Families Europe – COFACE response to the ESA consultation on Big Data

**1.** Do you agree with the above description of the Big Data phenomenon? If not, please explain why. Please also mention whether you consider that other characteristics are relevant to understanding the use of Big Data.

In broad terms, the ESA document describes the phenomenon in a correct way. However, certain issues deserve closer attention.

In point 5, it is interesting that innovation in the field of Big Data is deemed to serve the **sole** purpose of “deriving economic benefit” for the financial service provider, as if other types of benefit which could be essential in justifying it’s broader application, do not exist. For instance, Big Data could in theory, also provide a lot of “social benefit” by helping users to gain insight into their own behavior…

In point 7, and more generally, throughout the paper, the expression “meeting the needs of digital native clientele” or more generally, “meeting the needs of consumers” is constantly used as if its meaning was self-explanatory. It is important to underline, in this respect, that consumer “needs” are not necessarily inherent to consumers and can be also manufactured/implanted by businesses themselves. This has been clear for decades, especially since the essay published by Edward L. Bernays in 1947 entitled “the engineering of consent”[[1]](#footnote-1) and examined by many eminent researchers since then including Noam Chomsky.

Therefore, one must be very careful when assuming that financial service providers are “meeting consumer needs” as if merely passively obeying consumers, which carries an inherently “positive” connotation as opposed to the idea that consumers are being coerced through advertising/marketing/propaganda to “want” and therefore “need” a certain financial product.

In point 13, and at other times in the paper, it is stated that data could be “anonymized” to ensure ethical use of sensitive data. However, many researchers point to ways to “de-anonymize” data. This needs to be made clear in the paper[[2]](#footnote-2).

Finally, in point 15, there are again a great number of assumptions being made about the effect of “barriers to the use of Big Data” such as higher prices for consumers, or reduced choice. These are dogmatic in nature and do not reflect reality. On the contrary, barriers could enable consumers to get better prices by avoiding to falling prey to toxic financial products or being drowned under too many offers. For instance, interest rate caps clearly “limit” the offer of certain financial products, but in a very positive way, by preventing pay day lenders from emerging. In the FSUG paper on simple financial products, the FSUG has underlined that the offer of financial products is already too big for consumers to process and that the sheer number of options confuse consumer as opposed to empowering them which prompted for the creation of more “simple” financial products[[3]](#footnote-3).

**2.** Which financial products/activities are (likely to be) the most impacted by the use of Big Data and which type of entities (e.g. large, small, traditional financial institutions, Fintechs, etc.) are making more use of Big Data technologies? In light of ESAs’ objective to contribute to the stability and effectiveness of the financial system, to prevent regulatory arbitrage, do you consider that there is a level playing field between financial institutions using Big Data processes and those not using them (e.g. because they do not have access to data or the (IT) resources needed to implement Big Data processes) or between established financial institutions and potential new entrants (e.g. Fintechs) using Big Data processes? Please explain.

This question deserves to be reframed. While the “obvious” response to this question is that there is no level playing field since it is skewed on the basis of the access to Data, it is more important to ask on what terms the level playing field should be drawn.

At the moment, the ESAs seem to be concerned that new market entrants in the Fintech business may not have access to the data sets of incumbent banks (financial behavior of consumers, transaction history…) or of the GAFAM (Google, Amazon, Facebook…) and would thereby suffer from an uneven access to the market. The simplistic solution would be to guarantee equal access to data sets across the industry.

But this should **not** be the response from ESAs based on their objective to contribute to the stability and effectiveness of the financial system. The correct question to ask is for what purposes and under which conditions should any entity wishing to sell a financial product or service be allowed to access any type of data. This raises the question of the governance of the use of Big Data and algorithms. To go back to the example of pay day lending, if “leveling the playing field” and allowing any and all financial service providers to access data about consumers without discrimination results in pay day lenders to be able to more easily identify potential “victims” for their toxic financial product, then the ESAs will have failed in their mission of consumer protection. Thus the conditions under which data can be accessed for supporting the development of financial products and services cannot be disconnected from examining those products services and their impact on consumer protection. In the FSUG paper on Big Data, the FSUG has argued for the necessity to set up certain governance rules for the use of Big Data in financial services, and many other fields as well[[4]](#footnote-4). In addition, algorithms should be subject to auditing to review whether they contain inherent biases, whether they are fit for purpose, fair, accurate etc.

4. If you are a consumer or consumer organization, do you witness any of the uses of big data? In what fields?

* The National Consumer Law Council in the US has provided evidence of consumers being excluded from financial services due to mistakes in credit data. Discrimination of the poorest category of the population has also been examined in the news[[5]](#footnote-5).
* Equifax has been convicted in the US for selling consumer data to predatory lenders.[[6]](#footnote-6)
* Credit card details have been a frequent target for hackers[[7]](#footnote-7). Moreover, theft of non-financial data has also been on the rise, such as the recent LinkedIn hack[[8]](#footnote-8). It is not yet clear how data breaches may be monetized by hackers but there are several scenarios which can be envisaged: sale of the account details to third parties, sale of the data to third parties that may have an interest in the data set for data mining/analytics (advertising, predatory lending…), extortion (threatening victims to publicly release embarrassing data), impersonation (attempting to scam close friends and relatives)[[9]](#footnote-9).
* Changing terms of service or user agreement to “force” users into consenting to share their data with third parties for a variety of reasons (advertising…).[[10]](#footnote-10)
* Certain new financial service providers have faced substantial prudential risk such as Lending Club[[11]](#footnote-11).
* Google has recently gained access to sensitive information of patients (health related data) in the UK[[12]](#footnote-12). Sensitive health data is also subject to data theft and/or exploitation by insurance companies[[13]](#footnote-13).
* Online reputation management services have steadily developed over the last years[[14]](#footnote-14).
* Various tracking devices have collected/shared data about users with insurance companies (notably health trackers) with the objective to create individualized risk profiles[[15]](#footnote-15).

**6.** Do you agree with the above short, non-exhaustive, presentation of some of the main applicable requirements? If not, please explain why. Please also mention whether you consider that other legal requirements are essential and should be mentioned.

Additions to the applicable requirements:

* It is essential to add requirements which, although they do not directly target Big Data as such, have an indirect impact on Big Data use. These include:
	+ Interest rate caps and other financial product pricing regulation: by setting interest rate caps on loans, financial service providers cannot fully implement individual risk-based profiling and pricing. Either financial service providers are obliged to discriminate against any consumer whose profile would require (according to their internal, subjective creditworthiness assessment method) a higher interest rate by refusing to sell certain financial products, or by socializing risk by increasing the price of financial products for consumers who have a relatively riskless profile to allow the financial service provider to compensate for the higher perceived risk of more vulnerable consumers (the suitable scenario according to COFACE – Families Europe). Any other pricing regulation would have the same effect in areas such as health insurance, car insurance etc. Setting a “maximum price” or some other forms of regulation would impact the “utility” of Big Data since pricing could only be adjusted to a limited extent (something we see as positive as individual risk-based pricing would lead to massive discrimination).
	+ Right to access basic financial services: the EU Commission has already passed the right to a “basic bank account”. This means that any consumer, regardless of their “profile” have the right to access certain financial products, which are deemed essential for preventing social exclusion. So even if Big Data would suggest that certain metrics like “CLV” (consumer lifetime value) would indicate to the financial service provider that a certain consumer is “not worth” serving, they have no choice but to open a bank account for that consumer.
	+ Financial product ban: even if such a power is rarely used, regulators do have, in theory, the power to ban certain financial products altogether. Thus banning a financial product could also mean banning the Big Data analytics underpinning that financial product (especially if it represents a core part of the financial product).
	+ Anti-discrimination and other broad human rights based regulation: as we have seen since the Test-Achat vs. Council of Ministers case brought to the European Court of Justice, discrimination on the basis of gender has been forbidden. This is an important case law to reference since it could be applicable to many future cases where Big Data is used[[16]](#footnote-16).

In several points, and especially point 21, the document refers to the GDPR and the new sets of obligations that come with it. However, we would like to stress that many of those provisions will either be inoperable or are still very vague and will require considerable case-law and clarification before they become operable at which point, their interpretation may be quite different from the understanding we may have at present.

* Regarding the requirement of entities making use of Big Data to provide consumers with clear and transparent information about the data that is used and the profiling techniques, algorithms behind the scenes, this is already proving to be impossible for several reasons.
	+ Some information about consumers is more easily “explained” than others. For instance, collecting the age and address is rather straightforward. However, collecting mouse behavior, eye tracking, scrolling speed and other indicators like a website’s “heat map” are much more difficult to present in a transparent way[[17]](#footnote-17).
	+ Big Data also generates new, so called “derived” data, which is data that is created by the inferences on available data (for instance a comment about a sporting event on a social network and the likelihood that you like sports). These new forms of data are not directly generated by users but represent a perceived “knowledge” about you which can then be resold without including the user generated data. For instance, “consumer X likes football”, a “derived” data which has been generated from examining the consumer data and making inferences based on it. This data may then take on its “own” life and be used in algorithms regardless of whether its original “inference” methodology was sound or not. Again, it will be difficult if not impossible to fulfil the GDPR’s requirement of transparency, informing consumer or allowing them to “modify” mistakes in their data (since this “derived” data was not generated by consumers in the first place)[[18]](#footnote-18).
	+ Basic algorithms which are looking for statistical correlations in data are already difficult to explain. But algorithms generated automatically using neural networks and other machine learning/AI techniques are incomprehensible, even to those who designed them! There are several examples of this. In one instance, Google managed to instruct an AI to develop its own encryption mechanism. Once it was finished, two AI instances could communicate in an encrypted way without a third AI being able to decrypt the message. The message and encryption in itself was also incomprehensible to humans[[19]](#footnote-19). A second example is the scandal which involved Google’s image recognition system which labeled black people as “gorillas”[[20]](#footnote-20). This decision by Google’s image recognition algorithm could not be corrected and had to be shut down, simply because the algorithm was created automatically by “learning” from a pool of pre-labelled images. Thus if an self-learning algorithm has made a decision based on a certain “pool” of data, it is impossible to “change” the algorithm to make an alternative decision or make an exception for a **specific case.** To make the algorithm change its decision would imply retraining it with another data pool or reprogramming it altogether which would change **all** of its decisions.
	+ Current credit assessment techniques are not disclosed to consumers. Companies like Equifax or Experian do not share their algorithms and credit scoring methodology. It is doubtful whether these companies will be ready to disclose them once the GDPR enters into force.
	+ Finally, this provision does not address issues such as a single device being used by multiple individuals (like a family computer) where information generated automatically and not easily “editable” or “corrected” comes from multiple people. This touches on a broader issue of “data pollution” which we will come back to later, which encompasses situations where data about a user is “polluted” by incorrect or false data. This can be due to many reasons including a single device used by multiple individuals, hacking of a user’s devices which may corrupt the data generated by that user or even the user him/herself either intentionally or unintentionally feeding incorrect information about him/herself (for instance, GPS spoofing, bot generated form filling etc.)
* Regarding the right to data portability, this could also play against consumers. The consumer is very often the weakest link as he/she is in demand of a financial product or service and thereby, is often willing to sign or agree to anything in order to access the desired product/service. Giving consumers more “control” over their data could simply result in the further spread and sharing of consumer sensitive data. Consumers would give their consent to port their data from one provider to another in exchange for accessing certain financial services. An example of such a phenomenon is the Belgian public credit register which was designed to limit the access to consumer’s financial data. However, financial intermediaries in Belgium, who do not have the right to access such data, simply ask consumers for a print-out of their public credit register data as a pre-condition for accessing their services[[21]](#footnote-21). COFACE-Families Europe is therefore underlining the importance of governance of big data, and defining “rights” to data and protection to consumer data which cannot be waived by consumers (similar to human rights which cannot be violated even if a person renounces these rights).
* Regarding the right to ask for a human being to intervene in the profiling, this is also very vague and will require clarification by EU Institutions/EU courts. It could simply mean a “manual” approval of the algorithmic result which would equate to rubber stamping without any “formal” review. The opposite, which would be a thorough review of the algorithmic decision, could be impossible to implement in practice due to the impossibility of humans to “understand” decisions which are made by machine learning or AI (see above).
* Regarding the right to object to processing, this could also be a provision which will be inapplicable in practice, especially regarding credit. Both the Mortgage Credit Directive and the Consumer Credit Directive include the necessity for financial service providers to carry out creditworthiness checks, which require data processing to be carried out, thereby exempting financial service providers from this requirement. Even if consumers were able to enforce this provision, this could simply lead to discrimination from accessing financial services/products, which would defy the purpose of enforcing it from the point of view of the consumer.
* Finally, regarding the necessity to “prove” that the data and the profiling being carried out is “necessary” for providing the financial product/service, there is no difficulty in seeing that financial service providers and especially private credit bureaus or any actor in the business of selling services based on data collection/processing will provide ample “evidence” to justify their actions. ACCIS which represents, among others, private credit registers such as Experian and Equifax, have always argued that the more data they have the more “accurate” is the creditworthiness assessment (see the responses of ACCIS to the FSUG paper on creditworthiness assessment).

Point 21 finishes with a sentence which encourages “big data analytics by using anonymized or pseudonymized data”. However, such an encouragement goes against the direct interests of financial service providers. Using anonymized data is only useful to predict “general” trends and is therefore only useful for financial products which socialize risk fully among an entire population (for instance, health insurance mutualities). Using anonymized data would prevent financial service providers from carrying out individual risk based pricing. If, however, financial service providers are ready to shift their sales strategy from individual risk based pricing to socialization of risk across their entire consumer base, COFACE-Families Europe would happily welcome and support such an initiative.

In point 30, there is the mention of the principle of acting in the “best interests of clients” along with other expressions such as “understanding their client base”. These notions will have to be revisited in light of Big Data analytics.

As the paper has rightfully pointed out, the Unfair Commercial Practices Directive forbids unfair commercial practices which are likely to distort the behavior of consumers. Big Data in the form of targeted advertising, dynamic pricing/price optimization[[22]](#footnote-22) or simply due to its disciplinary power[[23]](#footnote-23) could fall under this provision and thereby nullifying such principles as acting in the “best interest of a consumer” or “understanding a consumer” since the consumer will have been manipulated. The major problem is that it will become very difficult to identify such practices, especially if they are carried out by autonomous, self-learning algorithms which leave no trace of the ads or price fluctuations they have displayed. (See the article from Le Monde referenced above)

Examples of such practices can be found in various products already including airline fares and Uber fares based on a users’ battery life (consumers with a lower battery life are willing to pay more)[[24]](#footnote-24). In the US, a study has found that over a quarter of insurers use price optimization in some manner[[25]](#footnote-25).

**8.** Do you consider the potential benefits for consumers and respectively financial institutions to be accurately described? Have you observed any of them in practice? If so, please provide examples. If not, please explain whether you are aware of any barriers that may prevent the above potential benefits from materialising?

Overall, we do agree with the potential benefits, but some of those benefits are less likely to materialize due to reasons explained below. There are also some additional benefits to the use of Big Data which we would like to point out.

In point 50, another potential benefit of Big Data, if it is developed, is for consumers to gain insight into their own online behavior, how much is known about them, what kind of ads (among which ads for financial services) would be shown to them, what is their risk profile according to different metrics/algorithms and data sets for various financial products/services and why (for instance, that they drive too fast, brake too hard, or that they are on an unhealthy eating diet). Such services would only serve as a way for introspection and insight into ones’ habits and lifestyle choices without the “punitive” element of being excluded from accessing financial services or paying discriminately high risk premiums. One of the most interesting elements would be to help consumers understand their “commercial” profile based on their data so that they are more resilient to aggressive advertising, making them realize that depending on their personal financial situation, they may be targeted for commercial purposes which go against their best interest (for instance, being targeted by pay day lenders if they often fail to pay their bills, being targeted by loan consolidation proposals if they have several loans and missed payments on these…) Furthermore, it could help consumers “break out” of certain behavior patterns which Big Data coupled with advertising has locked them into. Social networking and search engine algorithms have come under fire in recent years due to the so called “filter bubble” effect, where users, being shown more of what they like already, are prevented from discovering new content/ideas/products. But Big Data can also serve the opposite purpose, depending on how the algorithm is configured. For instance, users could choose an algorithm which shows them specifically new content which is unrelated to what they already like/know. In a similar way, algorithms in financial services could help consumers break out of certain financial/spending patterns by making them more apparent to consumers.

In point 53, the potential benefit may be less likely due to the impossibility to reach transparency/intelligibility regarding the decision making rationale behind algorithms (as was pointed out above). While it will be easier to maintain records of automated decision making outcomes, it will be near impossible to justify the reasons and motivations behind such automated decisions.

Point 63 and more generally, the idea that Big Data will bring about “new” business models and therefore might address banks’ falling profitability is dubious given the low economic growth environment. In a near zero growth environment, profits are mostly temporary especially if they are financial products, since such products depend on a healthy economy and do not create wealth in and of themselves, and therefore cannot be deemed sustainable.

One additional benefit of Big Data which is not mentioned in the report, is the possibility to design algorithms which can audit/test other algorithms for bias, discrimination or other principles. Thus government agencies including the ESAs could design algorithms to test whether the solutions set up by financial service providers respond to certain principles like anti-discrimination, fairness, etc. Such a tool already exists and has been designed for the Zest Fintech. “To test for bias, Zest relies again on machine learning, which the system uses to test its own results. It applies an algorithm the Consumer Financial Protection Bureau uses to check for discrimination, and also does other testing to find any unexpected correlations with factors that lenders are prohibited from considering. “[[26]](#footnote-26) This could be a best practice moving forward in testing algorithms developed by financial service providers.

**9.** Do you agree with the description of the risks identified for consumers and respectively financial institutions? Have you observed any of these risks (including other risks that you are aware of) causing detriment to consumers and respectively financial institutions? If so, in what way? If not, please explain why. Please also mention whether certain risks for consumers and financial institutions have not manifested yet but have the potential of developing in the future and hence need to be closely monitored by Supervisory Authorities.

Overall, the risks are described accurately. Here are a few additional points we would like to raise.

In point 44, we fully support the risk identified by the paper, namely that some consumers may have been sold a product/service without really needing or fully understanding the implications. Such is the case for young people (so called “millenials”), who are more willing to disclose their private data in exchange for a “discount” or a promise of a better, more “personalized” experience with regards to their (financial) product/service provider[[27]](#footnote-27). For these young people, it may be an advantage in the short term, but not necessarily in the long term and thus represents a clear situation whereby the inability to fully understand the implications of a certain course of action (namely sharing their data).

Point 60 discusses checking data quality to ensure that it is reliable/suitable for use by financial service providers. However, some forms of consumer data may become unusable if the industry fails to secure the trust of consumers. Data masking techniques, such as substitution or shuffling of data, are applied to protect data sets from hacking or theft. However, such techniques could be used by consumers themselves to “pollute” their personal data to ensure that it cannot be used for commercial purposes. Data masking is the process of mixing up a data set to protect its content and render it unintelligible should it be stolen without the proper “key” to unscramble the entries. The process of substitution or shuffling means that data entries are reordered inside the data set to “look” genuine, but in reality, the content has been scrambled and is no longer accurate. Similar techniques could be developed to “pollute” consumer data sets for instance by “faking” a browsing behavior, filling online forms with made up information or even auto-posting things on social media (social media bots have already been developed in the past). This could turn into a business opportunity for companies, selling tools to allow consumers to manipulate their credit scores. Online management reputation companies already exist and some are already proposing their services to help consumers get a better credit score[[28]](#footnote-28)!

The purposeful targeting of vulnerable consumers for fraud is also a clear risk. The FTC report on Big Data has flagged that certain companies may obtain lists of consumers which are more receptive to certain forms of enticements or “suffering seniors” who have Alzheimer’s or other such conditions to target them with toxic financial products[[29]](#footnote-29).

Another more general caveat is the argument that using “more data” is always better and more accurate for creditworthiness/product customization, or that so long as the underlying data is of good quality, then the results from Big Data analytics are trustworthy. It would seem that the methodologies of various financial service providers using Big Data arrive to contradictory conclusions which is clearly a risk for consumers and ensuring that they are treated fairly.

In an article from the Wall Street Journal, one Fintech using mobile data arrived at the conclusion that gambling is a “good risk”[[30]](#footnote-30) whereas another Fintech called Zest identified gambling as being “bad risk”[[31]](#footnote-31). Which is it then? Furthermore, Zest in the same article points out that using data from Social Media is “not very useful”, whereas companies like Kreditech operating in Germany rely heavily on data from Social Networks to carry out their creditworthiness assessment[[32]](#footnote-32). Finally, even data which has been used for a while by incumbent credit registers like Experian or Equifax may prove to be misleading. Such is the case for rental or utility bill payment history. The FTC report on Big Data has underlined that consumers have the right to withhold rent or utility payment if landlords do not provide heating or basic sanitation services. In these instances, the failure to pay on time is not due to financial problems[[33]](#footnote-33).

**10.** Is the regulatory framework adequately addressing the risks mentioned above? Bearing in mind the constant evolution of technologies/IT developments and that some of the above mentioned regulatory requirements are not specific to the financial services sector (e.g. GDPR), do you think further regulation is needed to preserve the rights of consumers of financial services in a Big Data context? Please explain why.

We do not believe that the regulatory framework in addressing the risks is appropriate. Much of the GDPR, as explained already, will either require substantial clarification (which entails legal uncertainty and therefore cannot guarantee consumer protection) or will be virtually impossible to implement given the way Big Data and especially, the algorithms underpinning it, evolve.

Our position is that specific governance bodies need to be created to deal with the risks stemming from Big Data in general, and in particular in financial services. These governance bodies should have a mandate to intervene directly in financial product development/approval/ban based on a very clear set of principles (for instance, non discrimination in accessing financial services, fair price/fair conditions, etc) and should have the mandate to audit the algorithms and data sets used by financial service providers. Examples of regulations which include overarching principles applied to all financial services include the United States Community Reinvestment Act which requires financial service providers to prove that they serve the interests and needs of the communities where they are established.

With regards to making regulation future proof or measuring its impact ex-ante, we do not believe that this is feasible or realistic. Policy making and regulation should seek to be as future proof as possible, but will inevitably need to be reassessed and adjusted based on observed changes due to innovation in the regulated sector. Innovation is not something that should be reserved to the private sector, it is just as necessary in policy making. In the case of Big Data and data analytics, we defend the view that a combination of both technical measures (employing algorithms to audit other algorithms) and overarching principles (anti-discrimination laws) should be used concomitantly and assessed on their effectiveness (observed impact on financial inclusion/exclusion, levels of indebtedness, consumer complaints etc), followed by an adjustment of the regulation to fill the gaps. While some would point that a predictable and stable regulatory environment is essential for financial stability, let us remember that the recent financial crisis were caused by under-regulation rather than over-regulation which have spread moral hazard in financial services. A stable and predictable regulatory environment is impossible given the disruptive and fast-paced innovation of the financial services sector. What **should** be predictable and stable are the underlying **values and principles** underpinning the regulatory framework which will make future regulations more predictable. For instance, imposing a ban on pay day lending should **not** have come as a “surprise” to pay day lenders, based on the values and principles of consumer protection regulation.

**11.** Do you agree that Big Data will have implications on the availability and affordability of financial products and services for some consumers? How could regulatory/supervisory authorities assist those consumers having difficulties to access financial services products?

We strongly believe that simple and clear regulations will provide much more legal certainty and a level playing field than very technical and special purpose regulations. We have already mentioned interest rate caps which are very clear and easy to set up and directly affect the products consumers are sold. A similar set of regulations should be put in place in financial services, restricting the scope of individual risk based pricing for instance, to ensure that some form of risk mutualization/socialization is safeguarded in the financial products being sold to consumers. Again, these are rather simple regulations but which will have a much deeper impact than very sophisticated regulations that deal with algorithms, data quality, legitimate uses of data processing and so forth. If a financial service provider cannot sell a product because it is clearly overpriced under the conditions of a single blanket regulation, then that financial service provider has no need/incentive to make use of costly (unethical) Big Data analytics solutions for assessing individual risks.

A further advantage of such blanket regulations is that financial service providers will use Big Data for more accurately hedging their systemic/global risk. Since they will not be able to fully discriminate users based on their personal profile, financial service providers will have to focus on ways in which they can mitigate the risk at a global level, increase their overall product efficiency. It will also favor prudential regulation as it is easier to manage global socialized risk over a larger pool of consumers than to manage an addition of more or less accurately calculated personal risks, which introduce a much higher likelihood of errors.

Finally, many financial service providers already engage in “positive” uses of Big Data within a risk socializing system. Health mutualities in Belgium for instance, offer a number of rebates for subscription to fitness memberships and carry out targeted campaigns to incentivize people from living more healthy lives, thereby focusing on preventing risk rather than a punitive/discriminatory approach where risky behavior is assorted with discriminatory prices/practices.

**12.** Do you believe that Big Data processes may enable financial institutions to predict more accurately (and act accordingly) the behavior of consumers (e.g. predicting which consumers are more likely to shop around, or to lodge a complaint or to accept claims settlement offers) and could therefore compromise the overarching obligations of financial institutions to treat their customers in a fair manner? Please explain your response.

As explained above, the predictive capacity of Big Data can be strongly damaged simply by tampering with the underlying data quality.

But there are other reasons why the predictive capacity of Big Data is an illusion.

First, Big Data is an inherently conservative vision of consumer behavior. It does not necessarily react quickly to changes in situation or mindset, especially at an age where consumers experience major life events which may radically change the way they consume/think etc.

Some examples which are not related to financial service providers but are nevertheless interesting include mothers which have lost their babies due to miscarriage and continued to receive advertising for baby products. Clearly, the algorithm was not designed to deal with their “change” in situation[[34]](#footnote-34). This is of course, very damaging psychologically for women. One could imagine a similar pattern for over-indebted consumers who would receive advertising for solving their debt problems even if they have managed to get over their financial difficulties or more troublesome, credit scores which do not adapt to a consumers’ present financial situation after suffering a hardship.

Furthermore, Big Data is based on a myriad of human grounded biases and assumptions which will inevitably lead to inaccuracies in their predictive powers. For instance, besides the fact that correlation does not mean causation, setting the statistical significance of a correlation and the assorted decision inside the algorithm is arbitrary[[35]](#footnote-35). The same with regards to the data which is examined and its relative “weight” in the process. What is more important: the level of income? Spending/saving behavior? Past behavior with regards to servicing debts? If a consumer has behaved irresponsibly in the past but has shown responsible financial behavior in the last year, how is that interpreted by the algorithm? This is true for human designed algorithms of course. The situation is even worse for algorithms which learn “automatically” from data via machine learning or other AI techniques. In these instances, it is impossible to measure its accuracy. It is akin to carrying out “live” social experimentation on society. And should algorithms “fail”, it might be less “obvious” but might have much more serious implications for society than Google image’s mistake which labeled black people as being gorillas.

Another cause of concern is making consumers pay for global phenomenon such as extreme weather and natural catastrophe due to climate change. As the technology to help predict and assess the impact of climate change in a granular level (by neighborhood or even individual houses), the risk is that individuals will be asked to pay higher premiums for phenomenon which are the result mostly of failed national/international policies and of private companies (individual consumption accounts for a small fraction of climate change whereas pollution from the private sector represents the biggest contribution)[[36]](#footnote-36).

Finally, Big Data is inaccurate and especially, unfair in its predictive power since it does not take interest in the reasons for consumer behavior. Big Data, from a philosophical point of view, pushes the responsibility for certain behaviors solely on consumers as if consumers were entirely in control of their lives and decisions.

It is no mystery that human beings are complex, and their individual behavior can be traced to countless factors such as their upbringing, their education, their social circle (close friends and family) etc. Big Data in that sense makes people pay for the factors which have influenced their behavior, even though these factors were outside of their control.

To give an example opposite of the way Big Data measures risk at present, we could imagine the following “alternative” Big Data analytics tool for calculating insurance premiums: an algorithm which is aimed at calculating the health risk of an individual for a health insurance would measure that individual’s exposure to unhealthy food marketing and would assess whether there is a correlation between such exposure and the individual’s eating habits[[37]](#footnote-37), especially if his eating habits over time have deteriorated due to such exposure to unhealthy food marketing. The food companies which were responsible for the advertising would then have to pay the extra premium to cover for the additional health risk stemming from unhealthy eating habits. Many studies, including the one quoted below, have underlined the correlation between advertising and eating habits, including the WHO[[38]](#footnote-38)! In essence, aggressive food marketing shapes consumers’ eating habits from an early age, and later on, when comes the time to seek health insurance, due to the data they have generated which points to an unhealthy eating habit, consumers pay extra health insurance premiums. Cynicism at its finest.

The same can be said about a myriad of other things. An over-indebted consumer due to overuse of credit? Measure his/her exposure to ads from shark loans or even the global consumerist environment which pushes for impulsive/frivolous spending instead of responsible and sound financial behavior.

The way Big Data is designed carries an important philosophical/ideological dimension which is that responsibility lies solely with the consumer for each and every decision (translated into data) that he or she makes. Big Data pre-supposes absolute “free will”. But this means glossing over the mounting evidence from psychology, psychometrics[[39]](#footnote-39), sociology, behavioral economics and countless other fields that human decisions are not made in a vacuum and that the context (over which a consumer has very little control) plays a big part in the decisions effectively made. Psychometrics especially, made the headlines due to their direct use in the Brexit campaign by the “leave” camp and by Donald Trump during his election campaign[[40]](#footnote-40).

To summarize the impact of this growing field of research, psychometrics allows to match very specific psychological profiles carried out by psychologists via lengthy, specialized questionnaires, and match them to patterns of behavior on social networks. This technique allows to determine the psychological profile of **all** users who frequently use social networks without the need for them to fill in medical questionnaires thanks to the matching of voluntary subjects who filled in the questionnaire and gave access to their social media accounts thereby enabling algorithms to “match” social networking behavior and infer a psychological profile from it. From a marketing perspective, this means displaying an ad which will directly leverage the imagery or messages which will resonate emotionally with consumers, thus encouraging behaviors not based on rationality but impulsiveness. Since it has only been applied twice, very little evidence can be gathered at this stage to measure the impact on consumer/voter behavior, but hanging on to the idea that consumer decisions are made out of their own “free will” with no third-party interference is crazy.

These issues have been covered by studies from major players in the Big Data and Artificial Intelligence field such as Microsoft’s chief economist Susan Athey[[41]](#footnote-41), underlining the necessary caution in using Big Data for making predictions and especially, devise policies based on such predictions.

**14.** Would you see merit in prohibiting the use of Big Data for certain types of financial products and or services, or certain types of customers, or in any other circumstances?

Yes, for all the reasons explained above.

Besides, some of these prohibitions already exist. As discussed above, the Test-Achat vs. Council of Ministers case clearly prohibits the use of gender as a parameter in car insurance premium calculations.

In the US, some States impose such bans on the basis of non-discrimination[[42]](#footnote-42). For instance, the use of marital status in setting the price of auto-insurance is prohibited in Hawaii, Massachusetts, Michigan and Montana. Other states prohibit such an indicator for homeowners insurance. The justifications for such bans are grounded in moral arguments rather than technocratic ones: “Is it fair for the victim of an abusive spouse who obtains a divorce to face higher auto insurance premiums for ending a dangerous relationship? Is it fair for a widow (or widower) to pay more for auto insurance after experiencing the loss of a spouse?”[[43]](#footnote-43)

Another practice which is forbidden is “price optimization” which adapts the price based on the consumers’ willingness to pay more or less (based on Big Data)[[44]](#footnote-44).

**15.** Do you agree that Big Data may reduce the capacity of consumers to compare between financial products/services? Please explain your response.

Yes. This has already been covered in the paper on Big Data which COFACE-Families Europe has drafted for the FSUG. Since offers would be tailored to a consumer, it would be very difficult to compare a similar product from another financial service provider since the data sets/algorithms used might differ. For insurance products, for instance, Big Data could automate the process of limiting the liability of insurers based on consumer data. For instance, if a consumer has been found to frequently play Basket-ball, health insurance would not cover broken fingers. On car insurance, if the consumer is found to brake very brutally, the insurance could exclude changing the brakes. And so depending on the data and the algorithm applied, offers could wildly differ.

**16.** How do you believe that Big Data could impact the provision of advice to consumers of financial products? Please explain your response.

The inherent vested interest in the advice given to a consumer will be harder to recognize. For instance, in a “budget management” app, what would be the thresholds selected for saving versus spending? If a consumer is to buy a product, will he/she be advised to buy such a product using a consumer credit or pay for it with his/her savings? Regarding savings and investment decisions, what will be the benchmarks and projections used to advise on a sound investment strategy? For instance, Better Finance publishes each year a book on the real returns of pension funds[[45]](#footnote-45). According to their research, many pension funds have a negative return once all costs have been deducted. Thus objectively, these products should not be recommended to consumers! Needless to say that the financial providers responsible for selling these products would not refrain from advising them.

Thus any and all automated advice will carry a number of assumptions and a “model” of what the “optimal” consumer financial behavior should be. And it is doubtful that the advice will ever encourage consumers to fight consumerism, spend as little as possible, save as much as possible, use credit on very rare occasions etc as this goes directly against profit maximization or other metrics such as CLV (consumer lifetime value).

The FSUG has also published a research on financial guidance which we would encourage the ESAs to read as it also discusses the issue of liability in automated advice[[46]](#footnote-46).

**17.** How do you believe Big Data tools will impact the implementation of product governance requirements? Please explain your response.

Product governance will be much harder to implement under the current regulatory framework, which is why we would encourage to move to a simpler regulatory framework as advocated for above.

**18.** How do you believe Big Data tools will impact know-your-customer processes? Please explain your response.

It will depend on the underlying quality of the data. As explained above, Big Data does not adapt to changes in situations very easily and also, data can be manipulated (online reputation management). Therefore, “traditional” KYC processes need to be upheld (official ID…) rather than relying on Big Data.

**19.** What are key success factors for a Big Data strategy (i.e. the adaptation of the business model/plan towards Big data driven technologies and methods)?

* A clear and simple regulatory/governance framework assorted with auditing of algorithms Big Data.
* A clear understanding of what Big Data should do and can do (what are the limits of Big Data).
* Addressing the identified risks and especially, the risk of discrimination and unequal access to financial services/products.

**22.** How do you see the development of artificial intelligence or blockchain technology in connection with Big Data processes?

Artificial Intelligence is a revolutionary and potentially dangerous development for our societies based on how it is currently used because it is inherently conservative. Since AI does not have a creative capacity, it only “spits out” decisions based on a given data set. It does not have the capacity to reflect on how to “transform” the situation (data set) to arrive at a different outcome. For instance, AI will be able to justify decisions to discriminate people that are poor or have had troubles managing their finances, but not provide insight as to the solutions to solve poverty.

To give a concrete example, AI is now used in predictive policing in the US. While it does help predict where crime is most likely to happen, it does not say anything about the causes of crime or how to prevent them, only to “catch the culprits” to prosecute them and throw them in jail instead of investing in the community, create jobs, fight discrimination and poverty… The dangers of AI lie in the value and trust humans give it, their belief that the situation identified by AI is an unalterable fact, and the human interpretation and actions taken on the basis of the results it yields. AI could therefore help cement certain discriminations or situations rather than providing impetus for change.

As regards blockchain technology, this does not necessarily link up with algorithms and Big Data. It is a technology which will mostly allow to improve the security and quality of data by making a database (which is then used for data analytics) resilient to unauthorized changes. However, more importantly, blockchain may very well enable peer-to-peer financial services without the need for third-party intermediaries and that could have a much bigger impact on the financial services sector. Blockchain technology might enable a “parallel” economy to develop, with transactions, exchange of goods and services, even contracts being done outside of the scope and reach of regulators. Hopefully, such a technology will enable two things: first, it will put more pressure for ensuring a fair and stable “official” financial system as a failure to do so will drive people towards a decentralized blockchain/virtual currency solution (a phenomenon observed already in Venezuela and akin to the “flight to gold” in the past), second, it might help stabilize the financial system altogether by providing a pluralism of currencies as it has been studied in Switzerland with the Swiss franc and the WIR[[47]](#footnote-47).

In that light, regulators should also avoid falling prey to the propaganda of those seeking to protect their private interest by blaming such technology as means of financing terrorism, money laundering or purchase of illegal drugs and other black market items. One should not forget that the current financial system still remains the biggest contributor to all of these things. Terrorism financing by the Saudi regime has been known for quite some time yet nothing has been done to curtail it[[48]](#footnote-48), major mainstream banks have been caught in money laundering scandals[[49]](#footnote-49), some even directly linked to drug cartels[[50]](#footnote-50). The priority is to protect consumers and make sure that any peer-to-peer solution developed via blockchain is robust and resilient to fraud and scams affecting consumers directly.

On the other hand, certain economists and academics have pointed to the inherent dangers of “apolitical money” and that virtual currencies could turn out to be mere Ponzi schemes[[51]](#footnote-51). Such dangers need to be taken seriously by regulators.

1. <http://www.mcnuttphysics.com/uploads/2/3/6/9/23694535/engineering_of_consent-edward_l_bernays.pdf> [↑](#footnote-ref-1)
2. <https://www.cs.cornell.edu/~shmat/shmat_oak08netflix.pdf> [↑](#footnote-ref-2)
3. <https://ec.europa.eu/info/file/46842/download_en?token=6PKyKy_a> [↑](#footnote-ref-3)
4. <https://ec.europa.eu/info/file/46888/download_en?token=Mgdghb0P> [↑](#footnote-ref-4)
5. <http://www.theatlantic.com/technology/archive/2016/04/how-big-data-harms-poor-communities/477423/> [↑](#footnote-ref-5)
6. <https://www.worldprivacyforum.org/2013/03/public-comments-letter-to-ftc-re-equifax-sales-of-consumer-info-to-predatory-lenders/> [↑](#footnote-ref-6)
7. <http://www.zdnet.com/article/how-hackers-stole-millions-of-credit-card-records-from-target/> [↑](#footnote-ref-7)
8. <http://money.cnn.com/2016/05/19/technology/linkedin-hack/> [↑](#footnote-ref-8)
9. <http://personalweb.about.com/od/makefriendsonfacebook/qt/facebkscammoney.htm> [↑](#footnote-ref-9)
10. <https://www.paypal.com/be/webapps/mpp/ua/upcoming-policies-full?locale.x=en_US> [↑](#footnote-ref-10)
11. <http://www.investors.com/news/how-will-lending-clubs-problems-affect-2016-fintech-funding/> [↑](#footnote-ref-11)
12. <https://www.theguardian.com/technology/2016/may/04/google-deepmind-access-healthcare-data-patients> [↑](#footnote-ref-12)
13. <https://www.theguardian.com/lifeandstyle/2015/apr/17/fitness-trackers-wearables-insurance-employees-jobs-health-data> [↑](#footnote-ref-13)
14. <http://www.latimes.com/business/la-fi-lazarus-20140815-column.html> [↑](#footnote-ref-14)
15. <http://www.smh.com.au/digital-life/digital-life-news/data-collection-wearable-fitness-device-information-tracking-your-life-20150416-1mmzbq.html> [↑](#footnote-ref-15)
16. <http://curia.europa.eu/juris/celex.jsf?celex=62009CJ0236&lang1=en&type=TXT&ancre>= [↑](#footnote-ref-16)
17. <https://heatmap.me/> [↑](#footnote-ref-17)
18. <https://www.treasury.gov/initiatives/fio/reports-and-notices/Documents/2016_FIO_Consumer_Report.pdf> p. 10. [↑](#footnote-ref-18)
19. <https://techcrunch.com/2016/10/28/googles-ai-creates-its-own-inhuman-encryption/> [↑](#footnote-ref-19)
20. <http://www.usatoday.com/story/tech/2015/07/01/google-apologizes-after-photos-identify-black-people-as-gorillas/29567465/> [↑](#footnote-ref-20)
21. Interview with the Test-Achat representative, sitting on the board of the Belgian Public Credit Register, 2015. [↑](#footnote-ref-21)
22. <http://internetactu.blog.lemonde.fr/2017/02/11/comment-prouver-les-pratiques-anticoncurrentielles-a-lheure-de-leur-optimisation-algorithmique/> [↑](#footnote-ref-22)
23. In some retail stores, Big Data is being used to measure the “performance” of the workers with metrics such as the “shoppers’ yield”. The existence of such metrics puts pressure on workers to behave in a certain way. Thus Big Data has disciplinary power. The same principle can apply to algorithms which determine the right/conditions for accessing certain financial products/services which will push consumers to act in a certain way in order to be judged favorably by algorithms. [↑](#footnote-ref-23)
24. <http://www.telegraph.co.uk/business/2016/05/22/uber-app-can-detect-when-a-users-phone-is-about-to-die/> [↑](#footnote-ref-24)
25. <http://www.insurancejournal.com/news/national/2015/11/17/389153.htm> [↑](#footnote-ref-25)
26. <https://www.technologyreview.com/s/603604/an-ai-fueled-credit-formula-might-help-you-get-a-loan/> [↑](#footnote-ref-26)
27. <https://ec.europa.eu/info/files/fsug-annual-report-2016_en> p. 23. Presentation by Ernst and Young at the FSUG Berlin meeting in June 2016. [↑](#footnote-ref-27)
28. <https://www.reputationdefender.com/blog/job-seekers/your-online-reputation-score-your-new-credit-score> [↑](#footnote-ref-28)
29. <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf> p. 10. [↑](#footnote-ref-29)
30. <https://www.wsj.com/articles/lending-startups-look-at-borrowers-phone-usage-to-assess-creditworthiness-1448933308> [↑](#footnote-ref-30)
31. <https://www.technologyreview.com/s/603604/an-ai-fueled-credit-formula-might-help-you-get-a-loan/> [↑](#footnote-ref-31)
32. <https://www.kreditech.com/> [↑](#footnote-ref-32)
33. <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf> p. 30. [↑](#footnote-ref-33)
34. <http://www.destinationcrmblog.com/2014/02/10/accidentally-marketing-miscarriage/> [↑](#footnote-ref-34)
35. A good example of this is the correlation made by Google to predict Flu outbreaks. Whereas the machine-learning algorithm appeared to predict flu outbreaks by region based on flu search terms, it failed to be accurate over time due the fact that people were more likely to search for the term “flu” if the local news ran a story about a flu outbreak. <https://www.ftc.gov/system/files/documents/reports/big-data-tool-inclusion-or-exclusion-understanding-issues/160106big-data-rpt.pdf> p. 11. [↑](#footnote-ref-35)
36. <https://www.treasury.gov/initiatives/fio/reports-and-notices/Documents/2016_FIO_Consumer_Report.pdf> p. 14-15. [↑](#footnote-ref-36)
37. <http://psycnet.apa.org/journals/hea/28/4/404/> [↑](#footnote-ref-37)
38. <http://www.euro.who.int/__data/assets/pdf_file/0017/322226/Tackling-food-marketing-children-digital-world-trans-disciplinary-perspectives-en.pdf?ua=1> [↑](#footnote-ref-38)
39. <https://www.gsb.stanford.edu/faculty-research/publications/computational-personality-recognition-social-media> [↑](#footnote-ref-39)
40. <https://motherboard.vice.com/en_us/article/how-our-likes-helped-trump-win> [↑](#footnote-ref-40)
41. <http://science.sciencemag.org/content/355/6324/483> [↑](#footnote-ref-41)
42. <https://www.treasury.gov/initiatives/fio/reports-and-notices/Documents/2016_FIO_Consumer_Report.pdf> p. 12. [↑](#footnote-ref-42)
43. *Ibid,* p. 23. [↑](#footnote-ref-43)
44. *Ibid,* p.10. [↑](#footnote-ref-44)
45. <http://betterfinance.eu/fileadmin/user_upload/documents/Research_Reports/en/Pension_report_2016_For_Web_-_Final.pdf> [↑](#footnote-ref-45)
46. <https://ec.europa.eu/info/file/55556/download_en?token=3K0_lWIU> <https://ec.europa.eu/info/business-economy-euro/banking-and-finance/financial-reforms-and-their-progress/regulatory-process-financial-services/expert-groups-comitology-and-other-committees/financial-services-user-group-fsug/fsug-studies-and-papers_en> [↑](#footnote-ref-46)
47. <https://www.veblen-institute.org/IMG/pdf/currency_pluralism_and_economic_stability_eng_oct_2011_.pdf> [↑](#footnote-ref-47)
48. <http://www.politico.com/magazine/story/2016/09/saudi-arabia-terrorism-funding-214241> [↑](#footnote-ref-48)
49. <http://www.thetimes.co.uk/tto/business/industries/banking/article4700489.ece> [↑](#footnote-ref-49)
50. <https://www.theguardian.com/commentisfree/2015/feb/15/hsbc-has-form-mexico-laundered-drug-money> [↑](#footnote-ref-50)
51. <https://www.yanisvaroufakis.eu/2013/04/22/bitcoin-and-the-dangerous-fantasy-of-apolitical-money/> [↑](#footnote-ref-51)