speak
I know just what you're saying
So please stop explaining
Don't tell me cause it hurts
Don't speak
I know what you're thinking

> *I don't need your reasons*
> *Don't tell me cause it hurts"*
> Stefani & Stefani, "Don't Speak," 1995 (No Doubt)

Around the general idea of trust, we can place a number of limits to keep Alice safe.

Identity! for Alice to run her *reward-analyse-decide-risk* process over Bob, she needs to know that the Bob in spin 1 is the same as the Bob in spin 2, 3 through 100. In short, she needs an *identifier* to the person she's trusting and a little internal database collecting all the RADR traces from that person. Trust by its very nature is a many-round game, and no trust builds up over a person if he's different every round.

Then, if we were trying to pigeonhole trust, we could say that *to Alice*, Bob's *Identity* is the sum of all the interactions she has done over *identifier* Bob. And, her *Trust* of him is the summed results of all of her RADR experiences. But this is a crude simplification, we know that people are more complicated than pigeons.

Precision. Alice might trust Bob in one question, but that will not relate to another question. For example, Alice might trust Bob with a loan of $10 but she might not trust him to look after her children. Or vice versa, there is no necessary relationship between the questions, only that they exist, and each has to be analysed independently.

Expectation. Alice only proceeds if her expected profit is in the positive. Cost(Decision + Risk + Results + Analysis) < Reward (Results).

Choice. She only *trusts* Bob when she *chooses* to do so. If Alice has no choice, all bets are off, she's in *compliance-mode* not trust-mode. If for example she asks the state registry for a Driving Licence, she is not choosing the state over some other supplier. The state registry is the only authority delivering the document, and so there is no choice.

Authority. Alice has rights. If she exercises these rights, and takes no risk, then there is no trust expended. Likewise if she makes no risky decision, then no new information comes back. E.g., when she gets her Driving Licence, Alice doesn't find herself trusting the state registry more; she has a right to drive, and that right will persist. In contrast, when she drives past a police car, she worries that she'll be stopped. She has no right to speed, her choices in driving then have a lot to do with her taking risks, and her trust of the police to enforce or not is a continual issue.

A definition of trust

> *"The word 'risk' derives from the early Italian risicare, which means 'to dare'. In this sense, risk a choice rather than a fate. The actions we dare to take, which depend on how free we are to make choices, are what the story of risk is all about. And that story helps define what it means to be a human being."*
> Peter Bernstein, Against the Gods: The Remarkable Story of Risk.
> Gunnar: twitter.

Let's summarise these limits:

- A decision is made
- to take a risk
- over a known person
- in substantial uncertainty
- over some limited question or context
- in exchange for some benefit.
- Voluntary choice is essential
- Based on the history or experience of the maker of the decision over that person.

Enough blahbla, can we come up with a definition? Let's try:

> *When Alice trusts Bob, she chooses to take a risk on Bob's actions in a limited context, based on her prior experiences, in order to gain some expected benefit.*

Like Alice, I take a risk on that definition - if I get it wrong, you will write at length how my definitions are untrustworthy; on the other hand, you may have to take a risk on that definition too, in order to try it - to find your own base of experience, to find out if it works. Trust me?

II.2 Dynamics of Trust

> *"How did you know, 'Cause I never told*
> *You found out, I got a crush on you*
> *No more charades,*
> *My heart's been displayed*
> *You found out, I got a crush on you."*
> The Jets, *Crush on You*

The foregoing was a closed system - Alice trusted Bob, once, in isolation. Then, Alice moved on to trusting Bob again and again. Adding Alice's Trust RADR, we see there is a distinction between a trust decision (discrete) and a body of trust (continuous).

Let's go further. What Alice does in trusting Bob is non-deterministic, non-modellable, non-normative. It's also dynamic. In the larger sense or continual context, we might say

> *Alice trusts Bob (always) to go shopping*

to mean that we expect Alice will take a risk with Bob in some future shared understanding. We can then see two different uses of the word trust, being over a particular question and instant, or, over a set of questions of related context, spread over time. I suggest that understanding the one depends on understanding the other; above we concentrated on the singular or discrete context, and now it is time to concentrate on the continuous.

How to break Alice's trust

> *"Perhaps your definition of your self-system lacks authentic boundaries. You've erected a precarious structure of personality on unconscious factors over which you have no control. That's why you feel threatened by me."*
> Philip K. Dick, _*Ubik*_ 1969

Above, we assumed that Bob and Alice were approximately equal in some sense. Let's now consider a multi-decision game where Bob breaks Alice's trust model.

Because of the supercharged benefit that Bob can extract from Alice if he can correctly predict Alice's decision, he tries. And occasionally at least, he succeeds, and Alice loses out.

Which means that Bob is now no longer an uninteresting and uninterested part of the model - Bob's actions on Alice's trust are part of Alice's decision making. There is no algebra of trust, but there might be a calculus of trust.

In each succeeding round, Alice is forced to defend her model from prediction. She is forced to bury her thought processes and decision making behind any tool she has: vaguery, politeness, ditzyness, lies, ivory tower logic, belief, excuses, stonewalling… in order to stop Bob abusing her model. Alice's model becomes very subtle, indeed, in time, Alice becomes too subtle for even Alice to know.

Then, **if we model how Alice trusts, we break it** - Alice is forced by extraction of benefit to both stop trusting in the beneficiary, and *to change her model*. Until she reaches a point where she is defended in her model.

Trust then is Heisenbergian - we can know that trust is there, but we cannot know what that mechanism is. We may be able to measure the trust, but in the act of revealing it we break it. Finally, a use case for quantum computing!

I, Troll #UASF
@brian_trollz

🤔 But if we can't trust technology....and I put technology in my head...how could I ever trust myself again?

3 Aug 04:31

Trust and the Machine

> 🤔 *But if we can't trust technology....and I put technology in my head...how could I ever trust myself again?*
> @brian_trollz, https://twitter.com/brian_trollz/status/892935917148807168

This observation above, that if we model trust we break it, leads us to some tantalising hypotheses.

Machines cannot do trust. If they could, they would have broken Alice's model, and would then become Alice's vulnerability - forcing her to change her model, or hard-fork the machine. She might not realise it at first, it might take a while, but the longer it takes, the angrier she'll be.

Trust can't be interposed. Imagine a perfect machine that captured Alice's trust within and analysed the activity to assist Alice. As the machine gets better, it would become more valuable. Bob would attack the machine - or someone else would.

We can see this spectre of interposed trust in today's social networks and shopping sites - first they gain Alice's trust - or repetitive custom - then they bombard her with adverts, then they sell her data to every other site who tracks her around and feeds her back her private buying habits. Eventually, Alice is

forced to realise she cannot trust the system that is providing her the channels - and Alice buries her trust another metre deeper into her psyche.

Humans do not trust machines. As it happens, this is quite reasonable. Humans *rely* on good machines, because they are reliable. They deliver at the level of knowledge, or they are disposed of. When Alice asks the machine, it will always tell her the same thing: assuming the humans have done the right thing, the machine will do the right thing.

Machines can do data collection and prediction and all sorts of other things, but there emerges a natural line that machines should not cross, else the line in the sand is shifted. By Alice, to put the machine back in the box - of unfit machines.

Machines can't be Human. It follows that machines can't trust other machines. Machines can do protocols, run algorithms over data, run errands to other machines, high frequency trading, play Copycat against CopyKitten. Machines can lose all your money, but they cannot enter into contracts, themselves, nor take responsibility for a risky decision, nor fall in love.

A trust system isn't. Therefore, the so-called "trust systems" or "reputation systems" can only record metrics - hard data - that feed into a trust decision by a human. Systems cannot "trust" nor can they replace trust nor "be trusted" in the same way that humans are. The systems can move some of the calculations and collection that Alice used to do into her more convenient iPad, but what is left for her to do is still called trust, what she put into the machine she calls algorithms and data (cite: Identity is an Edge Protocol).

We could of course explore the edges of these claims. Could Alice trust the weather to deliver rain? Could an AI be built that would be as vaguely hopeful as Bob? Does Bob trust his dog, does Alice's cat trust Alice? For the sake of this essay at least, we leave those aside.

Alice's machine calculates, Alice trusts.

Automata in games can show us how to break trust, but they do not show us how to build trust. If you spent some time playing "The Evolution of Trust" you will realise that CopyCat and her friends break their opponent and win quite regularly within certain bounds. These are coded strategies to break the game of others and when they win against others, trust cannot grow. In effect, they are strategies to break trust or to enable trust with like players.

But above we say that Alice rebuilds her trust model after each loss, and hides it deeper each time to protect herself. The simple game theory algorithms are small, simple programs called finite state automata and they are not capable of holding the memory that is needed for Alice's trust game.

Economics of Trust

> *"But, what, you ask, do I do when someone defrauds me? The neat thing about using financial cryptography on public networks is that you can use the much cheaper early-industrial trust models that went away because you couldn't shove a paper bearer bond down a telegraph wire. In short, reputation becomes everything. Like J. Pierpont Morgan said 90 years ago, '...Character. I wouldn't buy anything from a man with no character if he offered me all the bonds in Christendom.' In a geodesic market, if someone commits fraud, everyone knows it. Instantly. And, something much worse than incarceration happens to that person. That person's reputation 'capital' disappears. They cease to exist financially. Financial cryptographers jokingly call it reputation capital punishment. :-). The miscreant has to start all over with a new digital signature, and have to pay through the nose until that signature's reputation's established. A very long and expensive process, as anyone who's gone bankrupt will testify to."*
> -- Robert Hettinga (1998) http://www.nikkei.co.jp/summit/98summit/english/online/emlasia3.html

Another observation that emerges is that trust is expensive - damned expensive. It behoves to consider trust through an economics lens - how is it expensive, and how can we deal with the costs? Because at a minimum, Alice is impelled to seek economies in her trust.

Trust is expensive! To keep doing trade with Bob, she has to be able to extract sufficient benefit from Bob that their relative benefits are in profit, even as they both attack each other's trust models.

Therefore, *Alice searches for economies*.

Trust requires continual testing - improvement of information and refreshing of the experience is needed. If Alice doesn't invest in the maintenance of trust, by refreshed information, it might not be there when she needs it. If Alice 'trusts' on too few data points, it is too easily gamed.

Alice has to construct a model of Bob's trustworthy behaviour, then test it at her own risk, and then keep testing and trusting.

Society. Worse, as we discover in the following section, Alice needs to work with, that is, achieve net benefit with, many people. So, she is compelled to build a trust relationship with many people.

Not only with Bob, Alice must duplicate this process with all her counterparties = ∑Xpensive.

Alice's model is not transitive. As a consequence of Alice's mindful analysis of Bob and his circumstances, and her need to hide the model deeply, any decision to trust Bob is not trivially shareable with another.

Facts might be transitive - the fact that Alice trusts Bob might be used by Carol to trust Bob, but only within limitations - does Carol know Alice's mind, did she share Alice's experiences, can she duplicate Alice's trust thoughts, can she incorporate the fact into her own trust model? No, she cannot.

If Alice trusts Bob to behave, this bounty does not necessarily pass on to Bob's friends. Now Alice has to look into Bob's mind and into the minds of Bob's friends. Assessing Bob's friends is a distinct trust decision to whether she trusts Bob himself.

What someone understands as a fact is, when transferred, just an opinion shared. Unless that transfer is backed up by an appropriate foundation, it lacks the ability to impress: Bob may claim his friends are cool and safe, but Alice might decide that Bob is unreliable on this point, and in order to permit his friends in the house, Alice may require Bob to up the ante - Bob must back up Alice's decision to trust Bob with some additional offer. Bob might underwrite the risk, cover the direct costs, or place other assets on the line - either way, if Bob wants Alice to accept a risk to her from Bob's erstwhile trustworthy mates, he might have to put something on the table to even the imbalance.

In the language of JP Morgan, Alice might accept Bob's character and Bob's bond on a transitive risk such as Bob's mates, but neither alone. This does not mean that trust is now transferrable, or even sellable; more, it means that Bob and Alice can construct an exotic derivative that appears profitable. Bob gives Alice an option on remedy for any shortfalls by Bob's mates. Indeed, relationships might be better off expressed in the language of finance, if only we could recognise the shared risk of Alice and Bob as opposed to the winner take all profit-seeking of the legal entities in finance.

Prisoner's Dilemma. Alice is beset by others looking for personal gain. The easiest personal gain is to steal from each other, but this is a net loss game - Alice loses more than the thief gains, so society loses value over time.

We can create more together, and live in a net-profit game, yet we remain subject to temptation of short term theft. This is the Prisoner's Dilemma, and game theory says that the solution that places us working together for shared gain is always another round, in which the greater part of the reward is always in the future.

Therefore, if trust is the accounting for our solution to the Prisoner's Dilemma - we come together, aim for the future, we build a future together, accepting shared risks made of shared downsides and upsides.

Trust prefers an equilibrium. In order to trust Bob she must interact with him, take risks on him, and balance the books. Unilateral trust is possible - think of rock stars, messiahs, presidential candidates and the new generation of Satoshii at the helm of each blockchain. But the more natural state of affairs for trust is an approximate equilibrium in which both invest in each other. A lot or a little, as its human nature to game each other's model a little or a lot, and both learn enough of each other to make the micro- or single-round trust decisions needed to proceed to the next round.

Trust is the potlatch of relationships: we both of us have to destroy good time and effort in order to win. Bilateral or mutual trust is therefore the first, easy economy - while I learn about you, you are learning about me.

Trust has economies of scale. Notwithstanding that the model cannot be easily shared, the expense of the process pushes us for a continual search for easier ways. Sharing of something is one way:

- Database sharing. Alice can share her information with Carol. She can relate anecdotes and experiences, or share her contacts list.
- Derivative trust. Alice can share her decisions - she can tell Carol that she trusts Bob. At its minimum, this is just more information for Carol's information base; at the maximum, it is a proxy decision for Carol - I trust Bob because Alice trusts Bob.
- Mutual Trust. Alice and Bob can work the trust game together - each cycle can be a cycle for both.

No matter the simplifications achieved here by Carol, to her economic benefit, she has still chosen where to place the dial - listen to Alice, copy the data or adopt a decision by proxy. Unfortunately, we have no way to be more precise where the dial is set, because we are faced with a recursive problem - Alice can only trust that Carol accepts what Alice says as said, as both of them hide their true trust model too deep.

Trust has diseconomies of scale. And then, it's almost an obviosity at this stage, that if Alice has to invest substantial amounts of time to keep Bob and Carol's trust, she'll quickly run out of capacity. True trust is limited to a very select group - that group you've invested substantial time into, and been invested by, for natural reasons outside the strict goal of acquiring trust.

Islands of Trust. If Alice can invest in Bob, and in Carol, can she invest in Bob and Carol at the same time? Larger groups of trust should emerge - if the components of trust can be shared economically, then the mechanisms of that sharing should encourage groups to form.

Alice wants to share - whether your theory of humanity is based on evolution of the defensibility of groups, specialisation or Maslow's pyramid of needs, humans want to belong and want to contribute. But it goes beyond that - Alice needs to share her identity with those she trusts because that's the only way that her life can move forward. The groups she trusts, those that she feels comfortable sharing her identity with, are then the next essential step in our journey.

Before we turn in the next section to what a group means to Alice, we should review the 'systems' that have to date presented a solution to her. If you are familiar with the 'trust business' or social networks or similar, you might like to skip to the Conclusion, or straight to Part III.

II.3 Critiques of so-called 'trust' systems

> *"It is nice to trust, but it's better not to."*
> Old Italian proverb, recorded by Steve Wilson, "Abandoning identity in favor of attributes," 2014

It should then be apparent that there is no such thing as a 'trust system' as, if it were, it would be broken by definition. Before we dive into what can exist (Part III), let's review what these systems refer to when they say they are in the trust business. Let's find the best trust that money can buy.

The antithesis of Trust

> *"Scientists are easier to fool than children."*
> James Randi

The best anti-Trust that money can buy is a certificate from a Certification Authority ("CA"). It breaches most all the above: There is no choice available to the user other than the intractable choice of not using the system; likewise, the decision to take a risk on a website is not made by Alice, as HTTPS is seamless and 2 decades of user experience and testing shows that users do not notice any switch to or from cleartext HTTP. Although the user takes a risk, few will recognise that she is indeed taking a risk, and none will know what that risk is; most insiders in the industry don't know what it is either. The context could be narrow or broad - nobody knows because although the certificate might be tight, the browser is not. Finally, her only foundation for taking that risk - the null decision of doing what she was told to by her bank or browser - is that nothing went wrong before now. Surely, if that is a foundation, it's the same foundation of risk that makes turkeys think that Christmas never comes.

About the only thing that is certain about the use of the term 'trust' in the CA business is that it isn't 'trust' and whatever it is, it has screwed with people's understandings of what trust is and means.

Reliance - a close cousin of Trust

There is a close cousin in the PKI ("public key infrastructure") world called *reliance*. In the CA/PKI concept, we can build up a structure of contracts, claims, and verifications such that the user *may rely on* the claims made. For example, the certificate includes the name of the holder, and thus the user may rely that this is the name of the certificate holder; it is said that the user then becomes the relying party.

Because each step is reliable, the hope is that the result is more deterministic than, say, trust. Reliance in this context is the same relationship we have with a machine. Our car will carry us on a journey to work this morning because that is everything about what it does, in a statement: its job is to carry us to work and everything is tuned to that objective.

Reliance aims at belief or knowledge. In contrast trust is a higher order human decision about taking a risk on some decision *where we cannot rely*. In this sense the evolution of an automated system is typically about creating more reliance; a re-interpretation of the CA's marketing might be that you do not need to trust because you can rely. You can believe, there is no risk.

Trustlessness

> *4/ If you (or your child) is 16 years old, Bitcoin has been around for more than half your life--- which is "good enough" for trust.*
> Spencer Bogart, Tweet.

In additional analogue, the same could be said of Bitcoin - it has built a system of reliance, in which, you do not need to trust people to keep it running. Hence, *trustless,* a term used by Bitcoiners, is actually a fairly good approximation of some parts of the technical system such as the incentive system promoting a probabilistic finality on consensus over a block. Yet, as with Alice and Bob, the algorithm might be trustless, but the algorithm's friends might not be.

As we saw in the hard fork stories for Ethereum and Bitcoin, *trustlessness* is not reliable, neither in cryptography nor in life. While the mathematics might be unchallengeable, you can still lose your money to a host of enemies. And now, with forks, you can double your tokens, which might sound good for those hoping to win double the money, but it's death to business which earns double the liabilities through forked smart contracts.

As systems, Bitcoin and Ethereum have pushed themselves back down from reliable knowledge down to risky trust. Indeed, the blockchain mechanics don't even reach the challenge of the old Russian proverb - trust, but verify! - as you yourself cannot verify the mathematics nor the system.

Try as you might, you cannot escape the essential law of life: that which we can automate safely, we do, for everything else there is governance. The experiences of the blockchains just cast into stark relief what happens when one single innovation simplifies the governance requirements - does a wild-eyed mystical technocrati class believe they've solved everything? Or can we recast the governance to serve the members, utilising the benefit of the new invention?

This is less to criticise Bitcoin than to criticise the weaknesses of the system. Its implementation of consensus over shared facts, or, "*I know that what you see is what I see"* is a thing of great worth, but reliance on the blockchain simply moves the questions of trust to a higher level: Alice and Bob do not need to trust the system for their tokens, but they do need to trust each other, in community, that the tokens are worth something. Whether Alice can trust Bob beyond a shared consensus on value has not been addressed.

Blockchain engineers do not wear the iron ring. Alice cannot rely on the engineers because she has little or no choice or sway over their actions, they act much like any state registry issuing a licence.

The failure of PKI Reliance

Back to PKI. Why then could not the PKI / CA be called a trustless system when Bitcoin could lay claim to the new title? It is because the reliance in PKI is over nothing, and is therefore not reliable. A relying party in PKI relies on a claim with zero value, because nothing backs it up when it goes wrong. In Bitcoin, we all rely on the coin being acceptable - a Bitcoin is worth 1BTC is worth that Bitcoin to all who agree it is worth a BTC. That works, because we all back it, and the blockchain keeps the numbers solid. If we didn't enter into that compact, it would be worthless, like PKI. But we did, so it's not.

In part this is because Bitcoin has a tight feedback loop that is re-proven many times; you rely that your payment will stick, and as an essential act of commerce, if it fails, you are screwed.

You run the payment again, you seek other recourse, you get your money back, or you exit the system. With feeling.

In the CA system you might want the name to be correct, or the connection or whatever, but whether it is or not, is not related directly, haptically, financially to an essential act of commerce. The name is so filtered, the connection so wrapped in other systems, the errors so hidden, that the real act is disconnected. If you're screwed, who do you blame?

Let's characterise this as Alice layering her trust over components of reliance. The machine provides reliance at the lower level, and that which the machine does not provide is kicked up to the higher, trust or governance layer. Trust is the exclusive domain of humans. As described above, experience is collected and processed, decisions are made, risks are taken; all acts taken by humans. Without these acts, repeated and repetitive, no trust is possible.

In the end, certificates are not *reliable*, but the system is *trustable* - the user trusts the browser every time she goes to a merchant site. With this comes an inherent risk of failure. Phishing opens up the weaknesses in the system and easily slices the browser's security model apart from the certificate model. As the weakness is in the browser, the result is that Alice and her browser manufacturer are in a trust relationship. The only shortfall of this model is recognition - that Alice is ignorant of who she is trusting, and the browser manufacturers deny that their brand as über-CA is what Alice is trusting.

From this point of view, the secure browsing mechanism is a trust relationship and this is a bad thing - we should have been able to build a system that delivered reliance for websites by now.

A system of real trust

> *bIlujlaHbe' chugh vaj bIQaplaHbe'*
> *If you cannot lose, you cannot win*
> Klingon proverb

A system that evolves for the benefit of the user might be said to be one that identifies slices of the user's trust process, automates those subsets, and pushes them into her reliance layer. Each evolutionary step pushes a little more down into the machine, leaving the bulk of thinking still to be done by humans. And, even if the new system is a radical improvement in Alice's lot, it seems that

she finds new ways in which to use and abuse the system. Perhaps there is an inner need to gamble, a cognitive limit or a learning block such as Dunbar's Number (discussed in Part III), a limit on her trust on the inner machine, that humans just need a certain balance of trust and untrust in order to function?

Can we conclude that a system of real trust should seek the above happy medium, to find some identifiable mechanism, to capture some subset of the activity, to deliver it safely and reliably wrapped in metrics? Once so captured, Alice our user can discover and decide for herself how to proceed, internalise, and work with the new information.

To do this, Alice needs to have available to her events - memories - of the past actions of Bob, and preferably ones close to or analogous to her current pending decision. Alice needs to live her prior life before she admits any new information. To do otherwise is a denial, a collapse, a fleeing of all that she knows; then all new trust must build on what she knows.

With this in hand, let's continue our review of what people call 'trust systems'. If you are impatient, skip to the Conclusion, or straight to Part III.

There are approximately two popular ways to look at so-called trust systems (sometimes called reputation systems): bottom-up grassroots or top-down hierarchical systems. The former is generally characterised as web-of-trust (WoT), the latter by the Identity Document (ID). In between these two poles are both corporate variants and local variants.

The Identity Document (ID) - Top-Down

> *"If you work at Goldman Sachs in New York City and you want to tie up a woman and then have sex with her, there's a good chance you'll first have to speak to Rita. She'll insist on calling your office, speaking to the switchboard operator, and being patched through to your desk. Then she will want to check out your profile on the company website and LinkedIn. She'll demand you send her message from your work email, and require a scan of either your passport or driver's license.*
>
> *And you will comply."*
>
> Allison Schrager, "Trust and Crime" http://qz.com/621994/trust-and-crime/

The popular view of trust starts from a top-down government-issued identity document ("ID") or its close sibling the Identity Number. In this concept, people can look at Alice's ID, and decide whether that's good enough to trust her.

The foundation of this one view on *trust in documents* is a little bit difficult tease out, but history can be found in the Napoleonic code countries in which citizens are required to register with their city office. If a name is known, then the person can be found, as the registry has their address, by law. Lawsuits filed in court can then be delivered reliably by the postal service, and in some sense, history can follow a person from city to city. In European tradition, it is considered bad (illegal?) to engage with others under a different name or use a different address.

Curiously, IDs are considered as worthy of trust in non-Napoleonic code countries, yet they do not feature the above assists. In the anglo-world, it is substantially more difficult and risky to *serve summons* from a court. And, trading under any name or at any address is considered more acceptable, as long as fraud is not entertained. You do not need permission or registration to move to a city, you do not need permission to marry in another country, records do not follow you. People are more free, for good and bad - large countries and free countries tend to hide more scammers.

One confusion with the state-issued ID is that the state itself has a substantial advantage in relying upon it, whereas the ID is less usually created and issued with the purpose of any other entity (other than another state) relying upon it; this weakness can be seen in the use of the social security numbers (SSNs) by American issuers of credit. Even though such was an explicitly unintended use by Congress, and SSNs are not reliable in any technical sense, many rely.

Both systems are weak when borders are involved. Indeed it could be that reliance on IDs by the public is actually reliance on the fabric of law and society within borders - in this sense, by custom we all operate to the standards of the society, the country, the city or *el barrio*. Recording ID could then simply be a ceremony to remind ourselves of this fabric.

Web of Trust (WoT) - Bottom-up

The web of trust was first popularised by PGP in the early 2000s, so let's describe that system. A public / private key pair is created by a user, and she distributes her key[3] by some means or other. When Alice meets someone at an event, she conducts some brief and due diligence over Bob, and then decides to sign his key, and return the signed key to him. Hopefully this is mutual.

In this way, Alice collects signatures on her key, as does everyone. When Carol comes across Alice's key, she'll feel a bit more comfortable that it is indeed Alice's key because it has some signing traffic. Better, if she can find one of the signatories who she has also signed, this means

[3] Public Key. When we say key without describing which key of the public-private pair, we generally mean the public key.

someone she knows and already checked out has verified Alice in like fashion. In this way a *web of trust* is constructed across many people.

Alice can also put on a rating on her friend's key of *do not / maybe do / usually do / always do* trust the key to verify the keys of others. In this way, if Bob has signed Carol's key, and Bob is rated as "*usually do trust*" then Bob's signature over Carol's key can be accepted by Alice, at least to that level of *usually*. Whatever that means.

So how does this compare with our version above of trust? The first thing that strikes out is, we're unclear on the meaning: what does "trust" mean in the PGP context? And this is unclear. While it is clear that Alice "trusts" Bob by her rating, what does she trust him for? It's not stated, and if we've learnt anything by now, it is that overly broad expectations must fail.

Typically we would expect a statement of what the signature means, but there is none in PGP doctrine, and this is both formal and deliberate. As a result, we get considerably distinct schools emerging with different meanings within the same web of trust. Some people treat it as an identity check and carefully make sure that the identity encoded into the key is matched by say a passport. Yet other people refuse to look at documents, and insist that it is the meeting that is attested to, not the name. In this way, people in the second group have keys signed as "Mickey Mouse" much to the annoyance of the first group.

Without at least some indication, PGP's web of trust fails to deliver any reliable information and therefore cannot scale as a vector of trust.

Certificates signed by CAs (PKI)

The alternate school of thought was popularised by a startup called Verisign, which to cut a complex story off at the knees convinced the first browser manufacturer Netscape to let it sign the keys of users of SSL, a protocol for securing the web.

Thus, emerged the first widely scaled *public key infrastructure* or PKI. Originally, PKI was popularised by telecommunications companies who were looking to deliver email to customers over the one-family-one-landline scenario of the late 20th century. Customers would dial up, download mail, then disconnect from the telco. They could then read *and verify* their emails at leisure.

This model never came to pass for a multitude of reasons, but the technology lived on through the 1980s and into the 1990s for any opportunity that presented itself. And so it was that SSL dawned as a pilot to secure ecommerce in the uncertain, halcyon and financially charged times of 1994.

Fast forward to now. The business model for SSL certification authorities (CAs) is complicated, tedious and deceptive. But in short it reduces to: the CA checks the first user's identity through standard techniques (ID) and then signs a claim of identity over the key. When (other) users are browsing, their software lets the signed certificates go through and badly signed ones are blocked. The user now knows she is communicating with who she expects to.

Unlike PGP's 'trust model', about which there is almost no writing and therefore no consistency, there is no shortage of literature on the CA PKI model. Unfortunately practically none of it is accessible and/or useful. However, some criticisms standout:

The statement over which the signature is made is fairly consistent across the industry. It is in this sense better than no statement such as PGP, but only just. Unfortunately, the statement is over the ID, which as we suggest above, is more of a proxy of hopeful behaviour rather than anything reliably strong, so the baseline value of the statement is surprisingly weak.

Secondly, because the browsers refuse to differentiate the certificates, they are all the same, resulting in a race to the bottom, which leads to certificate manufacturing. Hence the structure of the PKI model fairly universally eliminates the ability to rely upon it. These and other problems pretty completely eliminate any potential for the browsing PKI to ever deliver something on which Alice can trust, but for different reasons to those perverting PGP's efforts.

Western tradition of Trust - Courts

CAs will point out that you are free to take your grievance to the courts and indeed this is the trust system that underpins much of rich, western civilisation. Indeed - you can walk the streets after dark, you can rely on a contract, you can avoid being shot, raped, mugged or extorted in much of the OECD world because of the courts, the police and the integrity of the system.

To learn just how powerful these systems are, you need to leave. In the third world, none of the above are necessarily true. In Nairobi, where we spent a few years, walking after night was not safe, and we heard stories all the time. I walked at night, but I also wore a huge knife on my belt - both obvious and illegal, but the police there would simply steal it not charge me. Corruption is the purpose of a job in government.

In the middle east, honour killings are still commonplace. In much of the world, women are not allowed property rights. In Mexico, the real law is the drugs cartels, the police and the military having been long since corrupted by the USA's war on drugs. In impoverished countries sitting on rich resources, no courts seem to trouble the steady flow of left-leaning revolutionaries or right-leaning western corporations. In the countries I am familiar with, the baton of corruption is fueled by the steady stream of powers delivered by the OECD under the guise of anti-money laundering. What the rich countries don't realise is that power to police is power to corrupt.

Western courts also have two other burdens that make them difficult choices for reliance even for us in the rich white world - they are both too expensive where we find them, and they do not work well across borders. Much of Internet life is across borders, making resort to the courts an expensive hypothetical - until the UN opens a global court of petty sessions, we have no reliable recourse for the vast mass of worldwide interaction.

As a sort of tiny footnote to the history of trust in the Internet, this inability makes the PKI offering of the CAs a joke. But the far more important takeaway is that without the reliable courts system that we assume, **no system, no business, no community has an easy to use 'trust' foundation available to it**.

CAcert

Perhaps in rebellion against the untrustworthiness of the PKI industry, a community known as CAcert formed around a sort of do-it-yourself version called assurance. Members check each other's ID against names, but they also check a few other things: that the person is indeed a member, and has agreed to the user agreement [4]. In the user agreement, there is a clause for Arbitration which all agree to, and CAcert runs its own forum of arbitrators to resolve disputes. Notably, the Arbitrator has wide powers including the ability to fine or eject, and thus there is a recourse available to relying parties which is not practically available to browsing users over CAs [5].

Once members have collected enough assurance, they can then get their keys signed by CAcert's CA as above. CAcert also signs OpenPGP keys. Assurers have to be assured, and they have to pass a test. In this sense, CAcert's community is hierarchical with two layers, assurers and non-assurers. Within the assurer layer, it looks more like a web of trust. Outside, it looks more like a PKI, but there is no strong barrier to becoming an assurer.

CAcert is a notable exception to the WoT and PKI limited systems. It resolves in principle many of the weaknesses of the above models. The certificates are slightly clearer, and they can be relied upon at least in documentation. For structural reasons, the delivery falls down: as a CA, it is not itself trusted by the browser suppliers, for reasons outside scope of this paper.

What is more significant for our present purposes is that, even though the certificates within CAcert have failed to deliver much if any trust neither internally nor externally, the people within CAcert have been remarkably trustworthy and trusting. We within the community put this down to **Arbitration - if something goes wrong, there is a system to deal with the failure**. It is not

---

[4] It is worth noting as *disclosure*, the present author spent many years with CAcert auditing and designing some of the mechanisms so described.

[5] It is also worth noting that Identrust, a company that manages corporate PKIs for large companies such as banks, has also solved this problem with a similar solution: their corporate customers and employee/users are bound into a common arbitration framework in order to provide recourse.

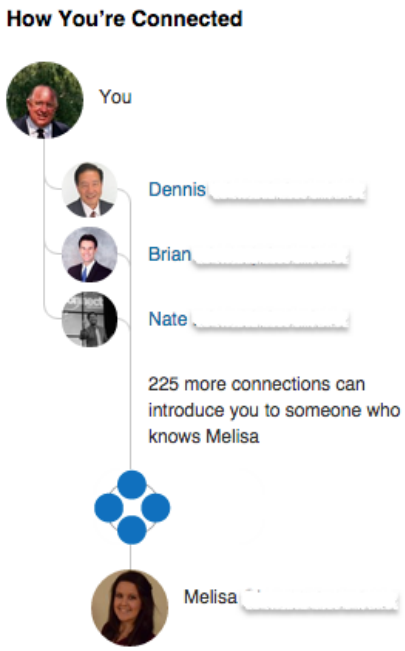
without its controversies, yet CAcert delivers trust - hundreds of resolved cases suggest that *recourse* or *dispute resolution* is both necessary and useful to maintaining trust.

Facebook

No description of trust would be complete without reference to Facebook. Their system is built on a combination of things, primarily the network of 'friends' by interlinking of users by the friending, commentary and photos, the combination of which is sometimes called *the social graph*. As the graph is combined of many weak links, it is relatively powerful in a Granovetter sense, at least strong enough for many companies to rely on it for low value transactions.

This social graph of weak links allows Facebook to avoid making much of a claim over links or the combination. A user finds out for herself who the other person is, using the many links or out of band confirmation. When Bob connects to Alice, he sees her photo and knows he is connected to Alice.

Facebook creates a pretty good social graph that was originally optimised towards the needs of the US campus females. As the film *The Social Network* lays out albeit opaquely behind a brilliantly deceptive script, the expression of the male-female relationship is captured in such a way as to give the female the substantial protection of distance and transparency - *"Facebook me!"* - reversing the imbalance previously trialled in the earlier experiment of Facemash. It so happens that males are not unduly disadvantaged by this system, and thus there is a stable equilibrium that fits the needs of both sides of the gender divide. Although narrowly constructed within the narrow domain of WEIRD (western, educated, industrialised, rich, developed) campus undergrads, the system also worked for wider, greater parts of society, and the rest is history.



Facebook is by far the most successful of these systems, both from a metrics and an identity point of view. It is notable that Facebook distinguishes itself from other systems by (a) collecting many weak links, (b) not making much or any claim about them, and helping those that are linked to talk. In this way, **it helps Alice with her life by making itself her memory**, but it stops short of directly helping her make trust decisions. Such a thing sits in marked contrast to competitors which try and fail to provide 'trust' information such as "see who your connections are connected to…" and "add a capability…"

Summary of 'trust' systems

Looking at the above, we can do all sorts of 2x2 graphs and views and venn diagrams and polemics. But let's just look at one thing: identity. The state model delivers an ID which hints at some limited view over identity. But at their core, ID relies on and is perhaps a proxy for local custom, so it does little good across borders and even less good online. Worse, applying these models to the developing world is just rubbing more salt in the wound - KYC/AML systems that demand ID are corrupted as channels to know your victim, extort him and money launder the profits.

The CA likewise delivers a certificate claiming some check over ID. Bank systems follow a compliance model delivered to them by far off bureaucrats without a clear meaning and without clear delivery of their stated mission, but like any religion, it is not necessary to understand and prove the model, belief is sufficient. The commercial companies selling trust similarly provide some basic tools then try and convince you that it will all work, as long as you trust them all.

None of these speak to identity.

CAcert created a system of identity pillared on protocols of assurance backed up by dispute resolution. Your Facebook Identity is the combination of weak links or data presented to another (name, photo but also friends in common and commentary on shared context). These speak to identity, but like the thesis suggested above, they are the weakest evolutions of reliance, the minimalist improvements on the machine. If our fate is to wait for these small improvements, then our fate is to be patient in extreme.

II.4 Where are we on the 'trust' thing?

It is somewhat clear that this whole 'trust thing' is far more complicated than any venture capitalist would have the patience for. Let's summarise.

- Alice trusts Bob:.
    - Each trust decision is particularly focussed on circumstances,
    - that builds from past information, and adds for the future,
    - To create a model in Alice's mind.
    - Trust then is both a momentary decision as well as repeated game.
- To define trust is risky, and indeed
    - Alice's model of trust is deeply hidden and personal,
    - we cannot model it nor interpose it.
    - If we could model it, we could break it. If we could break it, we will break it.
    - Trust is Heisenbergian - We can know it is there, but not what it is.
- Only people trust;

    - machines deliver reliance.
    - Machines are to deliver reliance over something worthy of relying upon.
- Alice's trust is economic:
    - Trust requires continual investment of a costly type,
    - Which means the protocol has to deliver benefits back to Alice,
    - And thus Alice and Bob are naturally led to an equilibrium of trust.
    - It fails to be easily transitive - We can share data, or opinions, but not trust.
    - Which leads to two economies of scale - sharing and mutual reward.
- Most systems labelled as trust aren't up to the job:
    - State IDs deliver reliance to the state - state ID is not intended for you.
    - Courts deliver for local (western) communities, but are broken for developing world & Internet.
    - CAs failed because they said nothing of value over documents that weren't meant for you.
    - Web of Trust failed because it didn't say anything.
- Some notable successes:
    - CAcert delivered trust internally through
        - Assurance over own members, and
        - arbitration - its own courts
    - Facebook facilitated trust between people through
        - many weak links - much memory.
        - less judgement - the machine that judges or sells is not worth our trust.

That all said - what can we say positively? The big 'trust' thing is that trust is in Alice's mind, and it is an integral part of her Identity. But it is also expensive, and economising means Alice wants to share the cost by sharing her identity, while maintaining the faith of Alice's defensive trust model.

Alice's personal network of trust is then a group. In order to understand Alice's Group, we turn to *Part III - On the proper upbringing of a young lady named Alice*.

Endnote. This document has received useful criticism from Ada Lovelace, Tatsu and Trishina.
Todo. Review and Integrate "Metaphysics of Trust - A Propaedeutic" by Michaël Suurendonk.