PayPal Under The Microscope

Research Dossier

Background Information and Additional Resources

September 2022

Carmen Amme
Alicjia Suchonska
Regilla Bastian
Anneke Büürma

Table of Contents

1 PayPal: Business Structure

1.1 PayPal's core business model

Online payment providers such as PayPal process rich transaction data, and they are the intermediaries between merchants and their customers. PayPal supports over 400 million active consumers and merchants worldwide. Every minute there are several thousand payment transactions. The relationships in this payment network are strong predictors of behaviors and great risk indicators that can reveal fraud.

There are 426 million active accounts on PayPal. In the year 2021 there were 19.3 billion payment transactions made using this service. They all added up to 1.25 trillion US-dollars.

In a study by Preibusch et al. (2016), 881 US web shops were analysed and the authors found that more than half of them share product names and details with PayPal, even about sensitive products such as medication and sex toys. The minimum data requirements for payment handling are the order total, the receiving merchant and an authenticated payment method. However, a much richer set of data items becomes available for online purchases, including an itemized bill or information about the buyer. Collecting data that is not necessary to process payment transactions is also a financial decision, because it is expensive to store data of 426 million PayPal users. This means that the collected data must be used in some way that brings profit to PayPal.

Data privacy concerns are here not only about the personal information, but also about consumer behavior. Payment providers are able to collect and connect purchase details at large scale to create a behavioral character of the consumer. This allows PayPal to build up comprehensive consumption profiles across the sites consumers buy from, subscribe to, or donate. Preibusch et al. (2016) further found that PayPal forwards shopping details to Omniture, a third-party data aggregator with an even larger tracking reach. PayPal also sends browser properties such as plugins and screen dimension, and therefore makes it possible for Omiture to track users by their browser fingerprints, even if they block cookies or use private browsing mode (Eckersley, 2010).

Moreover, PayPal collects the data about the user during most of the payments. This option can be disabled by the merchant, but most of them are not aware of the process, or don't know how to get rid of that option. In effect, every time when the user feels safe, because it can be double checked whether all payment details, such as purchase details, payment details, address, etc., are correct, the user sees all data that is forwarded to PayPal.
Most of PayPal's daughter companies are payment related, but PayPal also started buying companies that are more connected to marketing.

The following it a list of a few of PayPal's daughter companies:
Venmo, Xoom, PayLater, Paidy, Braintree, Hyperwallet, PayPalZettle, Simility, Chargehound, Happy Returns, **Honey.**

1.1.1 PayPal & Honey

In January 2020, PayPal acquired Honey Science Corporation, which is known for "developing a browser extension that aggregates and automatically applies online coupons on eCommerce websites".

PayPal itself claims that this acquisition was made because "we believe our acquisition of Honey will enhance our value proposition by allowing us to further simplify and personalize shopping experiences for consumers".

Users can link their Honey and PayPal accounts to redeem Honey Gold rewards points for cash back sent to their PayPal balance, discover deals and make purchases. Integrating Honey's rewards program directly into PayPal should bring more Honey users to PayPal's app, growing its user base and setting up cross-promotional opportunities for PayPal's other shopping and financial services.

This acquisition of Honey and the use of models in decision processes has immediately raised some ethical concerns about the extension being a security risk that sells personal information, because personalizing the experience means processing users' data far beyond the basic safety models. According to PayPal's materials for shareholders and form 10-K[1], they make use of models when facing a decision of whether to grant a loan to a seller and for fraud detection. If those were the only types of models that they used, they would mostly need users' financial information and they wouldn't be able to really enhance users' experience with that kind of information. PayPal's acquisition of Honey provides a new way for users to redeem reward points and get discount codes for online shopping. On the other hand, it also gives PayPal a massive amount of new information on customers.

1.2 PayPal's commercial strategies

PayPal is highly represented in different social media services. Providing a lot of commercial input adapted to each specific platform, the biggest collection of PayPal's marketing strategies can be found on YouTube where even the TV commercials are available. Their YouTube channel (link in appendix) per se seems to be intended to help all users (private customers, businesses etc.) and to provide helpful information on how to use PayPal's functionalities and what their advantages are, mostly independent of information about their privacy regulations.

Commercial strategies in general can be divided into addressing private customers and businesses. For the private customers' internal needs like simplicity, velocity, diversity, and safety are triggered. PayPal expresses their main advantages through those needs like highlighting their "fast checkout", the many ways the PayPal checkout can be used in (e.g., Amazon, Deutsche Bahn, small web shops, etc.) and their extensive buyer protection. Other than that, the promotion of new implementations like the new installment payment option shows that PayPal always keeps track of the current development of their competitors (in this case Klarna). One of many slogans for private customers says: "Willste? Kriegste. Das ist der PayPal Unterschied." (engl.: "You want it? You get it. That's the PayPal difference."). Links for private customer commercials can be found in the appendix.

For business customers and companies, the facts of intense competitions and roots of decreasing turnover rates seem to be inherent. Describing social norms as prescriptive rules to get the attention and to underline the advantages of using PayPal based on numeric visualizations and developments is the usual way to go. Their suggestions of increasing confidence in online shops and a better customer experience shall help to enhance the company's economic standpoint. A slogan that is used to represent the main goal says: "Einkaufen. Leichtgemacht." (engl.: "Shopping. Made easy."). Links for private customer commercials can be found in the appendix.

[1] https://investor.pypl.com/financials/sec-filings/sec-filings-details/default.aspx?FilingId=14673604

1.3 PayPal & Application Programming Interface (API)

Corporations can rent computation time, storage space, and licensing fees by running applications over the internet. Companies then don't have to invest in expensive hardware, software licensing fees, and personnel to maintain it. So called Application Service Provider (ASP) models allow for a monthly subscription fee and flexible contracts. Such cloud computing resources are especially beneficial for small businesses, because they then don't have to set up a whole infrastructure. PayPal introduced the merchant Application Programming Interface (API)[2] with a basic credit card payment feature. Through this, any small corporation can start their business with near-zero investment and accept payments over the internet via PayPal. If the user clicks on the check-out button in any online store, a request is sent to PayPal. The user enters the credit card information and this information is sent back to the web shop. The user is then redirected to the check-out confirmation page.

APIs are increasingly used in everyday items when using apps on our smartphones. The ethical concern that is raised is that outsourcing data storage to a cloud operator leads to data privacy issues. The more APIs are used, the more vulnerable our privacy gets when a cloud is attacked. While the GDPR places a legal requirement on companies to protect users' data, it needs to be pointed out that even the best and safest systems and APIs will never be entirely protected from potential attacks.

Conclusion

PayPal claims that they don't share payment details with the seller.

"We store your personal data in secure databases. In addition, we automatically encrypt your confidential information when it's stored and transmitted. Your data can't be intercepted when it's sent from your computer to ours. Plus, when you use PayPal to make a secure purchase, we don't share your payment details with the seller."[3]

However, when looking at what information can be obtained through APIs, it becomes obvious that PayPal does share sensitive and personal information with the seller when requested. Moreover, the recruitment of smaller marketing and payment services/providers (such as Honey) opens up a gate for Paypal for acquiring much richer datasets and a significantly more well rounded estimation/prediction/analysis of users' profiles.

[2] What is the API (Application Programming Interface) actually? Rough explanation about the API: https://blog.maqe.com/what-is-the-api-application-programming-interface-actually-36bf1045fd31

[3] How does PayPal store my data and keep my data secure? https://www.paypal.com/va/smarthelp/article/how-does-paypal-store-my-data-and-keep-my-data-secure-faq4468

2 Paypal: User Interface

2.1 Terms & Conditions

This is a comparison of the number of words written in each Terms & Conditions document of different companies offering online payment. The PayPal User Agreement is a document that consists of far more words than those of other companies, and therefore takes sufficiently more time to read. A questionnaire by *Marktwächter*[4] showed that users want to spend maximally 5 minutes on reading Terms & Conditions:

Which Document?	**Number of words**	**Duration (*on average, adults are able to read 240 words/minute)**
PayPal User Agreement (English)	20,586	1’26’’
Klarna Terms for Service	5,850	24’’
Amazon Pay User Agreement	10,748	44’’
Paydirekt	7,047	29’’

In January 2018, PayPal Germany published a new version of their Terms & Conditions in German. This document consisted of more than 20,000 words and 1,000 sentences, the longest of them having more than 111 words. Experts claimed that PayPal violates the principle of transparency and users are disadvantaged. Based on a scientific text analysis, PayPal’s Terms & Conditions are formally incomprehensible and the time effort that users have to invest is unreasonable. Users don’t have a realistic chance to read and understand this document, and PayPal tries to gain a better position within the market because of that. With more recent versions, PayPal made their Terms & Conditions shorter, but they still have the longest document compared to others.

Additionally, the sentences in the terms and conditions and privacy statements are often complicated and are phrased as passive constructions. Words and phrases as ‘possibly’ and ‘among other things’ makes it even more vague and harder to understand.

2.2 Outer Appearance

2.2.1 Background information on deceptive design - Dark Patterns

Dark patterns tend to influence users’ decisions into decisions they probably wouldn’t make otherwise. Often this influence is not in the best interests of the users since they get nudged or tricked into e.g., spending more money than they intended to (Narayanan et al., 2020). Primarily online service providers try to deceive users to give access to their personal data voluntarily (Bösch et al. 2016). All this happens through exploiting cognitive biases. There are many ways to turn behavioral facts into effective user interfaces, e.g., uncovering behavior changes through nudge movements and data driven optimization (expose different designs to many users and analyze which one serves the best). Out of these strategies Narayanan et al. (2020) fear an upcoming trend in which companies offer dark patterns as a service.

Bösch et al (2016) highlights the example of user logins at starting pages that show an obvious login with personal accounts like Facebook, Google; a less bright and sticking out “create a separate

[4] https://www.vzbv.de/pressemitteilungen/marktwaechter-mahnen-paypal-ab

account" part as well as a hidden skip button (often in upper right corner). The intention seems to be accessing e.g. friends lists, photos and likes, and therefore getting information that is basically not necessary for the service, by using **intentionally bad design**.

Patterns and strategies

There are different kinds of strategies and patterns to describe how interface patterns can be implemented as well as different approaches to analyze those interface strategies.

Bösch et al. (2016) describe the evolution of dark patterns as follows:

In the beginning principles for "Privacy Design" from which different privacy patterns have arisen the main intention was to achieve and improve the users' privacies by guiding the developers providing structured problem-solution descriptions using standardized templates that can be easily looked up. Since a significant number of parties try to exploit users' privacy on purpose for criminal or financial reasons using these patterns in an opposite way, Bösch et al. (2016) reverse engineered the evolution of privacy strategies and patterns into dark privacy strategies and dark privacy patterns (summarized in Figure 1 in appendix). With the privacy dark patterns that evolved from this, Bösch et al.(2016) want to address non-technical end-users helping them educate themselves.

Hoepman (2014) furthermore maps his strategies onto legal principles that conform with (at that time) existing Privacy laws (see Figure 2 in appendix) and puts it into a database schema to visualize their interplay (see Figure 3 in appendix).

Psychological background

It is known that some degree of awareness about the existence of Privacy Dark Strategies is spread among the users. Nevertheless, dark strategies are still effective by their frequency of occurrence which can be derived from the psychological investigations on these strategies. Their main focus seems to lie on cognitive information processing and the psychological constitution (Bösch et al., 2016).

Bösch et al. (2016) describes two thinking processes:

1. **System 1 thinking process:** automatically, unconsciously and with little effort (e.g., agreeing quickly, intuitively and automatically to a list of terms and conditions that are unnecessarily complex and not understandable.)
2. **System 2 thinking process:** controlled, conscious and effortful; driven by deliberative, effortful decision-making -> slow execution behavior

Other than that Bösch et al (2016) highlight the "humans' fundamental need to belong" as highly influencing the effectiveness of Privacy Dark Patterns. This seems to be logical because the human being strives to be an accepted member of a certain group and to avoid the reduction of their own well-being. Taking this into account Bösch et al. (2016) claim that peoples' readiness to provide personal data increases significantly.

Recommendations for dealing with dark patterns

Privacy Dark Strategies and patterns are effective due to a lack of motivation and opportunity of people to resist. Following the information from the psychological backgrounds, system 2 thinking processes need to be trained and strengthened. This can be done by increasing motivation by repeatedly bringing the negative consequences of Privacy Dark Strategies to the foreground.

2.2.2 Nudging

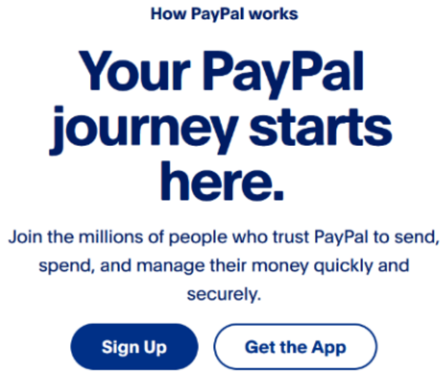
Nudging strategies use cognitive shortcuts observed in (irrational) decision making to make irrational behavior more explainable and even predictable. Since it employs descriptive social norms as prescriptive rules it can be used effectively, and is already adopted by instances like governments, employers, and businesses. (Narayanan et al., 2020)

An example suggested by Narayanan et al. (2020) on nudging the user into making quick decisions are sentences like this: *"7 people are looking at this right now!", "This deal expires in 15 minutes!"*. Those examples make users act with a sense of urgency by turning an abstract principle into a concrete nudge.

2.2.3 Dark Patterns and Nudging: PayPal

It is to say that PayPal doesn't use Dark Patterns and Nudging strategies as obviously and as perversely as many other services do. One can argue that this is due to their reputation, and they don't need to use these strategies as heavily. Some examples from the website in which some sort of dark patterns or nudging movements are used look like this (they provide different patterns depending on the language/country you select):

Private Customers:



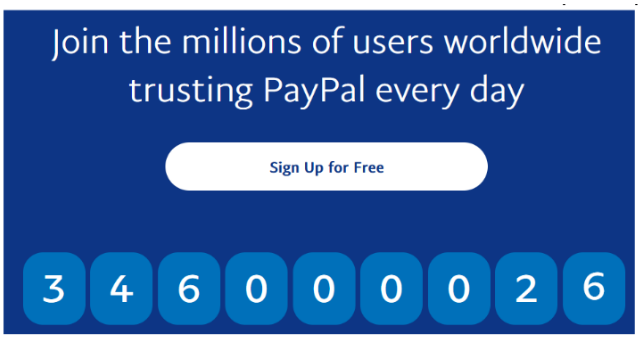
How PayPal works

Your PayPal journey starts here.

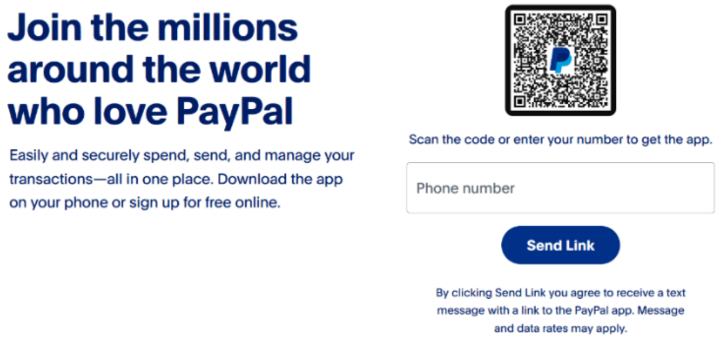
Join the millions of people who trust PayPal to send, spend, and manage their money quickly and securely.

An example of a dark pattern using different button colors to make the user sign up to PayPal.

PayPal addresses the users fundamental need to belong. The counter doesn't really count the actual number of users. It starts at 346.000.000 users every time one opens the website.



Join the millions of users worldwide trusting PayPal every day



Join the millions around the world who love PayPal

Easily and securely spend, send, and manage your transactions—all in one place. Download the app on your phone or sign up for free online.

Scan the code or enter your number to get the app.

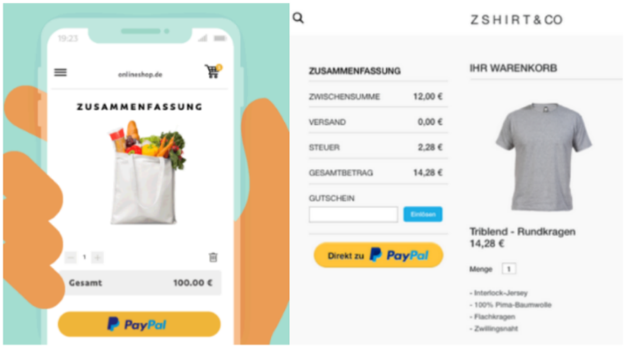
Phone number

Send Link

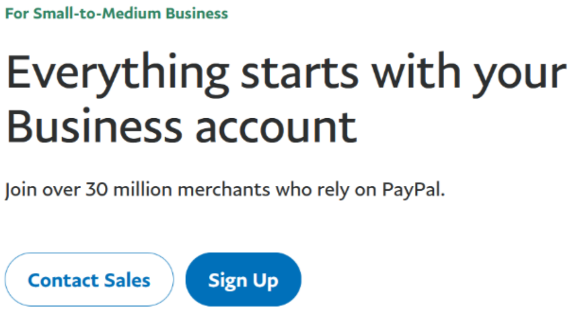
By clicking Send Link you agree to receive a text message with a link to the PayPal app. Message and data rates may apply.

PayPal forces the users to even get the App. Further they try to catch the user's attention and need to belong by telling them that many people already do love PayPal and highlight their advantages.

An example of the express checkout button using bright colors that stick out. It implies "PayPal security" and therefore increases the confidence in the shops.



Businesses:



For Small-to-Medium Business

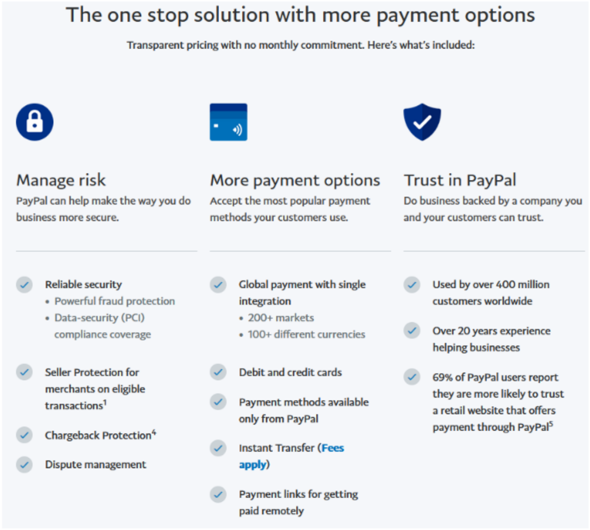
Everything starts with your Business account

Join over 30 million merchants who rely on PayPal.

Contact Sales   Sign Up

A dark pattern using different coloured buttons to make the business/online shop sign up to PayPal.

PayPal further provides many numeric examples and easy-to-use options that make one's business more profitable.



The one stop solution with more payment options

Transparent pricing with no monthly commitment. Here's what's included:

Manage risk

PayPal can help make the way you do business more secure.

- ✓ Reliable security
  - Powerful fraud protection
  - Data-security (PCI) compliance coverage
- ✓ Seller Protection for merchants on eligible transactions[1]
- ✓ Chargeback Protection[4]
- ✓ Dispute management

More payment options

Accept the most popular payment methods your customers use.

- ✓ Global payment with single integration
  - 200+ markets
  - 100+ different currencies
- ✓ Debit and credit cards
- ✓ Payment methods available only from PayPal
- ✓ Instant Transfer (Fees apply)
- ✓ Payment links for getting paid remotely

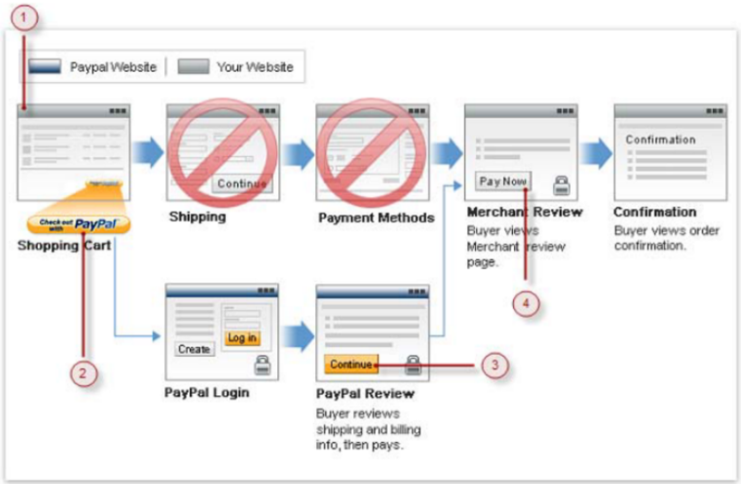
Trust in PayPal

Do business backed by a company you and your customers can trust.

- ✓ Used by over 400 million customers worldwide
- ✓ Over 20 years experience helping businesses
- ✓ 69% of PayPal users report they are more likely to trust a retail website that offers payment through PayPal[5]

Next to other options like developer tools (API's etc.) they also provide detailed information (PayPal, Inc. 2013) in which they help businesses to achieve the PayPal "advantages" by telling the customer that PayPal is available at their service. The guide explains in detail how to place the express checkout button precisely to keep the users' attention and to make the user buy items from their shop. They want to „blend digital and physical commerce to create seamless customer experiences".

FIGURE 1.1 Merchant's Checkout Flow



1. Add the PayPal acceptance graphic to the page.
2. Add the **Checkout with PayPal** button. Call the SetExpressCheckout API request. Redirect the buyer to PayPal.
3. PayPal redirects the buyer to the Merchant site. Call the GetExpressCheckoutDetails API request.
4. Call the DoExpressCheckout Payment API request.

3 Big Data and Privacy

3.1 Background

In today's world, personal data can be traded as a currency in Big Data. There are different opinions to what extent this situation is ethical. In the trading model of personal data, the transmission of personal data is a framework that offers people the opportunity to control their digital identity and create granular agreements of data sharing. The idea of open data centers around the argument that data should be freely available. However, willingness to share data varies by person.

The sum of an individual's personal data forms a **digital identity**. A digital identity has the advantage of quick access to online content and related services. Its use might generate discrimination based on the representation of a person according to their online data, which may often not correspond to the real situation. We seem to be caught in a process called "data dictatorship" in which "we are no longer judged on the basis of our actions, but on the basis of what all the data about us indicates our probable actions may be".[5]

A fundamental question in the ethics of Big Data research is, who owns the data? This involves the subject of property rights and obligations. In European law, the General Data Protection Regulation (GDPR)[6] indicates that people have ownership of their own personal data. The GDPR has changed the data landscape in the European Union (EU) since its implementation in May 2018. *According* to a survey of UK consumers by the Data and Marketing Association (DMA), 62% of consumers indicate that the GDPR will improve their confidence in sharing data with companies. The protection of personal data of an individual is based on the opinion that they provide a direct expression of one’s identity.

The relationship between those who provide the data and those who use it is often indirect. There is a big gap in the process between accepting terms or cookies that give away amounts and combinations of data that we cannot imagine, and the actual use of this data by companies. Third parties might be involved at various steps of the chain, or might be at the other end of the chain, waiting to receive data that we did not directly share with them. This data is crucial to the identity of a person and should really be handled carefully, which is usually not the case.

[5] Norwegian Data Protection Authority, 2013, from N. Sfetcu

[6] https://gdpr.eu

3.2 Paypal’s Privacy Statement *(6 May, 2022)*

In their privacy statement,[7] PayPal lists which personal data they collect from their customers:

- Name, address, phone number, e-mail, data for identification
- Amount sent or requested, amount paid for products or services, merchant information, including information about any funding instruments used to complete the transaction, device information, technical usage data, and geolocation information
- Personal data of every participant of a transaction
- If the user establishes a connection between their device or a social media platform and a user’s PayPal account, PayPal uses the users’ contact list information (such as name, address, email address) to improve the experience

They further explain why they collect personal data and for how long they store it:

“We retain personal data in an identifiable format for the ***least amount of time necessary to fulfill our legal or regulatory obligations*** *and for our business purposes. We may retain Personal Data for* ***longer periods than required by law if it is in our legitimate business interests and not prohibited by law****. If your account is closed, we may take steps to mask Personal Data and other information, but we* ***reserve our ability to retain and access the data for so long as required*** *to comply with applicable laws. [...] We may also* ***process your personal data*** *where we believe it is in our or others’ legitimate interests, taking into consideration* ***your interests, rights, and expectations****.”*

PayPal’s Privacy statement, and also their Terms & Conditions, are contracts where users have no chance to make a change and can only reject or accept. Users might therefore be less motivated to read those documents. Because users are the product and companies make profit by collecting their data, the companies are not interested in making an effort to make the documents more comprehendible. It’s their goal to make the user accept them quickly and use their service. This also shows that PayPal, opposite to what they claim in their privacy statement, doesn’t ask for and considers users’ interests and expectations.

In their privacy statement, PayPal further explains that they use their users’ personal data to operate the site, manage their business needs (which include analyzing, monitoring and improving their site), manage risks, and provide personalized services, also called ‘interest-based marketing’. With the user’s consent, PayPal makes use of cookies[8] and tracking technologies to customize the user’s experience and services, measure the effectiveness of promotions and prevent potential fraud.

In interest-based marketing, people are presented with ads that fit their interests based on previous searches and behavioral data analysis. When this is applied during political campaigning, people are presented with increasingly extreme opinions about political topics to influence opinions. This might eventually lead to a polarization which is not based on neutral facts, but which is due to an algorithm that presents the individual with topics and ads that seemingly fit their interests and where people end up clicking on.

In their privacy statement, PayPal further states that they share aggregated statistical data with third-parties about how, when, and why users visit their site and make use of their services, while promising that this data will not identify the individual user. They state that PayPal doesn’t share their users’ personal data with third parties for their marketing purposes without the user’s consent.

[7] This is **Paypal's Privacy Statement** where they explain which kind of data they take from customers. The language used is abstract in some parts and allows for a more open interpretation of what exactly they are collecting. https://www.paypal.com/us/webapps/mpp/ua/privacy-full#privacyRights

[8] This is an explanation of how PayPal uses cookies: https://www.paypal.com/US/webapps/mpp/ua/cookie-full

Lastly, they state that PayPal shares their users' personal data with other third parties for their business purposes or as permitted or required by law. PayPal determines if it is necessary to share a user's personal information with third-parties to prevent physical harm or financial loss, or when a user is connected with an investigation of suspected or actual illegal activity. PayPal describes nowhere what their business interests are and what personal data is necessary to fulfill this interest. They are intransparent about their business interests, creating a disadvantage for users, because it is not clear what data is collected and what this data is used for. Based on this part of their privacy statement, PayPal could collect any information they want and always argue that this is necessary to fulfill their business interest, which is also not obvious what that is. Consumer advisors criticize PayPal for collecting more data than necessary.[9] During registration and payment, between 4 and 13 individual data entries get collected. Up to 11 trackers are used to collect information about the consumer's behavior for targeted ads. The GDPR fails to protect users in such situations.

They finish their privacy statement by stating that PayPal may use automated decision-making for decisions concerning credit with the user's consent.

3.3 Collecting and Sharing of Personal Information

Based on the information they provide in the AppStore, PayPal collects the following data for advertising and marketing, analysis, personalization of the product and app functionality:

- Shopping history
- Contact data (e-mail address + name + surname)
- Search history
- Browsing history
- Identification
- Usage data (product interactions, other data)
- Financial information (information about payments, credit information, other)
- Localization (exact location, estimated location)

PayPal further states that they share their users' personal data with other members of the PayPal corporate family, which include Honey, Braintree, Paidly and Venmo, to provide the service and deal with possible fraud.

PayPal also shares their users' personal data with other companies that provide a service to PayPal to verify their users' identity, send advertisements about PayPal's services, or provide customer service. The users' data is also shared with other financial institutions to give benefits to the users that are associated with their cards and keep users' financial information up to date. This poses an ethical issue, because it implies a continuous "surveillance" of a user's financial status from many different sources, at different points in time and from a combination of sources.

Consumption patterns that are linked to a payment method are not pseudonymous but identifiable through offline details such as credit card numbers or bank account details. Therefore, if a company is interested in identifying the individual behind pseudonymous data, they are able to do so by gathering other information that can serve as a link between the pseudonymous data and the individual and even allow the linkage of multiple transactions across different logins or accounts.

[9] Kritik an Datenhunger von PayPal & Co: https://help.orf.at/v3/stories/2878112/

Consumer details are collected and merged across transactions, even for sensitive products and merchants.

PayPal’s data science team can help target people in a better way through so-called “Transactional Data”. Transactional data is information that is captured from transactions like time, place, amount of money, payment method etc. This kind of data is the most solid factor that helps data scientists predict people’s buying behavior patterns. **Transactional data remains the strongest pointer in predicting customer behavior at PayPal,** even stronger than data about how many people looked at a product, which website they visited, etc. but.

3.4 PayPal and the General Data Protection Regulation (GDPR)

In Article 5, the GDPR determines which personal data may be collected:

GDPR Article 5

„1. Personal data shall be:

Adequate, relevant and limited to what is necessary in relation to the purpose for which they are processed (data minimization).“

The GDPR states that the principle of ‘data minimization’[10] should be applied, which means collecting as little information as possible for a purpose. The problem is that the EU law doesn’t define what ‘adequate, relevant and limited’ data is. What data is needed and collected by the processor depends on the purpose, but a company should never aggregate more data than necessary for the purpose they want to fulfill. Data collection is insufficient if the purpose of the processing is not possible, and also, if a decision cannot be made due to a lack of understanding of the facts. When the data is used to make decisions, it is in the users’ interest to provide as much data as needed to make a fair decision. However, the downside is that the users never know if the collection of this particular information is truly necessary and what else could be done with that data.

As shown above, article 5 states that „personal data shall be adequate, relevant and limited to what is necessary in relation to the purpose for which they are processed (data minimization).“ **Because the GDPR does not further describe what „adequate, relevant and necessary data” is, PayPal may collect any data as long as they can justify it with their business interests, which are not obvious to the users.**

[10] What is “data minimization” under EU Data Protection Law?
https://medium.com/golden-data/what-is-data-minimization-under-eu-data-protection-law-b0e30fbb856e

3.5 Privacy and Communication

PayPal itself claims to be very interested in user privacy, but it doesn't clearly talk about how to protect privacy. After reading the Form-10k and materials for shareholders[11] it seems that they rather mean data security. It is very important for their business that the data they obtain, especially financial data like credit card information, doesn't get outside. However, they don't consider creating user profiles as something that goes against the users' data privacy. They gather as much of their users' data as possible and use it to "enhance their experience" and to lower the risk for themselves.

PayPal mentions that their *Global Privacy Program* is based on the following 8 principles:

- Management
- Notice & Transparency
- Choice & Consent
- Security
- Data Lifecycle Management
- Data Quality
- Stewardship
- Standardization

They claim to integrate "Data Hygiene by Default" and "Privacy by Design" culture throughout their company, and they do to the chosen extent. They also organize annual mandatory compliance, ethics training and education for all employees and contractors. It is impossible to say that they don't keep their promises about data protection. But this is true only because they choose to define these promises in the way that suits them and can even be used as a part of marketing their payment system as a more safe and secure option.

3.6 Connection to Theoretical Concepts

Based on the sections above, it becomes clear that PayPal collects data that goes far beyond the information that is necessary to process a payment. The more data gets collected, the greater the success for an artificial intelligence (AI) and machine learning (ML) algorithms to find hidden patterns in that data to predict our behavior and desired outcomes. Furthermore, a huge amount of data makes it easier to apply *Predictive Analytics* to get predictions either on the future development of the economy and market activity in general or on a person's individual future buying behavior. Besides, this can lead up to a prediction on people that are not using services like PayPal at all. With AIs becoming more powerful and an increasing availability of big data, *proxy discrimination* will pose a greater challenge to anti-discrimination regimes (Prince & Schwarcz, 2020).

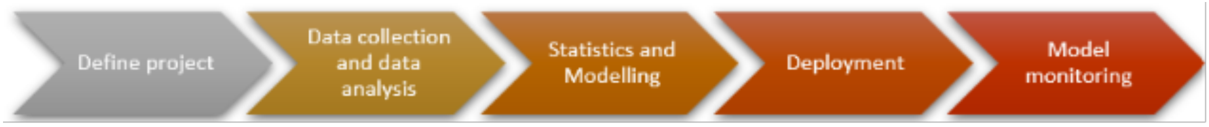
3.6.1 Predictive Analytics

Predictive Analytics describes a sort of pattern matching (Mühlhoff, 2022). Huge datasets are being analyzed automatically, helping organizations and companies to have a better understanding of how people behave in a certain situation (Mishra & Silakrai, 2012). Because curiosity and trying to predict future events are often defined as part of human nature, the prediction of future events based on previously observed and analyzed data through ML algorithms and determining the likelihood of the occurrence of a certain situation have evolved increasingly fast (Mishra & Salikrai, 2012; Kumar et al., 2016). Based on that, Mishra & Salikrai (2012) define the core of behavioral predictions as "capturing

[11] https://investor.pypl.com/financials/sec-filings/sec-filings-details/default.aspx?FilingId=14673604

relationships between explanatory variables and the predicted variables from past occurrences, and exploiting it to predict future outcomes."

Simplified, a prediction process looks like this (Ongsulee et al., 2018):



In the phase of *defining the project,* the outcomes and project directives are defined. That includes the companies' interests (also known as predictor) as well as their preferred and possible outcomes (Ongsulee et al., 2018). A predictor in this case paraphrases the variable to determine. As an example, credit card companies could be interested in the risk factor (= objective) that can be determined by defining predictors that should be considered: age, income, credit history, and more (Mishra & Silakrai, 2012).

Data collection and data analysis includes everything that happens to the data, starting with the source and the way of collecting data up to analyzing it with respect to the predefined objectives (Ongsulee et al., 2018). For the credit card company that would mean that the data analysis needs to consider data collections that include age, income, and credit history of a user (Mishra & Silakrai, 2012).

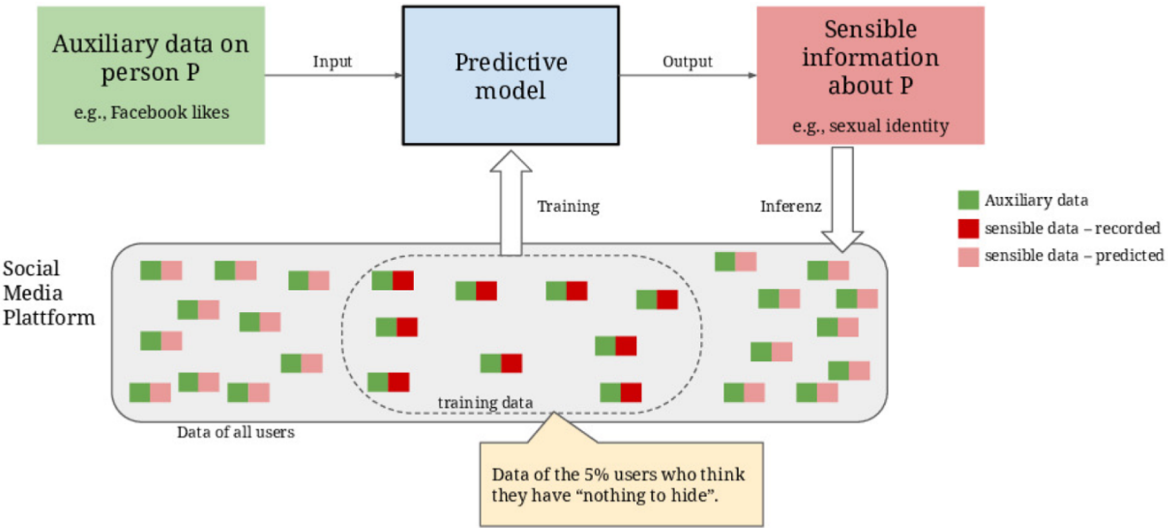
Conducting *statistics and modeling* enables the validation and testing of possible outcomes and then creating precise predictive models about future situations (Ongsulee et al., 2018). In the case of the credit card company, patterns will be found depending on the predictors to feed and train the model with (Mishra & Silakrai, 2012).

In the phase of *deployment,* the results can be transferred to everyday decision making, which also generates further outputs that can be reused to train the model to become even more precise. These results are again managed and monitored in the *model monitoring* phase to increase the model's performance (Ongsulee et al., 2018).

Next to credit card companies, Kumar et al. (2016) define more general applications for predictive analytics:

- Direct marketing and sales: Usage and customer behavior on the company's website can be analyzed to determine the likelihood of sale and create targeted advertising
- Customer relationships
- Pricing optimizations: Modulation of demand and price to achieve an efficient pricing strategy
- Health outcomes: Connecting symptoms with treatments and therefore increase reaction times making diagnoses of irreparable diseases (e.g., heart attacks)
- Insurance fraud: Identification of patterns using statistical models for prevention and investigations
- Improper public benefits payments and fraud: Less waste of taxes and ensure benefits to those who really need it
- Tax collections: Identification of owed taxes
- Predicting and preventing street crime, domestic abuse, and terrorism: Determine high-risk situations and places

Besides the positive effects mentioned by Kumar et al. (2016), predictive analysis brings some challenges that need to be considered. The main challenges mentioned by Mishra & Silakrai (2012) are privacy and ownership, as well as data ecosystems and exchanges. This means there is a high potential for data being misused in a way that the law doesn't catch any of it, so that it easily leads to strong social impacts like discrimination and inequality (Mühlhoff, 2022). Furthermore, it highly violates the sensible private space of all people by using easily accessible data for predictions (see figure; Mühlhoff, 2022).



It is hard to claim that PayPal is actively taking part in data collection, but by the way they navigate the market and approach data collection, they definitely seem to be acquiring large amounts of information on their users. By employing different daughter companies, they manage to create more well-rounded user profiles with a huge variety of information, from Google searches, over mobile phone metadata, to transactional data, credit scoring information and much more. This makes it possible to use these profiles in ways that we as users can hardly imagine, whether they are directly used by PayPal or sold to data brokers or other companies. This is the alarming part and should make us question why PayPal needs all this information on us and how they use it, possibly without informing us.

3.6.2 Proxy Discrimination

There are laws that prohibit the use of information about membership of protected classes (Prince & Schwarcz, 2020). 'Every person is protected from discrimination on grounds of race, ethnic origin, gender, disability, religion, belief or philosophy of life, age, or sexual orientation.'[12] The goal is to protect people from socially harmful and unfair outcomes that are based on group membership (Prince & Schwarcz, 2020). However, if one wants to make a prediction about behavior of a member of a protected class, a non-protected variable that highly correlates with the protected class, a so-called 'proxy', can be used to make such a prediction (Prince & Schwarcz, 2020). The next best information is taken to obtain a solid prediction. This is called 'intentional proxy discrimination', as the discriminator knows about the correlation between the protected class and another variable (Prince & Schwarcz, 2020).

Proxy discrimination can also happen unintentionally, simply because it helps to achieve a goal. Discrimination can then be an unintended by-product which is due to a correlation between a protected and an unprotected variable (Behrendt & Loh, 2022). Programmers and engineers are often only focused on the success and predictive power of an AI, not taking into consideration the 'black box', meaning the rules that the AI applies in order to reach a prediction (Prince & Schwarcz, 2020). Relying on proxies leads to using group characteristics that again amounts to stereotyping, because it associates

[12] https://www.antidiskriminierungsstelle.de/EN/homepage/homepage-node.html

individuals with the general characteristic of a protected class (Prince & Schwarcz, 2020). Already disadvantaged and vulnerable groups can become the target of increased exploitation by offering them loans with increased interest rates. This leads to a perilous feedback loop of worsening the economic conditions of those groups and eventually widening the economic and social gap between the rich and the poor (Favaretto, De Clercq, & Elger, 2019).

Our society produces a large amount of data every day, and this data still holds the historical bias against protected classes, such as race and gender, and is further used to train algorithms to make predictions (D'Alessandro, O'Neil, & Lagatta, 2017). The legacies of historical discrimination are strengthened because they are beneficial for accurate predictions that lead to increased profit of corporations (Prince & Schwarcz, 2020). This leads to over- or underrepresentation in the training data that is used to train an algorithm, leading to a maintenance of discrimination (Barocas & Selbst, 2018; D'Alessandro et al., 2017). Reversing this discrimination can be costly for the ones benefiting the most from it (Prince & Schwarcz, 2020). This shows, that proxy discrimination by AIs is normatively harmful, because it discriminates against people of protected classes that anti-discrimination laws try to protect (Prince & Schwarcz, 2020).

Simply denying access to proxy information is not a solution, as the algorithm will again make use of the next best information that serves as a proxy for class discrimination to obtain a prediction (Gillis & Spiess, 2019; Prince & Schwarcz, 2020). To account for biases in the training data and algorithms, sensitive data has to be collected and used in the modeling process to correct for the bias and prevent discrimination (Žliobaitė & Custers, 2016). In order to prevent discrimination, we have to pinpoint and name it (Gillis & Spiess, 2019).

3.7 Example - Credit Scoring

Credit scoring is an example that shows how aggregation of data at large scale can lead to unfair and discriminating outcomes because of proxy discrimination and predictive analytics. Credit scoring can impact people's lives without any direct and obvious connections to online footprints.[13] Oftentimes, credit scoring happens as a result of analyzing large amounts of data from one person in order to allow that individual to take out a loan, open a bank account, pay in rates etc. However, people are not necessarily aware of the background computations involving them and might never even find out.

AI-based credit scoring

Consumers' privacy and protection are at the core of efficient data collection.[14] The personal information of consumers is at high risk of being misappropriated. AI-based credit scoring utilizing smartphone metadata offers a mutually beneficial solution to both consumers and institutions.

AI is used to convert metadata into credit scores. Metadata refers to data about other data, the non-personal, binary (1s and 0s) version of the same data. AI-based credit scoring assesses smartphone metadata to detect predictive patterns and reliable credit scores can be generated from such alternative data. When models are used to calculate someone's credit score, they make use of proxies. People are placed in categories of other people that have behaved like them in the past. The

[13] Credit Scoring: Keeping Customer Privacy at the Forefront, Michele Tucci
https://www.credolab.com/news-press/credit-scoring-keeping-customer-privacy-at-the-forefront

[14] How safe is customer data?
https://www.credolab.com/news-press/credit-scoring-keeping-customer-privacy-at-the-forefront

problem here is, that the question that the model answers is not about how *I* have behaved in the past, but rather how *others like me* have behaved. This leads to stereotyping, disregarding the individual, and eventually to discrimination. If an individual is put into a category with other people that have behaved similarly to me in the past but have eventually not paid back a loan, the person will have to pay higher interest rates. The model doesn't check whether the individual is creditworthy, and the individual has worse chances and conditions simply because s/he behaves similarly to credit-unworthy people. Even though using proxies in automated decision-making is unethical and leads to discrimination, they do help in making accurate predictions, which again leads to higher revenues and security for companies. The problem is however, that there is no chance to feedback the system whether a decision was right or not. Such credit scoring models are black boxes, it is nearly impossible to find out why a system made the decision and which variables were used. People that are granted no loan or one with higher interest rates have no way to find out why and no chance to raise a complaint. Therefore, the vicious cycle of discrimination, unequal chances and the gap between poor and rich exacerbates (O'Neill, 2017).

The use of smartphone metadata also allows for real-time analysis. A credit score is generated within seconds of accessing the phone's metadata. As a result, AI and machine learning can enable faster processing time and, in some cases, an immediate (dis)approval. Credit scores are sometimes used during hiring processes. People with worse credit scores have less chances to get a job, which again exacerbates the poverty cycle. A good credit score is not just a proxy for responsibility and smart decisions, but also with wealth, which is again highly correlated with race (O'Neill, 2017).

"Positive & Negative" uses of data for Credit Scoring

The availability of big data sources is creating new opportunities as well as challenges for credit scoring.[15] These sources may be useful to score customers who lack borrowing experience (because it's their first loan or they recently moved to a new country) and would be automatically perceived as risky according to traditional credit scoring models which rely on historical information. Using these data sources also presents challenges. The first one concerns privacy. It is important that customers are properly informed about what data is used to calculate their credit score. An opt-out option should always be provided.

Furthermore, using social network data for credit scoring can trigger new fraud behavior whereby customers strategically construct their social network to artificially and maliciously improve their credit quality. Finally, regulatory compliance might also become an important issue. Many countries prohibit the use of gender, age, marital status, national origin, ethnicity, and beliefs for credit scoring. Since much of this information can easily be scraped from social networks, it may be harder to oversee regulatory compliance when using social networks or other data for credit scoring.

[15] Bart Baesens BDQ (Big Data Quarterly): Big Data for Credit Scoring: Opportunities and Challenges
https://www.dbta.com/BigDataQuarterly/Articles/Big-Data-for-Credit-Scoring-Opportunities-and-Challenges-109999.aspx

4 Collective Responsibility - The Bigger Picture

What does it mean to be responsible (Björnsson, 2019)? It is highly dependent on the context. It can mean:

1) A quality of moral character
2) Responsibility to serve in a particular character – ex ante
3) Being an agent that the action or its consequences can be attributed to – ex post

In big companies like PaypPal it is pretty easy to juggle the blame when something goes wrong, which in turn results in an approach to making decisions in big companies that can be catastrophical. Without the proper legislation there is nothing keeping the companies in line ethically. Customers need to be notified about all of the potential risks of their movements, what they are agreeing to etc.

- Some people will still decide to save some money and risk using less safe systems (even if they know all of the risks)
- Some people when given information and choice will not choose to risk it and instead will be even willing to lose money (or data currency)

Everyone in these companies is a part of the final product. Employees at big corporations such as PayPal are parts of a collective. They as a collective have the following important components of responsibility:

- Control (over the outcome)
- Knowledge (awareness of the consequences)
- Answerability (ex-post being available for questions, praise and blame, and ex-ante being answerable to the future)

Moreover, in cases involving Big Data, users are also entangled in this “responsibility vortex” in the sense that one user might expose other users or even non-users, through friend lists, contact lists etc. through social media, or, last but not least, through feeding predictive algorithms with data which can subsequently be used to predict behavioral patterns of larger populations of people. The danger of exposing data belonging to third parties - through blindly accepting terms and conditions or cookies - and subsequently giving access to, for example, our friends’ data on facebook, can lead to scary situations as we have already seen in the past (Cambridge Analytica). Therefore, responsibility is an important topic for both users and companies.

5 References

Barocas, S., & Selbst, A. D. (2018). Big Data's Disparate Impact. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.2477899

Behrendt, H., & Loh, W. (2022). Informed consent and algorithmic discrimination–is giving away your data the new vulnerable? *Review of Social Economy*, *80*(1). https://doi.org/10.1080/00346764.2022.2027506

Björnsson, Gunnar. (2019). Collective responsibility and collective obligations without collective moral agents.

Bösch, C., Erb, B., Kargl, F., Kopp, H., & Pfattheicher, S. (2016). Tales from the Dark Side: Privacy Dark Strategies and Privacy Dark Patterns. Proc. Priv. Enhancing Technol., 2016(4), 237-254.

D'Alessandro, B., O'Neil, C., & Lagatta, T. (2017). Conscientious Classification: A Data Scientist's Guide to Discrimination-Aware Classification. *Big Data*, *5*(2). https://doi.org/10.1089/big.2016.0048

Eckersley, P. (2010). How unique is your web browser? In Proceedings of the 10th international conference on Privacy enhancing technologies (PETS'10). *Springer-Verlag, Berlin, Heidelberg*, 1–18.

Favaretto, M., De Clercq, E., & Elger, B. S. (2019). Big Data and discrimination: perils, promises and solutions. A systematic review. *Journal of Big Data*, *6*(1). https://doi.org/10.1186/s40537-019-0177-4

Gillis, T. B., & Spiess, J. L. (2019). Big Data and Discrimination. *The University of Chicago Law Review*, *86*(2), 459–488.

Hoepman, J. H. (2014). Privacy design strategies. In IFIP International Information Security Conference (pp. 446-459). *Springer, Berlin, Heidelberg.*

Kroll, J. A., Huey, J., Barocas, S., Felten, E. W., Reidenberg, J. R., Robinson, D. G., & Yu, H. (2017). Accountable algorithms. *University of Pennsylvania Law Review*.

Kumar, S., Sharma, P., & Madhusudan (2016). Predictive Analytics: Cloud Computing

Mishra, N., & Silakrai, S. (2012). Predictive Analytics: A Survey, Trends, Applications, Oppurtunities & Challenges

Mühlhoff, R. (2022). Predictive Analytics: Prädiktive Privatheit: Kollektiver Datenschutz im Kontext von Big Data und KI

Narayanan, A., Mathur, A., Chetty, M., Kshirsagar, M. 2020. Dark Patterns: Past, Present, and Future: The evolution of tricky user interfaces. *Queue 18, 2, Pages 10 (March-April 2020)*, 26 pages. https://doi.org/10.1145/3400899.3400901

O'Neil, C. (2017). *Weapons of math destruction*. Penguin Books.

Ongsulee, P., Chotchaung, V., Bamrungsi, E. & Rodcheewit, T. (2018). "Big Data, Predictive Analytics and Machine Learning," *16th International Conference on ICT and Knowledge Engineering*

Preibusch, S., Peetz, T., Acar, G., & Berendt, B. (2016). Shopping for privacy: Purchase details leaked to PayPal. *Electronic Commerce Research and Applications*, *15*, 52–64. https://doi.org/10.1016/j.elerap.2015.11.004

Prince, A. E. R., & Schwarcz, D. (2020). Proxy discrimination in the age of artificial intelligence and big data. *Iowa Law Review*.

Sfetcu, N. Big Data Ethics. *Www.academia.edu*. Retrieved September 1st, 2022, from https://www.academia.edu/42114775/Big_Data_Ethics

Sfetcu, N. Philosophical Essays. In *www.scribd.com*. Retrieved September 18th, 2022, from https://www.scribd.com/book/567537100/Philosophical-Essays

Squires, G. D. (2003). Racial profiling, insurance style: Insurance redlining and the uneven development of metropolitan areas. *Journal of Urban Affairs*, *25*(4). https://doi.org/10.1111/1467-9906.t01-1-00168

Žliobaitė, I., & Custers, B. (2016). Using sensitive personal data may be necessary for avoiding discrimination in data-driven decision models. *Artificial Intelligence and Law*, *24*(2). https://doi.org/10.1007/s10506-016-9182-5

6 Appendix

This is a list of useful links and websites that we visited frequently to get general information about a variety of related topics.

- **What is "data minimization" under EU Data Protection Law?**
  https://medium.com/golden-data/what-is-data-minimization-under-eu-data-protection-law-b0e30fbb856e)

- **Visualizing the Length of the Fine Print, for 14 Popular Apps**
  https://www.visualcapitalist.com/terms-of-service-visualizing-the-length-of-internet-agreements/

- **PayPal-Website**
  https://www.paypal.com/de/home

- **PayPal YouTube Channel:** Their YouTube channel is intended to help all users (private customers, businesses...) where they again mention all their advantages and functionalities independent of any privacy issues/information.
  https://www.youtube.com/user/paypalde

- **Businesses Commercial:** The commercial is addressing businesses/companies and the intense competitions.
  https://www.youtube.com/watch?v=XqNxLf2XiGg

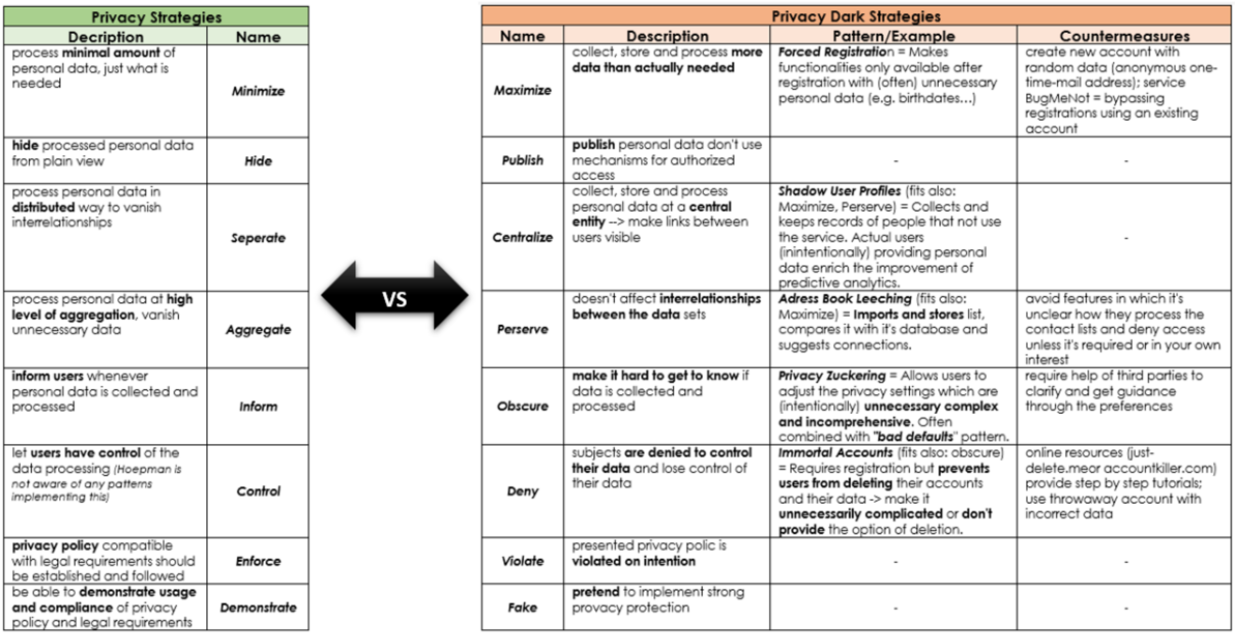
- **Private Customers Commercial:** The commercials are addressing private customers and depict all functionalities, advantages and new implementations of PayPal.
  https://www.youtube.com/watch?v=HFWkRfnb7OA
  https://www.youtube.com/watch?v=0nnRRe2rPnI

- Resistance to dark patterns should be increased by getting to know more about the **functionality of Privacy Dark Patterns** (Bösch et al. 2016). For that, websites like https://www.deceptive.design/ are highly recommended.

- **PayPal, Inc. 2013. Express Checkout User Interface Standards:**
  https://www.paypalobjects.com/webstatic/en_US/developer/docs/pdf/pp_ecplacement_guide.pdf

- **Info on how PayPal manage the large amounts of data that they collect from users:**
  https://www.projectpro.io/article/big-data-use-cases-how-paypal-leverages-big-data-analytics/231

- Figure 1 summarizes the **privacy design strategies** defined by Hoepman (2014) and then contrasts them with the **dark strategies** derived.

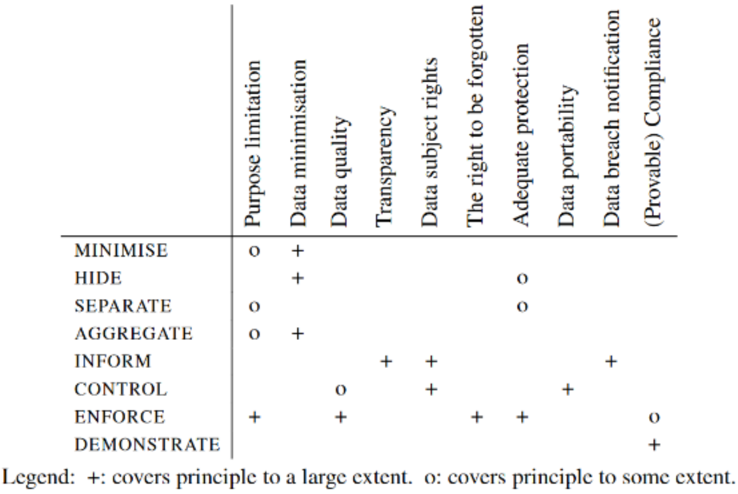


Privacy Strategies	
Decription	Name
process **minimal amount** of personal data, just what is needed	***Minimize***
hide processed personal data from plain view	***Hide***
process personal data in **distributed** way to vanish interrelationships	***Seperate***
process personal data at **high level of aggregation**, vanish unnecessary data	***Aggregate***
inform users whenever personal data is collected and processed	***Inform***
let **users have control** of the data processing *(Hoepman is not aware of any patterns implementing this)*	**Control**
privacy policy compatible with legal requirements should be established and followed	***Enforce***
be able to **demonstrate usage and compliance** of privacy policy and legal requirements	***Demonstrate***

Privacy Dark Strategies			
Name	Description	Pattern/Example	Countermeasures
Maximize	collect, store and process **more data than actually needed**	***Forced Registratio***n = Makes functionalities only available after registration with (often) unnecessary personal data (e.g. birthdates...)	create new account with random data (anonymous one-time-mail address); service BugMeNot = bypassing registrations using an existing account
Publish	**publish** personal data don't use mechanisms for authorized access	-	-
Centralize	collect, store and process personal data at a **central entity** --> make links between users visible	***Shadow User Profiles*** (fits also: Maximize, Perserve) = Collects and keeps records of people that not use the service. Actual users (inintentionally) providing personal data enrich the improvement of predictive analytics.	-
Perserve	doesn't affect **interrelationships between the data** sets	***Adress Book Leeching*** (fits also: Maximize) = **Imports and stores** list, compares it with it's database and suggests connections.	avoid features in which it's unclear how they process the contact lists and deny access unless it's required or in your own interest
Obscure	**make it hard to get to know** if data is collected and processed	***Privacy Zuckering*** = Allows users to adjust the privacy settings which are (intentionally) **unnecessary complex and incomprehensive.** Often combined with "***bad defaults***" pattern.	require help of third parties to clarify and get guidance through the preferences
Deny	subjects **are denied to control their data** and lose control of their data	***Immortal Accounts*** (fits also: obscure) = Requires registration but **prevents users from deleting** their accounts and their data -> make it **unnecessarily complicated** or **don't provide** the option of deletion.	online resources (just-delete.meor accountkiller.com) provide step by step tutorials; use throwaway account with incorrect data
Violate	presented privacy polic is **violated on intention**	-	-
Fake	**pretend** to implement strong provacy protection	-	-

Figure 1: Comparison of the Privacy Strategies by Hoepman (2014) and the reverse engineered Dark Privacy Strategies by Bösch et al. (2016). (Table inspired by existing table of Bösch et al (2016))

- In Figure 2 Hoepman (2014) maps his strategies onto legal principles that conform with (at that time) existing Privacy laws.



	Purpose limitation	Data minimisation	Data quality	Transparency	Data subject rights	The right to be forgotten	Adequate protection	Data portability	Data breach notification	(Provable) Compliance
MINIMISE	o	+								
HIDE		+					o			
SEPARATE	o						o			
AGGREGATE	o	+								
INFORM				+	+				+	
CONTROL			o		+			+		
ENFORCE	+		+			+	+			o
DEMONSTRATE										+

Legend: +: covers principle to a large extent. o: covers principle to some extent.

Figure 2: The eight Privacy Strategies mapped onto legal existing principles by Hoepman (2014), p. 457.

- In Figure 3 Hoepman (2014) puts his Privacy Strategies into a database schema to visualize their interplay.

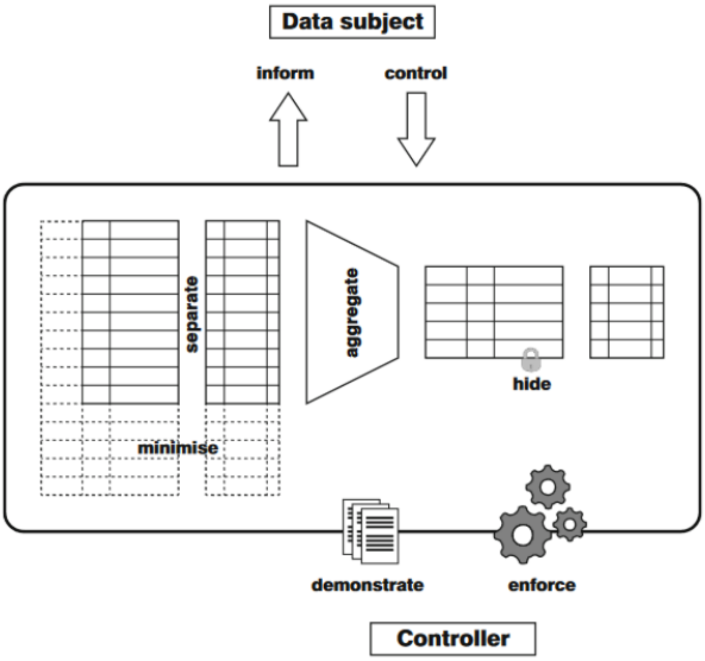


Figure 3: Schema of the eight Privacy Strategies by Hoepman (2014), p. 457.