UNLIKELY OUTCOMES? A DISTRIBUTED DISCUSSION ON THE PROSPECTS AND PROMISE OF DECENTRALIZED PERSONAL DATA ARCHITECTURES

/

SOLON BAROCAS,
SEDA GÜRSES,
ARVIND NARAYANAN
AND VINCENT TOUBIANA

PRIVACY DECENTRALIZED
PROJECTS SOCIAL USERS
CENTRALIZED SYSTEMS
DECENTRALIZATION NETWORKS
FACEBOOK DATA ARCHITECTURES

Introduction by Seda Gürses

Different communities see a wealth of social, legal, and political promise in decentralized architectures – architectures that are increasingly at the heart of debates among researchers, developers, artists, activists, and private enterprise alike. Much of this stems from a deeply held belief that decentralization has the potential to precipitate radical restructurings of power, even though questions remain whether a change in architecture can, on its own, address sumptuous issues like privacy, autonomy, and other fundamental freedoms on the internet, let alone the struggle for a just society.

There is a great diversity of opinion regarding how decentralization can address the problems that arise from the accumulation of power on the part of service providers like Facebook and Google. Some want to increase the number of service providers in the online commercial environment from which users can then choose (a goal shared by many in the Vendor Relationship Management [VRM] community); others want to create alternative networks that subvert centralized power, both politically and technically, and empower democratic participation (an objective shared by some of the more politically minded projects like Lorea and RiseUp).

The Unlike Us conference in Amsterdam in March 2012 brought to light some of the explorations of decentralization currently underway. It opened up a discussion that cut across many different communities: initiators of privacy-friendly decentralized architectures were able to share their work and, together with participants from various disciplines, engage in a critical dialogue about centralized and decentralized networks. The participants elaborated on the political economy and socio-political aspects of developing and participating in 'alternative' networks. The discussion thrived also because certain technical, economic, and value assumptions seemed to find consensus among the discussants.

Almost in parallel, and separate from the conference, a group of scholars presented a paper, 'A Critical Look at Decentralized Personal Data Architectures', that put under scrutiny many of these assumptions. The authors, Arvind Narayanan, Solon Barocas, Vincent Toubiana, Helen Nissenbaum and Dan Boneh, offer a historically informed assessment of a whole variety of decentralized systems, ranging from so-called 'infomediaries' to federated and distributed social networks. They argue that, despite much work and many different efforts, none of these projects have achieved widespread adoption and, so far, have not provided many of the supposed benefits of decentralization. In a sense, the authors argue that few of the values routinely associated with decentralization actually inhere in the architecture and that adopting a decentralized architecture in and of itself does not solve many of the problems that its proponents aspire to address. They explain that this is especially true given certain economic dynamics and the state of technology. The paper then goes on to describe some drawbacks of the current proposals for decentralization and explores some of the technical and cognitive factors that limit their likely success. The authors also tackle the issues of open standards and interoperability, finding a number of problems there, too. Altogether, the paper presents a series of critiques of both the viability and efficacy of some of the current proposals for decentralized alternatives. Surprisingly, the authors shy away from a political analysis, focusing, instead, on (market) economic, social, and technical matters.

Had this paper been presented at Unlike Us Amsterdam, I believe it would have added interesting layers to the debates that took place at and around the conference. To make up for this missed opportunity, I proposed a collaborative process: engaging members of these communities to see what they make of the paper, with the aim of drawing on their feedback to help structure and inform an interview with the authors. The objective of the exercise was to hone in on points of both agreement and disagreement and to connect the apparent attraction of decentralization to a deeper appreciation of its trade-offs, its practical viability, its promise for preserving privacy, and its overall emancipatory potential.

The collaborative process worked as follows: we invited people engaged in ‘alternative’ social networks, peer-to-peer research, identity management ecosystems, as well as interdisciplinary critical thinkers to respond to the paper written. Of our respondents, the following agreed to also publish their responses: SpiderAlex, Jula, and Hellekin Wolf of Lorea, Floren Cabello of GlobalSquare, and Elijah of RiseUp; Jan Schallaböck, one of the co-organizers of the W3C event on the Federated Social Web that took place in Berlin in June 2011, and Antonio Tapiador from Social Stream, a participant of the same event; Sonja Buchegger, Benjamin Greschbach, and George Danezis, all of whom are involved in research on peer-to-peer social networks; Michael Herrmann, Günes Acar, Leandro Doctors, and Ero Balsa from COSIC and from the SPION project. Further, from the VRM world, or the identity management ecosystem as it is often called in the context of EU Projects, we received responses from Jaap Kuipers, Markus Sabadello (Project Danube) and Scott J. David (UW). Finally, Nicolas Maleve from Constant VZW, the feminist media and arts collective based in Brussels, responded to our request to contribute.

Based on the responses we systematically extracted prominent themes that then became the basis of the questions of a written interview with three of the paper’s authors. Once the interview was written, we shared it with all the respondents and asked whether they would be willing to publish their responses along with the interview. While a short version of the interview is included below, a full version of the interview, together with the responses, are available online.[1]

[1] Available online through the Unlike Us webpage: networkcultures.org/unlikeus.

As the short description of our collaborative process makes evident, there are many parties in the world of decentralized architectures that we did not reach. We certainly did not address all the valuable issues that were raised by those who responded. And, surely we did not cover all that is at stake. Still, we sincerely hope that this contribution is as valuable as we aspired for it to be: another step in the ongoing and crucial discussions on the potential of decentralized privacy-friendly architectures. None of this would have been possible without those who responded to our request for comments, to whom we express our deep gratitude for their openness, trust in our process, and brilliant responses.

Seda*:* Tell me how you developed an interest in what you call decentralized personal data architectures and why you decided to write a paper assessing their prospects.

Solon*:* Arvind, Vincent, and I worked together on Adnostic, which showed that client-side profiling could substitute for extensive third party tracking in helping to target online advertising. Decentralization had played a crucial role in Adnostic; it meant that the tracking of users could be delegated to the browser itself, cutting out entirely the third parties that normally have to collect and centrally store user data. But we were somehow less convinced that decentralization could solve all the problems to which the architecture is now enthusiastically applied.

As part of a separate exploration of the various technical initiatives to make privacy policies more intelligible, I became aware of so-called vendor relationship management (VRM)[2] and I was surprised that none of the associated projects referenced the Platform for Privacy Preferences (P3P).[3] I had wanted to build a platform to distill privacy policies into more digestible units, but discovered that many past and current projects did much the same.[4] Realizing that there was a long history of stalled efforts, I thought it would be interesting and valuable to perform a survey.

At the same time, Arvind and Vincent had both become interested in decentralized social networks (partially, I think, because of the excitement around Diaspora). In discussing whether decentralized alternatives could really challenge the entrenched players (especially Facebook), we realized that there were some unexpected and interest-

[2] 'VRM is the customer-side counterpart of CRM […] VRM tools provide customers with the means to bear their side of the relationship burden. They relieve CRM of the perceived need to "capture", "acquire", "lock in", "manage", and otherwise employ the language and thinking of slave-owners when dealing with customers'. More specifically, VRM aspires to '[p]rovide tools for individuals to manage relationships with organizations', '[m]ake individuals the collection centers for their own data', '[g]ive individuals the ability to share data selectively', '[g]ive individuals the ability to control how their data is used by others', and '[g]ive individuals the ability to assert their own terms of service'. (Project VRM wiki contributors, 'Main Page', 4 September 2012, http://cyber.law.harvard.edu/projectvrm/?title=Main_Page&oldid=5867, accessed 1 December 2012).

[3] 'The Platform for Privacy Preferences Project (P3P) enables Websites to express their privacy practices in a standard format that can be retrieved automatically and interpreted easily by user agents. P3P user agents will allow users to be informed of site practices (in both machine- and human-readable formats) and to automate decision-making based on these practices when appropriate. Thus users need not read the privacy policies at every site they visit'. ('Platform for Privacy Preferences (P3P) Project: Enabling Smarter Privacy Tools for the Web', W3C, 20 November 2007, http://www.w3.org/P3P/).

[4] See, Parsing Privacy Policies, http://solon.barocas.org/?page_id=200.

ing overlaps to explore, specifically the common commitment to decentralization and personal data stores. We thought there might be important lessons to learn from these earlier projects, and we hoped to identify some recurring stumbling blocks.

Vincent: In my recollection, it all started with a discussion at a coffee shop between Arvind and Solon about whether or not Facebook was a monopoly. At that time, Google was under scrutiny for monopolistic practices in Europe. I think we debated whether Facebook was in a more monopolistic position than Google due to the network effects that prevented any new actor from becoming a realistic competitive threat. From there we started to briefly discuss plausible alternatives.

I was interested in this topic as I was quite enthusiastic about Diaspora when it started. I was at New York University (NYU) at that time and I remember several discussions during NYU's Privacy Research Group[5] meetings and in Helen Nissenbaum's class about how this could be a game changer and that a viable alternative to Facebook would discourage Facebook from changing its default settings again... I started to sketch in a slide deck how a caching-based system could smooth the transition from Facebook to a distributed system.

It was interesting to see that some communities reacted to Diaspora by saying that they had been working on this for months or years already. It revealed that several groups had followed the same path and that they had not yet taken off. Although Decentralized Online Social Networks (DOSN) were praised almost everywhere, some people started to criticize Diaspora – mostly because it got (involuntarily) hyped and overfunded. Maybe the project's initial objectives were overly ambitious, but the architecture itself seemed quite resilient in the face of critique. Beyond saying 'this is unrealistic', I thought reviewing common pitfalls would be a good thing.

Arvind*:* For some time I'd wanted to do a survey on a privacy-related topic, especially because of the number of different communities in which privacy is studied and the diversity of approaches they use. Also, I seem to be drawn to ideas that have been reinvented over and over again but don't quite succeed. I like to understand what the hidden pitfalls are. I feel this is a type of work for which academics are well-suited – taking the long view, looking for a common framework for understanding a variety of related projects, and examining them without a strong vested interest in the outcome.

My interest in decentralized social networking apparently dates to 2009, as I just discovered by digging through my archives. I'd signed up to give a talk on pitfalls of social networking privacy at a Stanford workshop,[6] and while preparing for it I discovered the rich academic literature and the various hobbyist efforts in the decentralized model. My slides from that talk seem to anticipate several of the points we made about decentralized social networking in the paper (albeit in bullet-point form), along with the conclusion that they were 'unlikely to disrupt walled gardens'. Funnily enough, I'd completely forgotten about having given this talk when we were writing the paper.

[5] See, Privacy Research Group, Information Law Institute, New York University School of Law, http://www.law.nyu.edu/centers/ili/privacy_research_group/index.htm.

[6] See, Social Network Security Workshop, Stanford University, 11 September 2009, http://crypto.stanford.edu/socialnetsec/.

Seda: Can you draw out the relationship you see between decentralization and privacy more explicitly? How does decentralization address privacy, and what does privacy mean in the various projects that you surveyed?

Arvind*:* Decentralization is inextricably tied to privacy in two ways. The first is that the 'threat model' of privacy differs in the two types of architectures. In a centralized system, surveillance is almost inevitably an important potential threat. In a more decentralized system, client-side software security and theft of devices are much more relevant, especially when data is stored on users' devices. The second is that the types of levers available for protecting privacy are different. Centralized systems are easier for governments to regulate. danah boyd has argued that 'Facebook is a utility; utilities get regulated'.[7] The levers in a decentralized system are more technological in nature and in my opinion may require more user vigilance, expertise, and effort.

Solon*:* I think there's a reasonably clear dividing line between those projects that adopt a model of privacy as confidentiality and those that adopt a model of privacy as control. Some aim to return a degree of opacity to people's lives by shielding their activities from view, while others try to give people the means to more actively decide when and with whom to share information about themselves. Still others aspire to do a bit of both. But this one line separates out what are really the two main and very different types of goals of these decentralization projects: to provide people with a means of escape from the existing commercial offerings or to empower individuals in their ongoing commercial dealings. A more general – and perhaps better – way to think of this is in terms of each project's relative suspicion of institutions: where one group wants to avoid the need for people to even rely on institutions, the other wants to improve how people interact with them. These commitments map quite nicely onto the confidentiality and control paradigms of privacy.

Unsurprisingly, privacy as confidentiality tends to resonate with those who hold more libertarian views, ensuring that the actions of individuals are shielded from outside (i.e, governmental) view. But this understanding of privacy also holds significant purchase with people who subscribe to various forms of collectivism, insofar as it enables communities to self-organize outside (the purview of) the state. Decentralization is key to this strange affinity because it seems to permit political associations and solidarity in the absence of centralized political institutions. And – more to the point – it allows these associations to remain confidential with respect to the existing institutions of power. Think of the crypto-anarchist movements, for instance: while some groups lean more heavily toward the radical individualism of libertarianism, others are more excited about the prospect of voluntary collectivism.

Talk of control gets a bit confusing in this case because, although many of the more politically motivated projects put user control first, they do so with a focus on administrative control over the platform rather than the practical control mechanisms that individuals would use to manage their privacy. The reverse is true in most of the commercial initiatives: control is a matter of how well individuals can look after their own data and rarely about administrative authority over the platform.

[7] danah boyd, 'Facebook is a Utility; Utilities Get Regulated', *Apophenia*, 15 May 2010, http://www.zephoria.org/thoughts/archives/2010/05/15/facebook-is-a-utility-utilities-get-regulated.html.

Vincent*:* While many DOSN have as a key objective to prevent information collection by a centralized entity, they also aim to help users easily transfer their data to prevent the establishment of walled gardens. Which is why, when we set off on this project, we were also interested in how monopoly interacts with privacy concerns, particularly how increased market share translates into more comprehensive tracking and data collection at the same time that it constrains new entrants and the move to alternative platforms.

When Facebook changed its default settings many people wanted to leave, but found that they had no obvious place to go. There were a few alternatives, but people had to do more than choose the social network that they wanted to join – they also had to make sure that their friends joined the same network. Unable to arrive at a common choice, users would have been scattered in different networks and, because of a lack of interoperability, they would have to maintain accounts on each network. People also realized that any centralized social network could follow Facebook's path once it captured enough users. Thus DOSN seemed like a very attractive alternative.

But as we started to dig into the existing projects, we noticed that a few projects focused mostly on the threat imposed by centralized entities and ignored other threats. For example, Arvind pointed out really early on that spam was a major issue (more on this later). The threat model that one has to consider when building a distributed social network becomes very complex. And then it becomes even more complicated when you start to consider human relationships: friends and acquaintances with whom you share data might become the entity collecting your information.[8]

Seda: Respondents repeatedly remarked that your analysis seemed to gloss over some crucial technical distinctions. For instance, some projects adopt a federated architecture, while others aim to be fully peer-to-peer. Certain arrangements involve single sign-on providers; others introduce a more general identity infrastructure. Other respondents pointed out the important differences in the way projects conceive of those who would use them: are they consumers, users, data subjects, or citizens – or even non-citizens? And then there's the commercial/non-commercial divide, a distinction that you do not make even though respondents frequently drew this contrast. Some even resisted the idea that business models should be at the core of their development efforts. What additional points of contrast would you add to your assessment in light of this feedback? How would attention to these details change your criteria for evaluating the success of these projects?

Arvind*:* There are great differences in decentralization projects. Some have broad adoption as a goal and others don't. Still, many of these systems target markets with very strong network effects, and as such it isn't clear how feasible it would be to serve a niche. Similarly, if we examine the VRM vision, we are talking about nothing less than a paradigm shift in the way we do commerce, requiring major buy-in from retail giants

[8] Benjamin Greschbach, Gunnar Kreitz and Sonja Buchegger, 'The Devil is in the Metadata – New Privacy Challenges in Decentralised Online Social Networks', Fourth International Workshop on SECurity and SOCial Networking, School of Computer Science and Communication, Stockholm, Sweden, 19 March 2012, http://www.csc.kth.se/~bgre/pub/GreschbachKB12_MetadataPrivacyDecentralisedOnlineSocialNetworks.pdf.

and restructuring of entire industry segments. So while some of these projects might be able to declare success even without mainstream adoption, I don't think that holds for the majority of the systems studied.

To me, the distinction between commercial and non-commercial projects is not so salient. The need for a business model is simply a fact of the world. A project might envision funding all their development, hosting, marketing, and other costs via volunteer donations, but that is also a business model! The type of claims that we make – i.e., that centralized systems benefit from economies of scale and have lower overall costs – are indifferent to whether a project is commercial or not.

Solon*:* It seems to me that earlier infomediaries (like Lumeria) and many of the more recent projects (commercial and non-commercial alike) attempt to deal with privacy in much the same way as P3P: by providing individuals with tools to better exercise choices regarding the disclosure of personal information to counterparties. P3P envisioned a world in which machine readable privacy policies would allow users to delegate the process of rendering these decisions to local 'user agents'; infomediaries positioned themselves as agents that could take on many of the same responsibilities, only remotely. A good deal of writing at the time made this explicit link between what some called 'negotiated privacy techniques' and infomediaries. That said, Lumeria – like other infomediaries then and now – had grander ambitions and actively sought to extend these techniques to all facets of online life by allowing users to pool their data in one location over which they would retain exclusive control.

For all their differences, FreedomBox and other initiatives that seek to extricate users entirely from commercial services, share this same interest in architectures that allow users to host all their own information. The point being that, in both cases, a personal data store is a prerequisite for meaningful control. But unlike infomediaries, FreedomBox and its ilk do not aim to facilitate more informed or more granular disclosure; they exist to avoid the need to disclose *any* information to third parties whatsoever. This is a crucial difference. I would still however stress that there's something quite remarkable about the fact that these communities both seize upon personal data stores, even though they adopt them for pretty much opposing purposes: one sees them as empowering consumers to more effectively engage with other market participants, while the other understands them as a way to avoid the need to turn to the market at all. The same architecture can accommodate very different politics.

With this in mind, it seems fair to ask whether it's right to even attempt to evaluate these projects on the same terms. This relates, I think, to the question about who these projects address and what it means to think of everyone who might partake of them as 'users'. We didn't set out to assess the size of these projects' user-base. We wanted, instead, to ask whether these projects had good chances of achieving their design goals and, further, whether the proposed (or existing) features were likely to produce the outcomes expected by their developers. As Arvind already remarked, however, attracting and actively engaging a large number of people is often crucial to the success of many projects, even on their own terms. But this is not always true, especially where the point of the project is not to compete with or complement the existing commercial platforms. Some projects propose to do something quite different: not vie for generic 'users', but rather serve the needs and interests of particular communities.

Seda: Are your criticisms peculiar to decentralized architectures? As many of the respondents pointed out, all systems have to grapple with scalability, reliability, and consistency. And hardware backdoors, security limitations, and the lack of refined access controls are by no means unique to these projects. Further, the P2P community faces problems that stem from limitations imposed by internet service providers and governments at the network layer. Is it sensible to talk about the challenges confronted by communities who want to adopt decentralized or distributed architectures only at the application layer?

Solon*:* These kinds of responses to our paper have been disheartening. We never intended to criticize decentralization as such. The point of our exercise was to show that many of the problems that we commonly ascribe to centralization actually carry over to decentralized systems – that these problems can affect *both* architectures. When some respondents pointed out that centralized services suffer the same problems, they were actually arguing our point: neither architecture easily remedies these problems or escapes them entirely. Some of the very serious problems we have with online privacy, for example, are not simply a function of the underlying architecture. Rather than solving these problems, decentralization might make them more difficult to address, both in theory and practice. Just think about the challenges in trying to ensure the appropriate flow of information between users on the same DOSN (the problem of 'lateral privacy', as it has become known) or the selective sharing of information in VRM (where people are burdened with the same – actually more – choices about what information they are willing to reveal to others).

Vincent*:* The problems that exist in centralized systems could be more severe in decentralized networks. Currently, in centralized architectures, the weakest element is probably on the client side: the browser or computer that is used to access a social network, for example. Even when a client is corrupted, safeguards can be implemented on the centralized system to prevent access to resources when suspicious behavior is detected (for instance, when a user starts downloading all his pictures and the pictures of his friends). If the computer that serves as a host in a distributed system is corrupted, these safeguards will be ineffective; in the worst case the attack could then be distributed to other peers.

The complexity and openness of federated systems in general make some problems harder to address. For instance, access control policies have to be supported across different systems. But I don't know if it's possible to translate 'provide access to friends of my friends' if I just know the system that my friends are using, but not the system that their friends are using*.* For example, in a federated system, your privacy is not subject to the nudges implemented by your system but to those implemented by the system that your friends use. In my opinion, this could create some serious misunderstanding, as an action that is discouraged in your network may not be discouraged in another network. For nudges to be globally effective, I think that designers would have to make sure that they enforce the same norms, but then they would lose some openness: federated systems that are not compliant would not be part of the federation. I think a good example is the warning[9] that is displayed on

[9] See, 'Google+ and Privacy: A Roundup', *33 Bits of Entropy*, 3 July 2011, http://33bits.org/2011/07/03/google-and-privacy-a-roundup/.

Google+ when I want to share a 'limited' (i.e., not public) post. If I post something on Google+, I can expect that my friends will have the same warning if they want to share it... But in a federated social network, I cannot hold the same expectations. And I don't think decentralized and federated systems can easily address these issues. Developers will have to engage in far more cooperation and dialogue, which would be great if it results in some standardization around nudges that increases respect for privacy norms.

Spam is another serious threat to federated systems as such systems would have to trust each other to filter outgoing spam. Blocking incoming spam would also be far less efficient. Consider the experience of email, which also relies on a federated architecture: when an email provider fails to block outgoing spam, there is a risk that others will blacklist it. One of the key advantages of a federated system – that any new interoperable actor can join the system – also poses the risk that hosts dedicated to spam could constantly re-emerge and pollute the network.

Arvind*:* We really had two different kinds of drawbacks in mind: some that we claim are specific to decentralized architectures, and some that have been claimed to be *advantages* of decentralized architectures but which we contend are drawbacks shared by both types of systems. Sadly we didn't do a good job of separating the two.

To pick up on Vincent's example, nudges are easier to implement and work better in centralized systems because they can impose a standardized user interface. They work best when they elicit a predictable behavior in typical users. Privicons,[10] a project I like a lot, is a good example of nudges in a decentralized medium: users who adopt the browser extension can signal to the recipients of their email how they would like the email to be handled (i.e., 'Delete after reading/X days'). But this only works if the person receiving the email has adopted the extension, too. And because email is federated, making this a general standard would require buy-in from *all* email vendors and efforts to explain to users how the feature works. Compare this to the Google+ nudge that Vincent described.

Seda: The projects that you've pulled together in your review employ very different tactics to enforce norms around the appropriate flow of information. Some adopt access control, others rely on legal contracts, while still others resist any such limits, seeing them akin to DRM. What do you see as the feasibility, advantages, and limitations of these approaches to instantiating 'user control'?

Solon*:* Consent still reigns supreme, I think. None of these projects adopt a more substantive guiding principle about the appropriate flow of information; they instead try to put users in a better position to understand and determine how information moves. And this – with very few exceptions – means implementing a more sophisticated set of choices. Access control allows users to implement these choices as persistent rules; the ability to draft contracts that specify the terms of exchange and use aim at something very similar. Allowing people to 'own' their data does nothing to ameliorate this situation; it just substitutes contract negotiations for the reading of privacy policies (and, very likely, one form of legalese for another). In all circumstances, the person

[10] See, www.privicons.org.

that developers have in mind when they build these systems is still a (hopefully more well-informed) rational actor with their own idiosyncratic tastes for privacy. But a more substantive notion of appropriateness, based on a sense that privacy serves a social value that exceeds the interests of the individual, is rarely baked into the design of the platform. Efforts to enforce choices through hard-coded use limitations that resemble DRM should not be confused for such principled design. As commonly proposed, they are adjuncts to notice and choice, not mechanisms to guide the movement of information according to contextual norms.

Vincent*:* My experience of access control lists in social networks is that they can be used in two different ways: 1) to limit the audience of my post (selective broadcast) and 2) to protect my privacy (control reading access on my Wall). Note that, in the latter case, the effect is quite limited because shared data is hard to control in practice. Unfortunately, when you see a post in your NewsFeed, there is no associated context so you have to guess the sender's intent (selective broadcast or privacy). Obviously, these critiques also apply to centralized networks...

Seda: What we think of as centralized services often employ various forms of decentralization, from the use of open source software, to the move toward distributed computing, to the leveraging of activities performed by consumers. What is the appropriate way to conceive of the relationship between decentralized architectures and administrative control? And, more generally, to what extent would companies actually like to delegate certain tasks or responsibilities to their customers?

Arvind*:* The distinction between architectural and administrative (de)centralization is an easy one to overlook. Now I don't think a system like Facebook is decentralized in any practically meaningful way. Sure, they might use PHP or Linux, but I don't see the underlying programming environment as having much to do with our analysis, as we're more concerned with the architectural decisions that affect personal data and user experience. Similarly, the rise of app platforms and user-generated content are all examples of successful 'decentralization' in a broad sense but not in a sense that is of concern to us. It is important to make the distinction here between delegating tasks and delegating control.

Vincent*:* Some centralized services have been designed to run on a centralized architecture and – to handle their rapid growth and localization – have distributed some of their components. The fact that these systems are controlled by a centralized entity alleviates most of the issues inherent to distributed architectures: redundancy is controlled, reliability is assured to a certain point, interoperability is not an issue... From a user perspective, these systems behave as if they were perfectly centralized. For distributed systems to replicate the same efficiency in managing distributed architecture, some parts of the system would have to be centralized or hierarchical and some constraints would have to be imposed on users (i.e., shared storage, on/off time, and connection quality).[11]

[11] As one of the respondents mentioned, there have been few distributed systems that impose shared storage constraints on users (i.e., Darknet/Freenet), which means that users would have to deal with legal and moral issues of hosting other people's content.

Arvind*:* Google (with OpenSocial, Google+ etc.) and Twitter are examples of companies that have made forays into, started with, or promised a limited level of decentralization. However, most such services seem to eventually turn away from that direction – the trend lately has in fact been to exert increasing control over APIs. Decentralization is possibly good for society, with some caveats, but probably not good for business.

Vincent*:* Still, things could work in the other direction: operators of centralized systems may be motivated to extend their architecture with devices owned by the users. Companies could delegate the management of personally identifiable information to users and just link profiles to pseudonyms so that they could still conduct business while feeling less pressure from Data Protection Authorities and other privacy regulators. Critical pieces of information could be stored by the user, and the centralized system could host perturbed, non-personal data. For instance, only thumbnails and inaccurate records (a truncated name and birth date and a perturbed social graph) could be hosted on the centralized system while providing a pointer to the system hosting the accurate records (for instance, the cell phone or the set-top box of the end-user). An idea along these lines was developed in Polaris,[12] but the objective of this project was to enforce users' privacy with only little consideration for service providers' preoccupations with respect to existing privacy regulations.

Solon*:* Efforts to devolve certain responsibilities to individuals (that might seem like instances of meaningful decentralization) can still serve the interest of the entrenched players. Enrolling people in the management of their own data shifts the burden of ensuring privacy from the platform to its users. This relieves institutions of the responsibility to look after their stakeholders in the name of empowering them. That VRM can present itself as a win-win situation for individuals and institutions alike indicates just how non-threatening decentralization can be. And although improved interoperability and data portability often figure in decentralization projects, decentralization as such does not insist on these features.

Seda: Respondents pointed out that there are a variety of ways in which decentralized and distributed architectures can serve the interests of privacy: they mitigate the risk of leaking users' data, reduce opportunities for profiling, and curb function creep. The limits that decentralization places on aggregation, analysis, and secondary use are features, not bugs. But you point out that these come with significant downsides, too. Can cryptography help to avoid some of these trade-offs that you highlight (e.g., spam)?

Arvind*:* We shouldn't confuse two unrelated things: security breaches are unequivocally bad, whereas secondary use arguably benefits society, even if individuals don't like it. It's an interesting empirical question how the risk from server-side data breaches compares to the risk from (say) lost or stolen personal devices. As for the privacy benefits of limiting secondary uses that happen in the centralized model, that's the raison d'être of many of these systems. The reason that this is also a *disad-*

[12] See, Christo Wilson, Troy Steinbauer, Gang Wang, Alessandra Sala, Haitao Zheng and Ben Y. Zhao, 'Privacy, Availability and Economics in the Polaris Mobile Social Network', ACM Workshop on Mobile Computing Systems and Applications (HotMobile 2011), https://www.cs.ucsb.edu/~ravenben/publications/abstracts/polaris-hotmobile11.html.

vantage is that these are often features that people want, and even if some or most users don’t want them, they exist in centralized systems because there is a monetary incentive for it. This takes us back to the argument about economic feasibility: an architecture that does not provide these features will have a tougher time competing in the market.

The limitation of crypto for enhancing privacy is a topic that I’ve studied and spoken about in some detail. I believe that the proponents of cryptography have massively underestimated the implementation costs and other practical problems with the paradigm of cryptographic privacy-preserving computations. To give just one example, a small change in what needs to be computed might involve re-doing the protocol (if not the math!), possibly affecting the database schema, and rewriting the code based on that. This is an extremely poor fit to the pace and style of modern web software development.

Seda: One respondent argued that building large distributed systems is simply very hard – much harder than building services that rely on centralized resources that benefit from economies of scale. Moreover, decentralized systems are not only confronted with interoperability issues, but also affected by the labor and cost it takes to replicate the desired application on different platforms. Are these indeed the main obstacles and are they sufficient to explain why decentralized alternatives can’t compete with their centralized equivalents? And is the solution to this problem, as this respondent suggested, ‘to focus on software components, infrastructure, and services that make the job of building privacy friendly architectures as easy as centralized ones’?

Arvind*:* My own views on the implementation difficulty and inefficiency of decentralized architectures are in fact stronger than expressed in the paper, so I find a lot to like about this argument. However, it may be a stretch to say that implementation obstacles are a sufficient explanation to the exclusion of economic and cognitive factors. There are examples of architectures that are highly complex and expensive in terms of development labor that evolve due to economic feasibility: witness the byzantine world of online advertising compared to the straightforward alternative of charging for services (such as social networks, mobile apps, etc.)

My view has been that building more technological components is not what is required, and addressing usability and economic issues is the need of the hour, but after reading this response I’m willing to rethink that. Perhaps the world is not ready for decentralized personal data architectures, and perhaps there are centralization-decentralization cycles as some have suggested, instead of linear trends. Building out the technological plumbing could prove to be a smart bet in case the equation changes favorably in the future in terms of economic feasibility.

Vincent*:* I think newcomers need differentiating features to attract users and encourage them to spend more time on their platforms. Many developers thought that privacy would be enough to attract new users but this ‘feature’ alone is not sufficient because, as was said before, these projects do not adopt the same definition of privacy. More generally, some users will find that Facebook privacy settings correspond to their expectations (or that they can use Facebook without revealing sensitive information) and would see no reason to leave Facebook. Others prefer to not use social networking at all.

I want to point out that I don't believe that decentralized alternatives cannot compete with their centralized equivalents. Centralized networks can also struggle with network effects. For instance, Google+ had a slow start although being supported fully by Google architecture. Had this project not been advertised and promoted by Google, I think it would have been quickly ignored.

Solon: I think there's a lot of merit to the argument that the technical difficulty of developing and operating decentralized services is *the* overriding reason for their lack of success, but I don't think this explains whether these projects would even improve the state of online privacy if they were able to meet their technical goals. A more elaborate supporting infrastructure won't ensure that these projects are any more likely to produce the intended or expected privacy outcomes. There are conceptual problems around privacy that are independent of the onerousness of getting a decentralized service up and running.

That said, there's no question that economies of scale can significantly reduce development and operating costs. Similarly, the ability to cobble together a supporting infrastructure for independent services based on modular, centralized parts encourages more experimentation and hastens the pace of innovation (how many projects owe their existence to Amazon's rentable storage and computing power?). And yet, for all that, developers don't seem particularly disinclined to build services based on decentralized architectures. Many, many, many[13] such projects already exist, some of which are even operational. And for all the development costs, work on them continues apace, including on those that attempt to disperse the costs of operating the service to local nodes (transforming this additional operational burden into something of a virtue). By spreading the responsibility of running the service across all users, they make up in community contributions what they lose in economies of scale (Bit Torrent serving as the obvious paradigm).

Given that these challenges have not dissuaded developers and have even given rise to clever ways of handling ongoing costs, the question is really how the relative difficulty of building decentralized services affects the nature and quality of the services on offer. This requires a shift from the perspective of the developer to that of the user – the idea being that people choose not to use decentralized services because they lack equivalent functionality or ease of use. But, as Vincent's point about Google+ makes clear, this alone does not seem to determine why people happen to adopt a service (or, more broadly, join a community). Equivalent features – and even the addition of novel privacy controls – don't produce sudden migrations. Google was only able to counteract the power of network effects through the *non-technical* work of promoting its service. And what this reveals is something that none of us fully address: how those players that are already a large part of our online lives are in an especially privileged

[13] See, Venessa Miemis, '88+ Projects & Standards for Data Ownership, Identity, & A Federated Social Web', Emergent By Design, 11 April 2011, http://emergentbydesign.com/2011/04/11/88-projects-standards-for-data-ownership-identity-a-federated-social-web/; Daniel Appelquist, Dan Brickley, Melvin Carvahlo, Renato Iannella, Alexandre Passant, Christine Perey and Henry Story, 'A Standards-based, Open and Privacy-aware Social Web', Harry Halpin and Mischa Tuffield (Eds) W3C Incubator Group Report, 6 December 2010, http://www.w3.org/2005/Incubator/socialweb/XGR-socialweb-20101206/; and 'ProjectComparison', Gitorious, 11 November 2012, https://gitorious.org/social/pages/ProjectComparison.

position to proselytize, both because they have the existing financial resources to do so and because they already have routine access to us.

Seda: In your paper, you make a strong case for the 'power of network effects'. And you argue that these effects are enhanced by what you call 'tighter integration'. What do you mean by this? And why did large social networks that should have benefitted from this (i.e., Geocities, MySpace, and Orkut) still fall victim to unraveling?

Arvind: It is inevitable in practice that administratively decentralized systems, with multiple competing interoperable implementations, would offer distinct experiences to users. It could be that interoperability is incompletely implemented, or some friction for users in utilizing interoperability, or something even subtler such as nudges implemented differently by different vendors. After all, if there were no substantive differences between implementations, the lack of diversity would defeat the point of administrative decentralization.

This is what I mean by tight integration in centralized networks and the lack thereof in decentralized ones. As we turn the dial from one end to the other, we can see that if there were little or no integration between different implementations, there would be no network effect, and if they were interoperable to the point of being indistinguishable, we would get a network effect that's as good as a centralized system. I'd expect that in practice it would always fall somewhere in between.

This is not to say that centralized systems don't have limitations to the network effect. The most familiar example is that different countries' populations are only weakly connected, so most services in markets with network effects have trouble breaking into some countries even if they absolutely dominate in others (i.e., Facebook and Google have trouble in Russia, where the incumbents in social networking and search are Livejournal and Yandex, respectively.) The evolution of the 'world map of social networks'[14] is very interesting. But where centralized networks have geographic islands, my claim is that decentralized networks additionally have islands of implementations.

This brings me to unraveling in social networks, a subject I find fascinating. The popular perception of what happened with MySpace (or even earlier, Friendster, classmates.com, etc.) is that a better or cooler service came along, and users switched. Obvious as this explanation seems, it's not true! Before Facebook, there was never a social network in the U.S. used by anywhere close to half of internet users. To simplify a bit, the way MySpace was supplanted by Facebook was that *new* users – those who weren't using social networks yet – mostly chose Facebook over MySpace because it was better. But actual *switching* didn't happen until Facebook was already about as popular as Myspace.[15] Due to the size of Facebook (measured as a fraction of the population in the markets it dominates), the dynamics by which Facebook supplanted MySpace

[14] Vincenzo Cosenza, 'World Map of Social Networks', Vincos Blog, http://vincos.it/world-map-of-social-networks/.

[15] See, danah boyd, 'White Flight in Networked Publics? How Race and Class Shaped American Teen Engagement with MySpace and Facebook', in Lisa Nakamura and Peter A. Chow-White (eds) *Race After the Internet*, London: Routledge, 2011, pp. 203-222. Available at, http://www.danah.org/papers/2009/WhiteFlightDraft3.pdf.

cannot be used to supplant Facebook! This is a subtle but crucial point. Obviously I'm not saying that Facebook cannot be disrupted, but the process will have to be different, and I'd wager it would be dramatically harder. It might take nothing short of serious mismanagement, like Orkut failing to crack down on spam.

Seda: Finally, do you think that alternative (decentralized) networks, even if their design principles and adoption are different from what the market is used to, can spur experimentation and innovation, diversify access to technology, and create potential grounds for social and technical change?

Arvind*:* I don't doubt that alternative architectures can bring societal benefits, although perhaps to a lesser extent than some of the portrayals I've read. While I remain skeptical of the likelihood of broad adoption, I acknowledge that some of these systems can be meaningful even when used by a niche group. Besides, the existence of these projects serves as a hedge against companies trying to usurp too much power too quickly, and keeps awareness of these issues in the public consciousness. So I'm glad these alternatives exist in some form even if they are never widely used.

Vincent*:* I am pessimistic about the future of decentralized networks. I think that technologies and devices are evolving in the other direction: towards more centralization. Although mobile devices are more and more powerful, storage capacity has remained stable over the last few years. Look at the capacity of Apple devices: we're quite far from high capacities that we used to have on computers. The storage capacity of mobile devices is not increasing; in fact, you can store less content on your devices now that movies and pictures are displayed in higher resolution and thus require more space. Although this trend might be caused by the shift to solid-state drives (SSD), the main reason is probably that device manufacturers are pushing to move storage to the cloud. So my fear is that the device manufacturers (which are often service providers), by reducing local storage in favor of storage in the cloud, will effectively prevent the development of decentralized services.

Arvind*:* Like Vincent I believe the gradual trend is toward more centralization. Lately I've come to see a future of 'digital feudalism' as the most likely possibility, where a few brands offer tightly controlled, vertically integrated platforms (hardware, software, apps and content) that are even less interoperable than they are today. I don't see this as dystopian, but there are obvious problems with this world, and it is close to the antithesis of the visions that have animated many of the efforts we've discussed.

There are two ways in which I see things evolving past digital feudalism, not mutually exclusive. One is regulation aimed at limiting the power that private companies have over us – our identities, our relationships and communication, our thoughts, our selves. This regulation might have the side effect of making it easier for alternative architectures to thrive, although I'm not holding my breath. The second is classic disruptive innovation. If the mantra is to be believed, it's something that will start out as a toy, so it's hard to predict where it might come from. But it will probably not be something that's conceived of as a frontal assault on the status quo.

Solon*:* I would never want to dismiss the significance of projects that demonstrate how things might otherwise be. They serve a social purpose that exceeds their immediate technical goals because they signal that there is no natural order of things.

Like Adnostic: a proof-of-concept that undermines the moral authority of those who claim that the apparent benefits of targeted advertising are only possible when third parties are allowed to track users. From this perspective, I also think that there is a lot to learn from the poor uptake that projects like Adnostic experience. They have the (sometimes unintended) effect of throwing into sharper relief the non-technical factors that determine why certain projects flourish and others flounder. Ironically, their failure in the marketplace can be a political success.

References

Appelquist, Daniel, Dan Brickley, Melvin Carvahlo, Renato Iannella, Alexandre Passant, Christine Perey and Henry Story. 'A Standards-based, Open and Privacy-aware Social Web', Harry Halpin and Mischa Tuffield (Eds) W3C Incubator Group Report, 6 December 2010, http://www.w3.org/2005/Incubator/socialweb/XGR-socialweb-20101206.

boyd, danah. 'Facebook is a Utility; Utilities Get Regulated', *Apophenia*, 15 May 2010, http://www.zephoria.org/thoughts/archives/2010/05/15/facebook-is-a-utility-utilities-get-regulated.html.

______. 'White Flight in Networked Publics? How Race and Class Shaped American Teen Engagement with MySpace and Facebook', in Lisa Nakamura and Peter A. Chow-White (eds) *Race After the Internet*, London: Routledge, 2011, pp. 203-222.

Cosenza, Vincenzo. 'World Map of Social Networks', Vincos Blog, http://vincos.it/world-map-of-social-networks/.

'Google+ and Privacy: A Roundup', *33 Bits of Entropy*, 3 July 2011, http://33bits.org/2011/07/03/google-and-privacy-a-roundup/.

Greschbach, Benjamin, Gunnar Kreitz and Sonja Buchegger. 'The Devil is in the Metadata – New Privacy Challenges in Decentralised Online Social Networks', Fourth International Workshop on SECurity and SOCial Networking, School of Computer Science and Communication, Stockholm, Sweden, 19 March 2012, http://www.csc.kth.se/~bgre/pub/GreschbachKB12_MetadataPrivacyDecentralisedOnlineSocialNetworks.pdf.

Miemis, Venessa. '88+ Projects & Standards for Data Ownership, Identity, & A Federated Social Web', Emergent By Design, 11 April 2011, http://emergentbydesign.com/2011/04/11/88-projects-standards-for-data-ownership-identity-a-federated-social-web.

'Platform for Privacy Preferences (P3P) Project: Enabling Smarter Privacy Tools for the Web', W3C, 20 November 2007, http://www.w3.org/P3P/.

'ProjectComparison'. Gitorious, 11 November 2012, https://gitorious.org/social/pages/ProjectComparison.

Project VRM wiki contributors, 'Main Page', 4 September 2012, http://cyber.law.harvard.edu/projectvrm/?title=Main_Page&oldid=5867, accessed 1 December 2012.

Wilson, Christo, Troy Steinbauer, Gang Wang, Alessandra Sala, Haitao Zheng and Ben Y. Zhao. 'Privacy, Availability and Economics in the Polaris Mobile Social Network', ACM Workshop on Mobile Computing Systems and Applications (HotMobile 2011), https://www.cs.ucsb.edu/~ravenben/publications/abstracts/polaris-hotmobile11.html.