It's time to disentangle the complex Aadhaar debate

Ever since the government issued a deluge of notifications making Aadhaar mandatory for one scheme after another, the Aadhaar debate has substantially heated up both in the public space and also in the parliament. However, the debate has more often than not resulted in inflexible binary positions without adequate analysis, making it rather difficult for an impartial, lay, but keen, observer to understand the full implications.

The issues have ranged widely from limiting the scope of Aadhaar, to its role in social welfare, to privacy and threat to democracy. There have been several conspiracy theories and debates on whether Aadhaar has helped welfare or welfare has helped Aadhaar. Many of the above issues have often been conflated in arguments, and, amazingly, the past and the present governments seem to have completely swapped their earlier strident positions on Aadhaar. Given the situation, one does not envy the honourable judges of the Supreme Court who are being urged to expeditiously decide on the pending issues.

The scope of Aadhaar

Aadhaar is undoubtedly the most ambitious digital identity project in history. The hurriedly passed Aadhaar bill has restricted its uses only to efficient and targeted delivery of services and subsidies, the expenditure for which is incurred from the consolidated fund of India, yet the government wants to use it not only to enforce better tax compliance, but also to disburse scholarships, to facilitate hassle-free train and air travels, for payment gateways and even for downloading of survey maps! The possibilities of Aadhaar are so enormous that it is inconceivable that its usage can remain limited to preventing leakages in welfare schemes - either the project has to be abandoned altogether or its usage must ultimately grow significantly beyond what is specified in the Aadhaar Act of 2016 or in the court restrictions. Whatever may be the final decision, one can only hope that it will be the outcome of an informed and high quality debate and not be based on strident dogmatic positions.

If we set aside the crucial privacy concerns for the moment, the following appear to be the main issues in deciding on the scope of Aadhaar usage.

First, the government clearly wants to use the unique identification of Aadhaar to enforce compliance in a variety of schemes by avoiding duplicates. As such, making Aadhaar based identity verification mandatory appears to be a strict requirement, and the opponent's insistence on making Aadhaar optional undermines this very purpose. It does appear that an optional Aadhaar will be severely limited, if not almost useless, in scope. A lot has been written about the requirement of *consent* in Aadhaar usage. But should individual consents be necessary for every public policy instrument, especially for those whose main purpose is to enforce compliance, provided due processes are followed to make the usage legitimate? For example, can one refuse to give consent if it is democratically decided that for larger good strict identity verification is necessary for employment or for tax compliance or for property registration? Notwithstanding the fact that it will be absolutely correct to argue that getting Aadhaar passed for limited purposes as a money bill without adequate debate and then recklessly extending its scope violating the Supreme Court's directives is not exactly what can be called *following due processes*, the *consent* and the *opt out* aspects of the debate appear to have been somewhat trivialised. What is perhaps required at the earliest is a detailed analysis of who all may have a right to verify the identity of a resident, to what extent, for what purposes and under what circumstances.

Second, why should Aadhaar be restricted to welfare schemes where the cultural capital required for high frequency digital transactions may be lacking? Admittedly there are substantial leakages in these schemes, but surely the need for de-duplication and strict record keeping and audit are as much and if not more in the domains of tax compliance, real estate transactions and property records and election funding? Why is it that these do not make for more compelling use cases for Aadhaar? Surely we need some analysis on the relative merits of insisting on unique identification in such domains as well?

Third, detecting duplicates is possibly just one aspect of an instrument like Aadhaar. Its scope can possibly be tremendously enhanced to facilitate linking of local ids in currently isolated verticals like census, education, health-care and immunisation records, birth and death records, land records, property registration, income tax, banking, loans and defaults, police verification and law enforcement, disaster management, security and intelligence and such others. Thus, Aadhaar may not only enable efficient design, delivery, monitoring and evaluation of services in each domain individually, but may also offer the possibility of using modern data analytics and machine learning techniques for finding large scale correlations in user data that may facilitate improved design of social policy strategies, including targeting, and early detection and warning systems for anomalies. For example, it may be tremendously insightful to be able to correlate education levels, family incomes and nutrition across the entire population; or disease spread with income and education. More generally, it may enable macro level analysis from high frequency micro level data, econometric analysis, epidemiological studies, automatic discovery of latent topics and finding both predictive and causal relationships across multiple domains of the economy. There have been very little analysis of such possibilities, either from the government or from independent researchers.

Aadhaar in welfare

It does seem intuitively obvious that the unique identification feature of Aadhaar can potentially plug leakages and facilitate better targeting in welfare schemes. Yet several experts with first hand ground level knowledge have severely contested this. Not only have they put forth that this is not happening on the ground as envisaged by the planners but some have also claimed that this can never happen, though the impossibility proofs are unconvincing. In contrast, the position of both the last and the incumbent governments has been that Aadhaar de-duplication can potentially result in huge savings in welfare and other schemes, but detailed analyses and convincing audit reports have been far and few. Proof by forceful assertions is perhaps not the best methodology to decide on social policy interventions?

There have been alarming reports of exclusion and disruptions caused by Aadhaar in social welfare schemes, but these have largely been denied or ignored by the government. These include denial of old age pensions, large scale deletion of names from MNREGA and alarming rates of failure of biometric identity verification resulting in denial of PDS ration. While this is unacceptable, there are hardly any thorough analyses of the reasons for these failures. It is not clear from either the reports or the responses from the authorities whether there are some fundamental reasons for these failures or these are fixable teething troubles. The reports are mostly based on small surveys and anecdotes, from both sides. Such anecdotal studies are definitely suggestive and ought to be taken seriously, but, as they say, the plural of anecdote is not evidence. What is required is a comprehensive audit of the situation.

Is it the case that the failures are due to the biometric technology itself, where to achieve

a low false match rate (FAR) of say 1 in 10000, the false reject rates (FRR) have become unacceptably high? If so, what are the possibilities of application specific tuning of these parameters? For example, is it possible to lower the FRR in social welfare while maintaining the FAR low for dedup? Is it possible that the failures are due to fixable process errors introduced by the local administrations, either due to faulty linking methods or due to faulty presentations during identity verification or that the hardships caused are due to faulty and thoughtless use cases (like making old people travel miles to reach their banks or register their fingerprints)? What are the rates of network or device failures, and is it possible to use local caching of biometric identity verification in a safe and secure way? It seems impossible to assess the viability of Aadhaar in welfare without any detailed analysis of such issues, either from the government or from independent researchers.

Also, apart from technology and processes, there are serious socio-economic concerns that are inadequately answered. What is the need for identity verification with every transaction in social welfare schemes? This appears to be a case of extreme mistrust of the welfare beneficiaries, even considering that there are significant leakages. Where is the analysis that such high frequency identity verification is required and doing so only once or twice a year, as is common in KYCs with more privileged residents, will not suffice? Where is the audit of the large scale deletions of names from MNREGA? Is it indeed the case that there were so many fakes, or were the deletions due to some process errors? It is highly unsatisfactory that there should be such large discrepancies in the versions of the proponents and opponents on issues that do not appear to be that hard to resolve.

Finally, acts of parliament such as NFSA and NREGA define rights to food and employment, and these rights are fundamental and unconditional. Nothing at all should come in the way, be it leakage prevention or any perceived need for improving efficiency. All such requirements are secondary, and the government should exercise extreme care in Aadhaar deployment at least in welfare schemes till all issues are satisfactorily sorted out.

Privacy and security in Aadhaar

It is undeniable that registering every resident in a central database, and recording all their transactions with the state can potentially give the government of the day unprecedented access to information and power over its citizens which can be misused, and unrestricted electronic mass surveillance on such transactions does not augur well for civil liberty and democracy. After all, *emergency* did happen in this country. However, the Aadhaar privacy debate has made little progress. On the one hand the government and the UIDAI have been claiming, rather unconvincingly, that Aadhaar is *safe, secure and robust* from privacy attacks. On the other hand the privacy advocates have been claiming, ad nauseam, that biometric based identity verification and linking Aadhaar UIDs to services must necessarily result in loss of privacy. Both positions appear to be extreme, especially without adequate analysis based on any careful modelling of an attack surface.

In the modern digital era privacy protection does not demand that data should not be collected, stored or used, but that there should be provable guarantees that the data cannot be used for any purpose other than those that have been approved. Indeed, there may be techniques from the areas of cryptography and security that can be used to ensure i) that user data and transaction logs are completely protected from manual inspection even by the maintainers, ii) that all transactions, investigations and analytics can be carried out in a safe way only through audited, pre-approved and provably tamper proof computer programs,

and iii) that the programs are true to the legal and policy frameworks, do precisely and only what they are supposed to do, and maintain tamper proof logs of all authorisation chains and results (see here for a first order analysis, much more is required).

Systems with unbounded specifications cannot possibly be analysed, and it will be helpful if the privacy advocates can carefully study the use cases, identify the vulnerabilities and try to suggest the technological and legal requirements (bounds) that can make Aadhaar safe. Only then it may become possible to ascertain whether such standards can be met or not. Alarming it may sound, but it may not be possible to specify requirements of privacy laws related to digital instruments like Aadhaar without getting into issues such as differential privacy and K-anonymity, but that cannot be an adequate reason for throwing Aadhaar out. Also, as has been pointed out here, privacy concerns related to digital databases should not be restricted only to Aadhaar, and a more comprehensive view of privacy protection is required.

The official position that Aadhaar is perfectly safe is also surprising because it does not appear that the confidence emanates from any careful modelling of the attack surface and evaluation of the design against such a model. The claims that Aadhaar is safe because UIDAI collects minimal data, and the application specific data resides in various application databases and that the Aadhaar numbers have no intelligence built into them are simplistic and untenable. These completely overlook the possibility of privacy attacks through correlation of identities across application domains even if one assumes that the Aadhaar databases are well protected. It is well known that providing aggregate statistical information about the data may reveal information about individuals, and that it is not difficult to identify personal information by linking two or more separately innocuous databases unless special care is taken to prevent such identification (for example, see here and here). Things do appear to be loose, especially because i) there are several violations of UIDAI's own stipulation that the Aadhaar number should not directly be used in any transactions to prevent unlawful identification of individuals, ii) there appear to be no standards, or even guidelines, on how various service providers (AUAs) are supposed to collect and store the Aadhaar numbers, and, as such, the Aadhaar number of an individual becoming public is a definite possibility making it easy to track the individual across application domains, iii) there seem to be a great deal of confusion among both general public and application providers as to whether Aadhaar is a digital identity (which it is supposed to be) or an identity card (which it is supposed to be not); rather than keeping the Aadhaar number fiercely private, they are often being presented printed on cards and embedded in easily readable QR codes, iv) the possibility of insider attacks, perhaps at the behest of some powerful elements within the state machinery itself, doesn't seem to have been considered at all, at least in any publicly available documentation of a threat model, and v) as the recent reports of leakages of biometric data from collection/PoS devices show, there seem to be no provable guarantees that such leakages and biometric replay attacks are not possible. It does appear that the government and the UIDAI need to analyse the situation more carefully and pay more serious attention to privacy protection.

Overall, the Aadhaar debate needs to get much more analytical to be able to make the most out of the initiative and to prevent it from becoming wayward.