Minding the Identity Gaps

SO OLD AND SO NEW, SO SPECIFIC AND SO BROAD

A nearly universal feature of human cultures is that people have names, and being able to distinguish one person from another by name has long been a key ingredient of social and economic life. Even isolated tribes that do not use personal names, such as the Machiguenga of the Amazon, refer to each other using genealogical and biographical information, such as “sister, the one who slipped in the river.” In modern times, the United Nations Convention on the Rights of the Child declares that every child has the right to a name from birth.¹

False and mistaken identities have long been a trope in entertainment, from Shakespeare (The Comedy of Errors) to Hitchcock (North by Northwest) to Monty Python (Life of Brian). They are a particular bugbear in the digital age, because so many more interactions happen remotely, with no direct visual cues or physical contact. As one dog in a well-known cartoon told another, “On the Internet, nobody knows you are a dog.”²

Being able to carry your identity with you into cyberspace is therefore a precondition for a range of activities and services available in the digital domain, which is why digital financial inclusion is so dependent on the development of appropriate digital identity solutions. Lacking formal proof of identity creates many practical problems in the lives of poor people, (see box, page 32).

More broadly, digital identities are needed to direct personal communications, assert ownership over digital assets, to manage access rights to resources and services, and to manage government programs.

A key theme of this article is that we should not get caught up in the idea that digital identity is fundamentally—or even primarily—about your full name and a legal ID number. Your digital identity is clearly about you, but it is less clear exactly what about

---

Ignacio Mas is a Senior Research Fellow at the Saïd Business School at the University of Oxford. He has drawn on generous funding from the Bill & Melinda Gates Foundation for this work. David Porteous is the CEO of BFA, a consulting firm based in Boston, Massachusetts. This paper is inspired by and draws many ideas from the writings of David Birch.

© 2015 Ignacio Mas and David Porteous

innovations / volume 10, number 1-2

27
you it is meant to represent. Ultimately, what notion of identity we need will depend on the context and on what issue it is supposed to help resolve. For instance, identity is often used to demonstrate a person’s right to things they own and services they are entitled to. A key to a safety deposit box, for example, can be a form of identity, because it identifies the holder as the person entitled to open an otherwise unmarked box. Clearly, identity is not always proven by giving a name or showing an ID.

The topic of digital identity presents some deep philosophical and technological problems, but people experience it primarily as a practical challenge. Less advantaged people may have a hard time establishing their identity if they have no approved photo ID, whereas many who are more privileged feel they have too many identities, particularly in the digital domain—think usernames and passwords. A not insignificant minority has experienced the hassle of having their online or credit card identity stolen. We may resent having to reveal our name or ID number to do things not specifically related to that information, but at the same time we fear that the all-pervasive Internet will combine our every action into one big trawable, marketable database.

Identity, which is the basis for self-determination and autonomy, lies at the core of notions of democracy, private property, and government services. However, although establishing and managing our identity, digital and otherwise, should help us function
effectively as citizens of a modern society, it can feel like a sequence of hurdles we need to jump over and arouse some deep-seated concerns about privacy and control by the State.

This paper looks at identity from two opposite but complementary perspectives. The first is a reductionist or biological perspective in which identity is associated with a person as an autonomous organism with a unique genetic makeup. The legal basis for identity tends to be based on this perspective, and it raises questions about the confidence with which a person can assert and confirm their identity. The second perspective, which is information based, views individuals as a complex web of personal information and attributes. Digital markets tend to take this second view of identity, which characterizes customers more in terms of defined attributes, preferences, and transaction histories than on intrinsic uniqueness. This perspective raises fundamental questions about what information people can legitimately be expected to reveal to prove their identity and what information they have the right to keep private.

These perspectives bring together the two key aspects at the heart of most identity problems: security—the confidence with which one can establish an identity and privacy—the control over what personal information is revealed and disseminated. When security and privacy are not handled appropriately, a lack of trust arises between social entities, between customers and providers, and between citizens and the state. In this article, we explore the nature of this lack of trust and what can be done about it. We do so by introducing three types of gaps that arise when the expectations and needs differ between the two parties involved—the disclosing party that provides the information, and the relying party that acts on it. The security gap relates to the certainty with which the relying party can establish the identity of the disclosing party. The privacy gap refers to the confidence with which the disclosing party can control what information is revealed to or discoverable by the disclosing party. The reputational gap is the general fairness and accuracy with which the parties project the information they have available today to project future behavior and performance by the other party.

IDENTITY BY PROXY AND THE SECURITY GAP

The issue of identity presents two conceptual questions, even in its narrowest sense (i.e., are you really you?). First, there is the definitional question of what constitutes your individuality. A sensible starting point might be your genetic code, although it also could be a combination of personal attributes, biological or psychological, or on what you have experienced in your life so far.

Second, there is the empirical question of how identity can be verified through observation. This has practical implications, as it largely involves making tradeoffs between reliability (including availability and accuracy) and cost. If the essence of your individuality is not directly observable or is too costly (such as running a DNA test each time you use an ATM), then identity must be verified through more readily observable proxies.
In traditional societies, the most common proxies are physical attributes. A proxy also might be a bit of information, some fact from a common past that you can provide to convince the concerned party of your identity. A more sophisticated version involves passing cyphered messages with a secret code only you and your counterpart know. A proxy is often simply another person: if someone I know will vouch for you, you will then be known to me.

You might notice that matters of identity have not moved far, even in our modern society. We now speak of factors of authentication, which are taken from three broad domains:

- **Something you have**—a national ID issued by the state, a bank card, or a phone with a unique SIM card
- **Something you know**—a membership number, a personal identification number (PIN) associated with a bank, or the name of your first pet, which you’ve previously disclosed to a website

---

**Figure 1.** The identity security gap
Something you are—the photo on your ID or your fingerprint

Each entity that wants you to prove your identity will rely on the authentication factors it chooses to work with—at the ATM, for example, you will be asked to enter your PIN. The factor an entity chooses then becomes your credential when dealing with it. In practice, then, when asked to prove your identity you are really being asked to prove your command over certain qualifying credentials. The need to prove this introduces the potential for a gap in trust that we call the security gap: Is the person presenting the credentials really the person they are meant to identify? Have the credentials been compromised? Figure 1 depicts the trust gap between identity and credentials, and breaks down some key considerations by type of credential.

The security gap generally stretches over three steps in the identification process (see Figure 2):

- **Original identity verification.** This is about establishing a mechanism for users to prove their identity, typically on a one-off basis. In banking circles, this step is referred to as know your customer (KYC). It might entail asking the user to show a valid public credential, such as a national ID, undergo a face-to-face interview and answer certain questions, or produce a letter of introduction from a reputable local authority or eminent person.

- **Credentialing.** This is about giving the user a new identification mechanism that replaces the original identification process. It might entail giving the user a unique credential (e.g., a new bank card with an assigned username), linking the user’s profile to an available credential (e.g., a mobile phone number or password), or agreeing on a set of questions/answers that provide an identification protocol. Providers requiring a high level of security will insist that the customer present credentials from at least two different domains, such as a card and PIN. Providers that require only one credential may want to provide a backup in case a customer loses or forgets their primary credential.

  Once given, credentials are generally available for repeated use. To mitigate the risk of a PIN being stolen if it is used frequently, credentials tend to have a defined validity period, after which a new card is issued or a PIN is reset. When security standards are very high or when credentials are communicated over insecure channels, they may need to be reset after each use. Such credentials, called one-time passwords, are often delivered via a text message to the user’s mobile phone.

- **Authentication.** This process establishes the user’s identity based on the chosen credentials on an ongoing basis, rather than having to undergo identity verification every time.

  It is important to note that the security gap sits squarely on the side of the provider, as problems can arise in the original identity verification stage: Is your staff checking IDs vigilantly? Do you trust the credentials you are imposing on
your users? Are your cards secure enough? Should your customers’ PINs contain a
greater mix of characters? The provider has the power to tweak these elements, but
doing so has implications, not only for its policies and costs but for the customer
experience, principally in terms of convenience and usability.

Moreover, if a customer does not protect their credentials adequately, the problem
may stem from a deficiency or limitation in the service offered by the provider which
forces the customer to behave in ways that may create unintended security risks. For
instance, a customer may share the PIN on their mobile account because they want a
friend to collect money on their behalf from the ATM, and thus decide to go for conve-
nience over security (see box, page 41).

Systems will always incorporate an error rate and some risk of manipulation, thus
providers must build their identity systems with some threshold of reliability and ac-
curacy in mind. Therefore, on a verification or authentication challenge, the provider
must determine when someone has “good enough” probability—that a system will rec-
ognize, rather than prove, a user’s identity. This notion of “good enough” identification
is particularly apt in a data-rich online environment, where providers can establish
original identity verification without asking the customer to show an ID or conducting
a face-to-face interview. There are three mechanisms for doing this:

• **Delegated KYC.** In Kenya, M-Shwari is an account issued by the Commercial
Bank of Africa (CBA), which is technically and operationally tightly bound into
Safaricom’s M-PESA mobile money platform. If an M-PESA customer wants to
open an M-Shwari account, Safaricom will share the customers’ KYC information
with the CBA so the customer need not be put through a second KYC process.

Delegation of KYC can occur without the explicit agreement or collaboration
of the entity that has conducted the original KYC. PayPal, for example, relies on
new customers’ KYC with other banks. PayPal asks a new customer to prove they
are entitled to use the bank account they have linked to their new PayPal account,
and then sends two tiny amounts of money to that bank account without disclos-
ing the amounts to the customer. PayPal then challenges the customer to log into
their bank account and reveal the amounts it sent. If this is done successfully, Pay-
Pal and the authorities are satisfied that the customer’s identity can be traced back
to the customer’s KYC documents at the bank.

• **Social validation.** While social networks initially were comfortable with and even
played up users’ anonymity or “pseudonymity” as a form of liberty, many now see
value in users representing themselves recognizably and truthfully. A network can
induce (but not force) customers to reveal their true identity by offering a desir-
able service that only makes sense to those who identify themselves properly. For
instance, why put your CV on LinkedIn if you are not prepared to be known? And
how do you expect your old friends to find you on Facebook if you hide behind
an assumed name? In general, the more you reveal on a social networking site,
the harder it will be to conceal your legal identity or to develop a persona that represents you only partially.

- **Data- (or cloud-) based KYC.** This “big data” solution is the practice of analyzing customers’ usage patterns and other data they provide and correlating this information with other publicly available information sources. The purpose is to infer customer identity. The practice is consistent with the informational rather than the biological view of identity.

  This approach can incorporate elements of the previous two: the online lender Lenddo, for instance, will check out your social media presence and assess how likely you are to be the person you claim to be. They do so by checking whether the personal information you put on Facebook tallies with the information you put on Twitter, or with information they get elsewhere. This can be done iteratively, and they also will check how many people you are connected with and how real they are likely to be, on the principle that the more real people you are connected with the harder it will be for you to lie about yourself.

It is important to recognize that the identity question can involve multiple stakeholders, and that what is good enough for one may not be for another. Banks want to have a good handle on who their customers are for the purpose of credit assessment; they particularly want to capture all credit reports pertaining to a borrower. Real-time payment providers that do not assume credit risk primarily want to be certain they make money available to the true account holder; however, they may not feel they need to know who the true account holder is. In this case, they will not invest so much in identification systems but will take credentialing and authentication systems very seriously. The state's tax and law enforcement authorities may have different priorities, and therefore may insist on proper identification so all people and their money movements can be tracked and traced. They will therefore regulate the customer identity requirements (KYC, in bank-speak) that all payment providers need to fulfill.
A number of policy agendas thus converge around the issue of identity, and they don’t all have the same understanding of the security gap. The risk is that they ratchet up to the highest standard and the inclusion agenda gets subordinated to other agendas, which often puts identity requirements beyond the reach of low-income customers at least.

Various providers’ and regulators’ identity verification needs are reflected in the level of stringency applied to each of the three steps in Figure 2—whether they are performed at all. The following security gap models are archetypal (see Figure 3):

- **Anonymity.** Anonymity is the absence of identity handles. All three steps are skipped, and providers are not able to trace individual interactions back to any specific user. Anonymous transactions may be recorded, but they leave no trail that leads back to its counterparties.

- **Persistent pseudonymity.** Persistent pseudonymity means that the user presents an assumed identity at each interaction, one not linked to a legal identity. They skip the first step, original identity verification, but credentials are assigned for subsequent authentication purposes and thus become persistent over time. It is possible to link the transactions of a given pseudonymous user, but that pattern of transactions cannot be linked to a legal identity.

  This is a prevalent model for nonfinancial service providers on the Internet, which require that customers register once (credentialing with user name and password) and log in with each use (authentication). At registration, users may be asked to provide personal identifying information, but that is usually unverified—hence the proliferation of funky and arbitrary user online names.

- **Conditional pseudonymity.** Conditional pseudonymity means users can act pseudonymously in the course of normal social or business interactions, but there is an exceptional mechanism for law enforcement to unmask these users if they breach the law, or to facilitate the investigation of alleged crimes. This requires that a record be kept that links pseudonyms to legal identities. This can be implemented by breaking the steps in Figure 2 into two chains handled by separate entities. It entails a KYC process not unlike the PayPal example mentioned earlier.

  To implement conditional pseudonymity, you establish identity (KYC and credentialing) with an independent party we can call a digital identity provider. Once you have a mechanism to authenticate with the digital identity provider, you can request that it confirm to anyone through a secure message that you are a known person and your KYC details are on record, but without revealing who you are. This allows you to go to any site, register with any pseudonym you like, and ask your digital identity provider to confirm that it knows you. More formally, the digital identity provider would send the site a unique number that represents the link between your chosen pseudonym and your known identity.
This mechanism allows users to remain anonymous with Internet service providers that don’t need to know their actual identity while confirming to providers that you are a real person. They can thus maintain a record of all your transactions. Law enforcement authorities would be able to trace all transactions back to a certain individual if they produce a court-approved request to a digital identity provider to confirm a suspect’s real identity.

- **Full identity verification.** Full identity verification means that each provider insists on full customer identification. This may be done by applying all three steps in Figure 1, as banks do: KYC and credentialing at account opening, and authentication based on those credentials at each subsequent interaction.

  Some providers may skip credentialing altogether and require original identity verification at each customer interaction. This would make sense if a user’s interactions were less frequent and occurred in a face-to-face setting, and in a location where everyone carries a national ID. This reduces complexity but is a stricter and costlier process.
What is needed are systems that draw people into improving their digital identities over time, rather than perpetuating the stark distinction between the identifiable and the faceless. If there is no centralized government identity system, a more decentralized system is needed to allow the issue of identity be resolved within the communities where excluded people live, shop, and work. Identity could be established in the first instance via local authorities, an employer, or a longstanding member of the community.

Once a person has established their identity, they can attest to the identity of others in their social circle. These circles can overlap to build greater credibility, much as one’s Facebook identity is more credible the more people one is connected to. This is the notion of social identity—even people with meager resources can help each other overcome their limitations to establish an official identity. This approach is hard to reconcile with the way governments and formal institutions tend to handle identity. The digital era calls for flexible notions of identity that build layers of information and verification through both social networks and bureaucratized processes.

MULTIPLE PERSONAL FACETS AND THE PRIVACY GAP

So far we have treated identity as a binary concept with a single purpose: to determine whether you are who you purport to be. However, the notion of identity quickly gets caught up in personal information in two main ways:

- **Identifying information.** As noted, proving identity often requires exposing some personal information that is used explicitly as an identity proxy, such as a personal history question or people who vouch for you.

  Service providers often trawl through personal information to prevent fraud, particularly theft of identifying credentials such as your credit card use history to determine whether a new transaction is suspicious. The information requested often seems disproportionate to and of dubious relevance to the situation.

- **Correlated information.** Many online sites ask you to assert your identity because they want to know not only whether you are you but something more about you. Facebook, for instance, is keen to get you to reveal your affinities, as that information is of value to advertisers.

  Why do we resist service providers’ attempts to extract personal information and distrust organizations that use information to build profiles of us? Oxford University philosopher Luciano Floridi argues that people want to control their personal information because that is central to how we shape our identity—by withholding personal information we control how we project ourselves.9

  Identity includes three intertwined aspects: intrinsic characteristics that constitute who you are, your self-conception of who you are, and how you are perceived in society. Each of these influences the others. Figure 4 illustrates that what you think of yourself affects your behavior, which in turn affects how you are perceived.
SECURITY IS REINFORCED IF THE SERVICE DOES WHAT CUSTOMERS WANT IT TO DO?

We often hear anecdotes about how some people who are new to banking or mobile money share their PINs or forget them, which is taken as evidence that PINs may not be an appropriate authentication mechanism for the poor. So, while we marvel at how the poor are sophisticated portfolio managers per Portfolios of the Poor, we question whether they can manage four digits. Those who share their PINs are often thought to be illiterate, ignorant, or irresponsible.

There may be other reasons why people share their PINs. It may be entirely harmless for them to do so because they use their account only seldom and/or have very little money in it. When the account does not hold treasure, the PIN will not be cherished. The implication is clear: give customers other reasons to value your service, and thus their PIN.

People also share their PIN because they feel it is useful for someone else to know it. For instance, because they have to go to a distant ATM or agent to collect their cash, they take turns with friends or family and collect payments for each other. Why not?

When people do something with your product that you don’t want them to do, the reaction shouldn’t be “educate!” but to see how the product can be changed to accommodate the customer better. Take customers who cash out money for each other. Under “withdrawal” in your mobile phone service, you could offer an “in person” and “through a friend” option. The latter would generate a code that you share with your friend who is going to fetch the cash for you. This would be similar to how “sending money to non-customers” or “ATM withdrawals” work on many mobile money systems, and would give customers a safe way to have someone else collect their money.

by others; it also shows that you adjust your behavior and your self-concept based on how you sense you are perceived.

We manage our personal information and thus our identity chiefly by compartmentalizing the different roles or personas we assume in different circumstances. You don’t present yourself in the same ways to your employer, your family, and your friends, for example, and you are able to control which persona you are enacting in each circumstance.
Ignacio Mas and David Porteous

I am confident, so I think of myself as such, and others come to see me as such.

People think I am self-confident, and that helps me feel more confident; over time, I become more confident.

**Figure 4.** Intertwined aspects of identity

**Figure 5.** Types of personal information
The various facets of your persona can be construed as distinct identities, each with a different set of attributes. Who we are is the combination of these distinct identities, and we want to be able to alternate between them as we go through our daily life.

Figure 5 depicts different personal information sets that correspond to an individual’s different identities. Personal information can include personal attributes such as age, personal history such as what school you attended, transactional histories, and relationships. Of course these attributes can overlap, and you may not want every person who needs to know your identity to know all of these things. In essence, you want to ration the information you disclose, some of which will be used to identify you. This creates a tension between identification and privacy. We want to be in control not so much of how much we reveal but more fundamentally of how we represent ourselves.

The issue of privacy, therefore, is a twofold question of personal choice: which identity we choose to reveal in any given interaction, and what information about that identity we wish to make known.

Data privacy issues are particularly prominent and sensitive in today’s digital world, for three main reasons. First, with the spread of the Internet and electronic sensors, vast amounts of data are “born digital” and hence can be captured by providers at extremely low unit cost and often without users even knowing about it. Second, with the falling costs of digital memory and the growing power of analytical tools that operate on large and unstructured datasets, providers no longer need to be selective about which information they choose to store; the business default rule is generally to keep whatever information is available. Third, the value of personal information is reinforced by quantity: the more types of information, the longer the history, and the more frequent data points a provider has, the more useful each data point can potentially be. These are the core drivers of big data.

A breach of one’s informational privacy, beyond being “a form of aggression towards one’s personal identity” (as Floridi phrases it), raises some legal and ethical concerns in three specific situations:

- **Inappropriate use of private information.** This is where a provider uses customers’ personal information for its own business advantage in a way that is detrimental to those customers. For example, a provider might use customers’ personal information to drive targeted marketing campaigns that expose that private information to third parties.\(^{10}\) Or a provider might use its knowledge of its customers to target its proposition so narrowly as to limit customers’ choices.\(^{11}\) A provider also might use personal information to set pricing that discriminates against the interests of specific customers.\(^ {12}\)

- **Unauthorized sale of private information.** The larger threat is that a provider might sell its customers’ personal information to third parties that want it for their own marketing purposes. This practice exposes customers to unsolicited marketing calls and creates a less transparent business logic, whereby products are sold to customers and customers are sold to other businesses.\(^ {13}\)
• **Data leakage or theft.** A provider that amasses valuable customer information could be the target of cyberattacks, through which customer information could pass into the hands of those who might use it for nefarious purposes.\(^{14}\)

The privacy gap is therefore a lack of trust between customers and providers. Customers’ desire to control how and what they expose about themselves can conflict with providers’ desire to know as much as possible about their customers. So how can customers protect their information while obtaining the services they need and want?

An extreme approach to dealing with this conflict would be to give customers the right to engage in digital transactions anonymously when they feel they need to protect their privacy. No identity information is shared in an anonymous transaction, which therefore cannot be linked to the user.\(^{15}\) Anonymity leads to the atomization of user information, since each data point exists independently, and it may be adequate for low-quality web interactions.

However, solving the privacy gap in this way comes at the cost of aggravating the security gap: providers cannot be sure who they are dealing with and authorities have no way to track illegal activities that occur in cyberspace. Anonymity thus creates an unattractive tradeoff between security and privacy, which is made worse by the fact that users with ill intent are more likely to choose anonymity, using claims of privacy as their cover.

The opposite extreme is to insist on proper proof of identity with every digital interaction and to put strict constraints on what providers can do with the information. These constraints could include laws and regulations that lay out the basic rules, supported by industry codes of conduct for providers that are supplemented by companies’ own privacy policies.

A purely rules-based privacy regime would be hard to enforce because companies’ digital data practices can be opaque, and they will get increasingly complex in the age of big data. Moreover, customers are rarely in a position to observe how providers manage, store, and use their personal information. While the security gap would be minimized by the proof of identity, a significant trust gap would remain around the handling of private information.

A third approach to managing privacy conflicts is to give people a technological solution that lets them unbundle not only their personas but also the information linked to each persona. This would give them effective control over their identity(ies) and personal information while allowing providers to track individual users and comply with legal KYC requirements.

This can be done by adding attribute-based credentialing to the notion of conditional pseudonyms mentioned earlier. Attribute-based credentialing enables entities that know certain things about you to confirm those attributes for third parties, under your express instructions. For instance, you might ask the passport office to certify your name and ID number to a creditor who wants to check your credit status, your bank to certify that you are over age 18, or your school to confirm your education, and so forth. You could link these attributes to your unique identity or to your condition-
<table>
<thead>
<tr>
<th>Known identity</th>
<th>My legal identity is MYZ</th>
<th>Provider does own KYC</th>
<th>IS details and all public information linkable to your legal identity</th>
<th>Legal rules, industry codes of conduct, company policies, ethical restraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conditional pseudonymity</td>
<td>Call me X; my identity is on record with Y</td>
<td>Delegated KYC, proving that I am a real person, a law enforcement can find me if necessary</td>
<td>Info voluntarily supplied. You can link my transactions to each other, but not to me</td>
<td>Above rules + fragmentation of pseudonyms + attribute based credentialing</td>
</tr>
<tr>
<td>Pseudonymity</td>
<td>Call me X</td>
<td>No KYC; registration only</td>
<td>As above</td>
<td>As above</td>
</tr>
<tr>
<td>Anonymity</td>
<td>Don't ask, don't tell</td>
<td>No registration or KYC. Different transactions cannot be correlated or traced to one user</td>
<td>None</td>
<td>Not necessary</td>
</tr>
</tbody>
</table>

**Figure 6.** The relationship between identity and privacy mechanisms
al pseudonyms if you want to compartmentalize how you present yourself digitally. None of these actors would disclose, or even necessarily know, any other information about you.

While using pseudonyms allows you to fragment your digital life and conceal your unique identity, attribute-based credentialing allows you to unbundle the information you choose to share, based on what you consider relevant and appropriate to disclose, strictly on a need-to-know basis. Used together, these two notions can help to eliminate the tension between the security gap and the privacy gap.

**Figure 7. Depicting the security and privacy gaps**

Figure 6 presents a summary of the main approaches we have discussed for dealing with the related challenges of ensuring proper security around identity verification.
WHO ARE TODAY’S DIGITAL IDENTITY PROVIDERS?

Digital identity provision seems a logical extension of services offered for entities that already collect significant amounts of customer data and are trusted to expand. This includes:

**Governments.** Many governments issue their citizens smartcards and digital certificates that they can use to prove their identity online in any setting. A particularly celebrated case is that of Estonia, which has used it successfully to drive a broad-ranging set of e-government initiatives. Interestingly, the Estonian system is set to be extended to foreigners living outside of Estonia in an attempt to create a global, government-standard digital identity.¹⁷

**Banks.** Banks are mandated to collect and verify a substantial amount of personal data on their customers as part of their normal KYC process. Swedish banks have developed a shared standard called BankID that is used by five million bank customers from ten different banks (which have 11 million registered online banking users) to access a range of public and private services. Outside of Scandinavia, banks have generally been quite reluctant to confront the consumer protection issues and assume the legal liabilities involved in managing identity for third parties.¹⁸

**Online payment providers.** Money in digital wallets is essentially protected by customers’ authentication credentials, so the distinction between money and identity online is getting thinner. PayPal already offers online merchants the possibility of letting their customers sign in to their website through a login button using their PayPal credentials.¹⁹

**Mobile operators.** Most governments have insisted that mobile operators collect basic KYC data on their customers at point of sale, although this is generally of a lower standard than bank KYC. The GSMA has now created a standard called Mobile Connect for network-based authentication via a phone message pushed to a customers’ registered SIM card.¹⁹ This provides a sign-on procedure that mobile phone customers can use to validate their identity for any digital service on any communications channel or device; the authentication channel always occurs through the mobile phone.

**Credit rating agencies.** Experian offers an online identity verification system called ProveID that lets businesses assess the veracity of basic customer-supplied information such as their name, address, or age. The system tries to match the customer’s information against various verified data sources. The
while protecting the privacy of personal information. The figure makes clear that the solution adopted determines the degree of privacy afforded.

We now explore the simultaneous closing of the two trust gaps, using the conceptual framework presented in Figure 7. The vertical axis, labeled “privacy,” shows whether the information captured from users in the identification process is too narrow or too broad for the intended purpose, whereas the horizontal axis labeled “security” shows whether the verification process used is too burdensome or unreliable, given the risks involved. Identity solutions should be aimed at the center of Figure 7, where

Online social networks and personal data storage providers. Facebook, Twitter, LinkedIn, and Google+ are all seeking to entrench themselves as single sign-on (SSO) providers for other websites. They have the customer base, but may not have sufficient knowledge of their customers’ true identity to do so. So far they offer convenience rather than a security service. Online private data storage providers such as Mydex have a higher level of trust embedded in their service, which is easier for them to leverage as a self-standing digital identity solution.

Large traditional retailers. Traditional retailers have been reluctant to enter this space, even though they may have a large customer base and significant customer data established through their loyalty programs. The post office in the UK was one of five companies selected to provide digital identity services under a government scheme.21

Trusted digital security solution providers. Companies like Symantec (which bought the VeriSign security business) and Digidentity have developed a business selling and managing trusted solutions that enhance the security of online activities. They offer the technology platform on which trusted authentication can happen (verifying digital certificates, safe storage of public keys). Many are exploring user identity services, generally in partnership with other companies from the previous categories.

Specialist startups. This includes miiCard, which leverages bank KYC as the original identity verification source. To accomplish this, customers must share their online banking login details with miiCard (using the Yodlee platform). From then on it uses one-time passwords sent via SMS to the phone to authenticate customers.
the information disclosed and the verification standard applied are proportionate to the underlying risks and context. In the central position, both gaps are zero, while positions away from the center represent solutions with either a positive gap (excessively intrusive or burdensome) or a negative one (insufficiently documented or verified). Of course, judgments on the appropriateness of privacy and security arrangements will depend highly on context and thus should be understood relative to the nature of the transaction that requires proof of identity and information exchange.

There is an obvious and implicit tradeoff between security and privacy when the two extreme solutions presented in Figure 6—full identity and anonymity—are relied on too heavily. Being required to prove your legal identity at a pub may assure the vendor that you are not underage but it exposes customers to intrusive information sharing with the bar staff. Being able to conduct cross-border transactions anonymously would protect privacy, but it would undermine the enforcement, for example, of anti-money laundering laws.

Pseudonymous solutions can also create disproportionality. Some websites insist on long customer registration forms even though none of the information is verified and hence is of dubious value. Security also is calibrated to identify repeat customers but is not good enough to catch rogue users who use the website for illicit purposes. Sometimes the regulators impose identity solutions that create a lack of proportionality vis-à-vis privacy and security, such as with financial integrity rules that apply the same KYC standards to all financial services and service providers, regardless of the underlying risks they create.

The balance between security and privacy considerations is best met with conditional pseudonymity, plus attribute-based credentialing. With conditional pseudonymity, multiple identities can be attached to a unique and uncontroversial identity, which can be as simple as a unique number attached to personal biometric measurements, but this link is revealed to others only with your explicit consent, or if law enforcement authorities demand it through a formal legal process. With attribute-based credentialing, customer information can be limited to the attributes of an identity that are relevant to the provider. These tools enable providers to aim for the center of Figure 7, but they need to be supported by an appropriate institutional structure and technological platform.

Technical solutions are based on a clear separation of roles between three types of trusted entities:

- **Unique numbering authority.** The foundational element of identity is the assignment of an arbitrary but unique identifier to each person by a unique numbering authority that implements a mechanism that allows people to be uniquely identified, most likely biometrically. It holds a database that matches unique numbers to unique biometrics. It also can hold other personal information, such as name and date of birth. When queried about a unique number and a set of biometrics, the unique numbering authority would only validate whether they match; it would
never disclose the biometric or other personal information it holds, except to law enforcement agencies that present a court order to identify the person behind a given unique number.

- **Digital identity providers.** While you would have one unique number, you might want to present a variety of identities or pseudonyms to different people and in different situations. Digital identity providers would bind each of your chosen identities or pseudonyms to your unique number. You could request that your digital identity provider confirm to a third party that a valid unique number matches your pseudonym. Digital identity providers would never disclose the unique number behind a pseudonym, except to law enforcement agencies that present a court-approved order to identify the person behind the pseudonym.

- **Attribute providers.** Attribute providers have the capacity to certify certain user attributes beyond identity and link them to your chosen digital identity or pseudonym. You could then request a given attribute provider to share or confirm an attribute of yours, such as being over age 18, with a third party.

A trusted system that links these types of players is needed for them to operate, which includes those that validate or certify the providers of identity and attributes, shared data frameworks, and identity federation standards (such as OpenID) that enables users to use their identity information across disparate entities.

**CORRELATING INFORMATION AND THE REPUTATION GAP**

Identity often is thought of as an enabler for social and economic inclusion, but in fact it can be entirely the opposite. Socioeconomic interactions that are based on identity may reinforce the established class hierarchy. Systems based on anonymity might in fact be the most equitable and inclusive in the sense that they ensure equal participation by all. However, anonymous systems also carry a high cost in terms of efficiency, as they would make it impossible to establish reputations or enforce contracts and much more difficult to disrupt illicit activities.

It is unlikely that the poorer segments of the population would be better off if the economy worked on an anonymous basis. Digital identities make it possible for people to participate in the economy in many ways—by opening accounts, receiving money, asserting their rights over digital assets, etc.—but establishing that identity can be a major hurdle, especially in countries without digitized national ID schemes.

It is ironic that the difficulty of establishing identity in the first place is often what locks so many lower-income and rural people out of the digital service domain. Identity systems with selective coverage create a double whammy of inequality by helping the “haves” engage in every socioeconomic interaction they desire while preventing access and negating the visibility of the “have nots.” Therefore, it is of prime importance that
“UNFORGETTING” AND UNFORGIVING CREDIT REPUTATIONS

In *Spent: Looking for Change*, the recent documentary about financial exclusion in the United States, there is a moving segment about a young man named Justin who is determinedly rebuilding his life after having obliterated his credit rating by failing to pay his credit card debt. He says, “People often judge me on the choices I’ve made, not knowing the options that I had.” Maybe if we were aware of Justin’s limited options we would agree that he had taken the most appropriate action by not paying his debt. When his situation had improved, we might even feel comfortable offering him a new loan for a new beginning.

Economists say that credit bureaus are about solving information gaps between creditors and borrowers, but they don’t say which gaps they are talking about. For example, no credit bureau explained to financial institutions that Justin was forced to scratch out a living on his own from age 16 but that those days are now behind him; they only propagated information on his history of nonpayment.

As David Graeber argues in his sweeping history of money, *Debt: The First 5000 Years*, we take it for granted today that all loans must be repaid, fully and on time, but that has not always been the case. Debt moratoria, renegotiations, substitution for assets or labor, even wholesale debt cancellation are common themes; it is only fairly recently that paying off debt has become a test of character summarized in a three-digit score that has become something personal that is disconnected from your circumstances and options.

With the digitization of finance, we face the daunting prospect of “the system” having an unforgiving memory of repayment of formal debt while knowing little else of significance about them. You can blow whatever positive attributes you’ve demonstrated all on one unpaid debt. One strike, you’re out. Big data can become the basis for exclusion.

Exclusion is often the result of ignorance, which creates prejudice. Moreover, in the absence of concrete information, people generalize, and big data is fundamentally about generalizations that will work well on the average for many. But in the process, many Justins will be pushed further into financial untouchability, silent casualties of a system that may be commercially satisfactory but falls short in helping those who need help the most. A way to prevent this unfortunate dynamic is for financial service providers to offer multiple avenues to obtain credit so that people like Justin can get the credit they need.
In sum, besides collecting more data on prospective borrowers to use in making credit decisions, lenders can seek ways to build more trusted relationships with local players in each community, people they can count on to make credit recommendations and help channel credit within their network of friends, customers, savings circles, and business associates.

the same privileges and standards of identity are extended to all. Governments should take responsibility for ensuring that all citizens have a digital identity; a powerful example is the Indian government’s Aadhar initiative to provide everyone in India with a unique number ID linked to biometrics.

While some form of identity is the bedrock of digital financial services, it may not be sufficient to gain access to services that tie the notion of identity to reputation. For example, a person who is well known and reputable is likely to find someone to attest to their identity more readily than others. But what is the true link between identity and reputation? Reputation is grounded on others’ assessment of your characteristics and past behavior. Managing your reputation thus begins with controlling your identity and your personal information, and as such is subject to the security and privacy gaps referred to earlier.

But reputation is more than the information you have chosen to expose about yourself, as your reputation is determined by others and involves their inescapably subjective interpretation of your personal information. It is a social construct through which people interpret your behaviors and actions in ways that project your identity and personal information into the future. This introduces what we call the reputation gap, which is the gap between how others judge you and how you judge yourself. The size of this gap determines how hard it will be to establish mutually profitable relationships in both social and business contexts. The box on page 51 illustrates how reputation can affect a credit score, for example, and thus a person’s financial future.

Establishing a mutually trustworthy reputation is of course a key element between providers and their customers. This might be useful in providing specific services like credit, but it can also provide a lens through which all customer issues and complaints are handled, as described in the box above.

SUMMARY: THE EXPANSIVE NATURE OF IDENTITY

In this article we have given expansive treatment to the notion of personal identity. We started with the purely objective view that identity is provable in theory but can only be ascertained probabilistically in practice, which creates a security gap.

We then developed a fuzzier, more intimate notion of identity as a set of personal attributes and history linked to the distinct roles we play in life. In this view, asserting identity is a tug-of-war between the personal information others have about you and the information you want to reveal, which creates a privacy gap.
APPLYING REPUTATION TO BUILDING TRUST IN THE DIGITAL FINANCIAL WORLD

For the adoption of a digital financial service to lead to regular use, clients must come to trust it. —Not only that an instruction will be carried out within an expected timeframe, but also that any difficulties that arise will be resolved quickly. First-time clients typically experiment with a small amount of money, and if the service does not perform as expected it can taint the user’s perceptions of digital payments and even affect an entire local market.

As digital payments become interoperable and electronic value moves between clients at different service providers, the risk of failed transactions grows—and so do the time and cost involved in resolving a query or dispute. With a high volume of small-value payments, providers cannot investigate every problem thoroughly. However, if they don't provide a quick and easy resolution, a lack of trust will likely develop and lead to low adoption and use.

To overcome this, financial service providers could assess a complainant's reputation in terms of their use of digital services and agree on a process for settling a query or dispute accordingly. Hence, a new user may be given more latitude than one who has used digital services repeatedly, particularly one who makes frequent complaints. The user’s pattern of complaints would be analyzed to address systemic issues and prevent fraud. A clear notion of identity would have to underly the user’s reputation to prevent clients from establishing multiple identities that would enable them to take advantage of settlement rules.

It may be that the cost and effort required to develop a systemic approach to resolving problems and failures quickly, thereby building new customers’ confidence in digital payments, is more cost effective than spending substantial money on general customer education.

Finally, we talked about identity as a set of assumptions people make about you. While reputation may be grounded on past and current information, its purpose is to make inferences about your likely future behavior—essentially projecting essential elements of your individuality that may not correspond at all with your own view of yourself, which creates a reputation gap.

Trust between providers and customers will depend on these three gaps being adequately bridged. As a client I will want to know a provider can properly identify me without undermining my ability to define myself to others, such as by publicly disclosing my purchase history, and that the provider is not judging me in an unfair or injurious manner, such as limiting my access to certain products. While the three identity
gaps are separable, they all need to be resolved for trust to flourish between client and provider.

What makes this difficult is that the gaps sometimes create conflicts as different stakeholders place different demands and constraints. The best way to reconcile varied needs is to unbundle the notion of identity: a person can have multiple personas (pseudonyms) used for different occasions; a person has a collection of attributes that can be asserted independently (attribute-based credentialing); privacy rights may be different when dealing with providers than with law enforcement authorities; and a person’s reputation need not be reduced to a single variable.

In the end, the best way to reconcile these gaps is to take a social view of identity. All human societies have juggled solutions to these three gaps and it thus would behoove authorities to introduce formal standards of identity that build on established social norms and behaviors. The process can then be framed as a journey from an informal, socially based identity to one that is formalized and legal.

1. The source for the last three sentences in this paragraph is the Wikipedia entry on personal names. Available at https://en.wikipedia.org/wiki/Personal_name.

2. The cartoon (see https://en.wikipedia.org/wiki/On_the_Internet,_nobody_knows_you%27re_a_dog) is by Peter Steiner and appeared in The New Yorker in 1993, before Netscape was even founded. Since practically no commerce was done over the Internet at the time, the identity problems it highlighted were of an entirely social nature.


4. This is not to say that identity confers ownership; it just establishes the means for asserting ownership if there is a contract or other proof of ownership.

5. This paper provides a broad perspective on identity issues, focusing more on the social and business aspects than on the technical. It does not provide a systematic review of available technology options behind the concepts discussed.

6. PINs communicated over SMS are generally thought to be unsafe because sent messages are saved on SIM cards by default (and on the telcos routing servers as well), so a phone thief could gain access to the two factors of authentication simultaneously (the phone and the PIN). With OTPs, any saved messages that contain old passwords are rendered harmless. Another common way of delivering OTPs is via small hardware tokens or dongles that automatically generate a unique number using a timed algorithm; this number is displayed on a small LED screen for the user.

7. This box is drawn from Ignacio Mas, “My PIN is 4321,” CGAP blog, September 13, 2012. Available at http://www.cgap.org/blog/my-pin-4321.

8. On the financial side, an example of pseudonymity would be the Swiss numbered bank accounts of yesteryear. Cryptographic currencies like Bitcoin apply pseudonymity to money as well, since users must have a unique public address (or key) in order to receive Bitcoin from others, as well as a matching private key that proves they are the legal owner of any money assigned to that address.

10. A real-life situation was when the U.S. retailer Target offered coupons for baby products to a young woman whose earlier purchases coincided with those of pregnant women; her father learned about her pregnancy through those coupons. Available at verdict.justia.com.

11. A real-life situation is where a customer’s TIVO (a digital video recorder) assumed its owner was gay based on something he’d watched, and started highlighting gay content. Available at http://www.wsj.com/articles/SB1038261936872356908.

12. Amazon, for instance, has been accused of this practice, known formally as price customization. It is not technically illegal as long as price discrimination does not occur on the basis of gender, ethnicity, or similar, but it can greatly upset customers. Available at http://edition.cnn.com/2005/LAW/06/24/ramasastry.website.prices/index.html?_s=PM:LAW.

13. The definition of what constitutes personal information may not always be clear-cut. Who owns transactional information: the buyer, the seller, or both? If buyers keep tabs of which sellers are reliable and offer good products, might it not be reasonable for sellers also to keep tabs of which customers are valuable and what each looked at and bought? If I own a website and am paying for the infrastructure that runs it, do I not then own the information that flows through my servers about those who are “surfing” there?

14. In the United States in the last year alone, retailers such as Home Depot, Target, Staples, and Neiman Marcus have all reportedly had their customers’ credit card data compromised. Available at http://www.cnet.com/news/staples-probes-potential-theft-of-customer-credit-card-data/. Even banks such as JPMorgan have been subject to cyberattacks, which may have compromised financial and non-financial data such as customers’ social security numbers. Available at http://dealbook.nytimes.com/2014/10/02/jpmorgan-discovers-further-cyber-security-issues/?_r=1.

15. The analogy is with cash-based transactions, which do not carry any memory of who the cash was gotten from. Transaction histories do not attach to either cash itself or its users.

16. David Birch calls this approach “real reputations, not real names,” which he describes as “keeping [your] real identity tidy locked away at the bank while [you] navigate [your] way around the new economy using attributes. The more you are required to give up your real identity on the web, the more likely it is to be compromised.” Examples of initiatives that implement this are the IRMA project (for “I reveal my attributes”; see https://www.irmacard.org/), U-Prove (see http://research.microsoft.com/en-us/projects/u-prove/), and Idemix (see http://www.zurich.ibm.com/idemix/). See David Birch, “Banks and Blockchains and Clouds, Oh My,” Tomorrow’s Transactions blog, December 10, 2014. Available at http://tomorrowstransactions.com/2014/12/banks-and-blockchains-and-clouds-oh-my/.


18. For more on Bank ID, see here [this link doesn't work, please provide updated url]. Lloyds Bank in the UK has run a pilot with the government’s Cabinet Office to use their identity systems to enable their customers to access online government services (see pilot report here[1 got a virus warning on this link, please provide updated url]).
19. PayPal offers a range of developer tools to facilitate easy integration of the log in with PayPal button (see [I got a virus warning on this link, please provide updated url] here).

20. For more information, on how Mobile Connect works, see Marie Austenaa, “Personal Data: Enabling Trust and Creating Value from Digital Identity,” GSM Association, presentation made at the Mobile World Congress 2014 (available here [I got a virus warning on this link, please provide updated url]).

21. The Register, September 3, 2013. Available at http://www.symantec.com/index.jsp. The UK post office’s traditional role related to identity has been the verification and certification of photocopies of identity documents (see here [I got a virus warning on this link, please provide updated url]).