

АНДРЕАС ВАЙГЕНД

# BIG DATA

ВСЯ ТЕХНОЛОГИЯ В ОДНОЙ КНИГЕ



Top Business Awards

Андреас Вайгенд

**BIG DATA. Вся  
технология в одной книге**

«ЭКСМО»

2017

УДК 004.6  
ББК 32.973-018.2

**Вайгенд А.**

BIG DATA. Вся технология в одной книге / А. Вайгенд —  
«Эксмо», 2017 — (Top Business Awards)

ISBN 978-5-04-094117-9

Как жить в мире, где информацию о тебе может получить кто угодно и когда угодно? Ведущий мировой эксперт по будущему Big Data Андреас Вайгенд знает ответ. Он разрабатывал стратегию данных таких крупных рыб, как Alibaba, Goldman Sachs, Lufthansa, Thomson Reuters. Работал руководителем по Big Data в компании Amazon, является Директором компании Social Data Lab, преподаёт в IT-школе Калифорнийского университета Беркли. Андреас уверен, что этот «дом со стеклянными стенами» несет в себе не только большие риски, но и потрясающие возможности как для корпораций, так и для обычных людей. Из книги Вы узнаете, как с помощью Big Data инновационные компании: следят за поведением пользователей; определяют круг интересов человека; формируют мнение потребителей.

УДК 004.6  
ББК 32.973-018.2

ISBN 978-5-04-094117-9

© Вайгенд А., 2017  
© Эксмо, 2017

# Содержание

Пролог	6
Введение	9
Глава 1	19
Глава 2	41
Конец ознакомительного фрагмента.	65

# Андреас Вайгенд Big Data

*Вся технология в одной книге*

Andreas Weigend

Data for the People

© Andreas Weigend, 2017

This edition published by arrangement with Levine Greenberg Rostan Literary Agency and Synopsis Literary Agency

Серия «Top Business Awards»

© Богданов С., перевод на русский язык, 2018

© Оформление. ООО «Издательство «Эксмо», 2018

\* \* \*

*Посвящается п., ф. и с.*

## Пролог

### Когда зафиксировано все

*Информация как таковая становится самой значительной отраслью экономики, и базы данных знают о каждом конкретном человеке больше, чем известно ему самому. Чем больше информации о каждом из нас попадает в базы данных, тем в меньшей степени мы существуем<sup>1</sup>.*

**Маршалл Маклюэн**

В 1949 году мой отец, в ту пору двадцатитрехлетний молодой человек, получил место учителя в Восточной Германии. Приехав в город, где ему предстояло работать, папа решил, что ему очень повезло: прямо на вокзале он встретил человека, который тоже искал себе жилье и соседа по комнате. Они нашли себе квартиру, но буквально через пару дней сосед исчез. Папа был озадачен. Спустя несколько дней он был уже не на шутку обеспокоен.

Как-то утром, когда он готовил себе завтрак, в дверь постучали. Папа обрадовался – он решил, что сосед вернулся! Но, открыв дверь, увидел каких-то незнакомых людей. Они сообщили, что ему присуждена премия за успехи в деле народного образования. Премию будут вручать в торжественной обстановке, а их прислали, чтобы сопроводить его на церемонию. Папа не слишком поверил сказанному – уж больно угрюмо выглядели эти мужчины в одинаковых плащах. Но выбора у него не было. Когда его затолкали в ожидавшую на улице машину, он с ужасом обнаружил, что ее дверцы не открываются изнутри. Его арестовали советские власти.

Отца обвинили в шпионаже в пользу американцев. Основанием для обвинения послужило его знание английского языка. Ни семья, ни знакомые не знали, где он. Для них он исчез с лица земли. Его бросили в камеру-одиночку тюрьмы, где он протомился следующие шесть лет. Он так никогда и не узнал ни причины своего ареста, ни причины своего освобождения.

Доступ к личной информации человека – реальная угроза его безопасности, поскольку эти данные могут быть использованы ему во вред. В моих глазах этот риск выглядит особенно очевидным и пугающим, в частности, потому, что я знаю, как собирали и использовали личную информацию против моего отца.

Лет через десять после распада ГДР я попросил дать мне возможность ознакомиться с информацией, которую Министерство госбезопасности, Штази, собирало о моем отце до и после его тюремного заключения. Я был далеко не единственным – с момента падения Берлинской стены с просьбами предоставить доступ к досье Штази на себя или на своих близких обратились почти три миллиона человек<sup>2</sup>. К сожалению, в письме от комиссии по архивам Штази сообщалось, что все материалы, касающиеся моего отца, утрачены.

Но в конверте с письмом обнаружилось кое-что еще – фотокопия обложки досье Штази на меня самого. Я был поражен. Штази вела досье на меня? Я же был просто студентом-физиком. Тем не менее агенты госбезопасности начали собирать информацию обо мне еще в 1979 году, когда я был подростком, а датой последнего обновления значился 1987 год, когда я уже переехал в Штаты. От досье осталась только обложка, и я вряд ли когда-нибудь узнаю, что именно собрала на меня Штази, зачем это было нужно и как использовалось, если использовалось вообще.

---

<sup>1</sup> McLuhan, Marshall, with Wilfred Watson, *From Cliché to Archetype* (Berkeley: Gingko Press, 2011), p. 13. Первое издание этой книги увидело свет в 1970 году.

<sup>2</sup> Pidd, Helen, "Germans Piece Together Millions of Lives Spied on by Stasi", *Guardian*, March 13, 2011, <http://www.theguardian.com/world/2011/mar/13/east-germany-stasi-files-zirndorf>.

Во времена Штази получение информации о «гражданине, представляющем оперативный интерес», было непростым делом. Сначала нужно было собрать данные – организовать слежку, фотографирование, перлюстрацию почты, опрос знакомых и прослушку в доме. Затем все полученные данные скрупулезно анализировались. Работы было столько, что к моменту краха ГДР один процент всех граждан, занятых в народном хозяйстве, являлись штатными сотрудниками госбезопасности. Но для сбора информации Штази требовались еще большие ресурсы<sup>3</sup>. По данным германского федерального правительства, негласными осведомителями властей являлись примерно 200 000 жителей ГДР<sup>4</sup>.

Сегодня собирать данные стало намного проще. Вспомним лишь несколько из наиболее известных примеров. После многомесячных протестов и судебных разбирательств борцам за тайну личной жизни удалось одержать небольшую и неполную победу в деле об упрощенном порядке предоставления Агентству национальной безопасности (АНБ) информации о частных телефонных разговорах<sup>5</sup>. Тем не менее лишь очень немногие решили отказаться от услуг мобильной связи, хотя совершенно очевидно, что метаданные телефонных звонков могут быть доступны АНБ – и не только ему. Так, женщину – торгового агента из Калифорнии уволили с работы за то, что она удалила со своего смартфона приложение, позволявшее менеджеру отслеживать ее местонахождение как в рабочее, так и в нерабочее время<sup>6</sup>. Когда стало известно, что Facebook тщательно исследует распространение настроений пользователей, поднялся шум по поводу «манипулирования» чувствами<sup>7</sup>. Однако на популярности сети это практически не сказалось, и она продолжила эксперименты с данными пользователей без их предварительного согласия по той простой причине, что это крайне необходимо для дизайна платформы. А в 2015 году аффилированная с торговым гигантом Alibaba компания Ant Financial запустила в Китае пилотную версию сервиса Sesame Credit, рассчитывающего рейтинг кредитоспособности частного лица на основе анализа его покупок – как если бы выдачу кредитов американцам одобряли на основе истории их покупок в Amazon<sup>8</sup>. Этот рейтинг моментально стали использовать и в других сферах, в том числе в качестве опции профайла на самом популярном китайском сайте знакомств<sup>9</sup>. Признаков массового отказа от мобильных телефонов, электронных адресов, навигаторов, аккаунтов в социальных сетях, покупок в интернет-магазинах и прочих цифровых услуг не наблюдается. Ведь все эти технологии сильно упрощают жизнь.

---

<sup>3</sup> Koehler, John O., *Stasi: The Untold Story of the East German Secret Police* (Boulder, CO: Westview Press, 1999), p. 8.

<sup>4</sup> Федеральный уполномоченный по архивам Государственной службы безопасности б. Германской Демократической Республики, “What Was the Stasi?”, Bundesregierung, [http://www.bstu.bund.de/EN/PublicEducation/SchoolEducation/WhatWasTheStasi/\\_node.html](http://www.bstu.bund.de/EN/PublicEducation/SchoolEducation/WhatWasTheStasi/_node.html).

<sup>5</sup> Crocker, Andrew, “EFF Case Analysis: Appeals Court Rules NSA Phone Records Dagnet Is Illegal”, Electronic Frontier Foundation, May 9, 2015, <https://www.eff.org/deeplinks/2015/05/eff-case-analysis-appeals-court-rules-nsa-phone-records-dagnet-illegal>.

<sup>6</sup> Kravets, David, “Worker Fired for Disabling GPS App That Tracked Her 24 Hours a Day”, ArsTechnica, May 11, 2015, <http://arstechnica.com/tech-policy/2015/05/worker-fired-for-disabling-gps-app-that-tracked-her-24-hours-a-day>.

<sup>7</sup> В главе 3 я остановлюсь на экспериментах в социальных сетях более подробно. См. McNeal, Gregory S., “Facebook Manipulated User News Feeds to Create Emotional Responses”, Forbes, June 28, 2014, <http://www.forbes.com/sites/gregorymcneal/2014/06/28/facebook-manipulated-user-news-feeds-to-create-emotional-contagion>; и Booth, Robert, “Facebook Reveals News Feed Experiment to Control Emotions”, Guardian, June 29, 2014, <https://www.theguardian.com/technology/2014/jun/29/facebook-users-emotions-news-feeds>.

<sup>8</sup> Sesame Credit – один из восьми пилотных проектов, призванных расширить сферу кредитования в стране к 2020 году. См. Shu, Catherine, “Data from Alibaba’s E-Commerce Sites Is Now Powering a Credit-Scoring Service”, TechCrunch, January 27, 2015, <http://techcrunch.com/2015/01/27/data-from-alibabas-e-commerce-sites-is-now-powering-a-credit-scoring-service>.

<sup>9</sup> Hatton, Celia, “China ‘Social Credit’: Beijing Sets Up Huge System”, BBC News, October 26, 2015, <http://www.bbc.com/news/world-asia-china-34592186>.





Обложка досье Штази

Шок, испытанный при виде досье Штази на себя, мог бы превратить меня в фанатичного ревнителя тайны личной информации. Отнюдь. На самом деле записи Штази – пустяк по сравнению с тем количеством информации о себе, которую я добровольно предоставляю всем желающим изо дня в день.

С 2006 года я публикую на своем личном сайте план всех своих выступлений и лекций, а также всех авиаперелетов, вплоть до номера забронированного в салоне места<sup>10</sup>. Я делаю это, поскольку считаю, что реальная польза от предоставленной о себе информации выше, чем связанные с этим риски. Эта информация создает возможности для получения и оптимизации знаний. Главное – обеспечить, чтобы интересы тех, кто использует эту информацию, не противоречили нашим собственным.

Как можно этого достичь? Через понимание того, какая информация доступна (и наверняка будет доступна в будущем) и как эта информация анализируется и используется компаниями. При всем уважении к Маршаллу Маклюэну я считаю, что чем больше личной информации о нас накапливается в базах данных компаний, тем в большей степени мы существуем и тем больше узнаем о самих себе. Реальная проблема в том, чтобы сделать компании, собирающие личные данные, прозрачными для нас в той же степени, в какой мы прозрачны для них, и обеспечить себе право голоса в вопросах использования этих данных. В этой книге рассказывается о том, как достичь этих целей.

<sup>10</sup> Увидеть, чем я занимаюсь, можно на <http://weigend.com/past> (прошедшие мероприятия) and <http://weigend.com/plans> (текущие и будущие мероприятия).



## **Введение**

# **Революция социальных данных Как можно заставить информацию приносить пользу людям?**

*Всякая революция начиналась с мысли одного человека; а когда та же мысль овладевала другими людьми, она становилась главенствующей для своего времени<sup>11</sup>.*

*Ральф Уолдо Эмерсон*

В 6.45 утра меня будит сигнал будильника в моем мобильном телефоне. Я бодро переминаясь вместе с телефоном на кухню, чтобы начать день с просмотра электронной почты и ленты уведомлений в Facebook. GPS-приемник реагирует на мои перемещения на несколько метров на север и на восток, которые записываются в память телефона. Я наливаю себе кофе и начинаю вести себя более активно. При этом акселерометр телефона отслеживает, насколько быстро я двигаюсь, а барометр фиксирует мой подъем вверх по лестнице. Поскольку на моем телефоне установлены приложения Google, все эти данные попадают в базы этой поисковой системы.

Позавтракав, я отправляюсь на работу в Стэнфордский университет. Энергосбытовая компания установила в моем доме «умный» счетчик, который фиксирует снижение потребления электричества по мере того, как я выключаю свет и отключаю зарядные устройства моих гаджетов. Когда я открываю двери гаража, счетчик отмечает расход электричества, характерный именно для этого события. Поэтому к моменту, когда я выезжаю на улицу, у моей энергосбытовой компании достаточно информации для того, чтобы понять, что я не дома. А когда сигнал телефона переходит к другой вышке сотовой радиосвязи, это понимает и мой мобильный оператор.

Камера, установленная на углу, сфотографирует номерной знак моего автомобиля в случае, если я проеду на красный. Но сегодня я веду себя пайнкой, поэтому появления в почте квитанции со штрафом не предвидится. Тем не менее по пути мой номерной знак попадет под камеры наблюдения еще не раз. Некоторые из этих камер принадлежат районным властям, другие – частным компаниям, анализирующим данные для выявления закономерностей в перемещениях. Результаты их анализа – продукт, который покупают полицейские управления, девелоперы и прочие заинтересованные лица.

Приехав в Стэнфорд, я оплачиваю парковку с помощью приложения EasyPark в моем телефоне. Деньги автоматически списываются с моего счета, а факультет и банк теперь знают, что я приехал в университет ровно в 9.03 утра. Когда телефон перестает перемещаться вместе с машиной, Google решает, что это место парковки, и записывает координаты, на случай, если я вдруг забуду, где оставил машину. А еще пора свериться с приложением страховой компании Metromile, которое считывает данные о моей поездке с бортового компьютера автомобиля. Оно мгновенно сообщает, что расход бензина сегодня был ниже (один галлон на девятнадцать миль) и поездка обошлась мне в 2 доллара 5 центов.

После занятий я планирую повстречаться с одним новым знакомым из Сан-Франциско. Виртуально мы уже встречались, комментируя в Facebook пост одного общего приятеля, и у нас

---

<sup>11</sup> Emerson, Ralph Waldo, *The Prose Works of Ralph Waldo Emerson*, vol. 1, rev. ed. (Boston: James R. Osgood, 1875), p. 220.

обнаружилась общность точек зрения на затронутую тему. Оказалось, что у нас больше тридцати общих знакомых в Facebook – более чем достаточная причина для личного знакомства.

Google Maps прогнозирует, что я окажусь на месте к 19.12, и, как обычно, этот прогноз оказывается точным плюс-минус пара минут. Оказывается, что квартира моего нового знакомого расположена прямо над магазинчиком, который торгует табачными изделиями и разнообразными принадлежностями для употребления марихуаны. GPS-приемнику моего телефона разница между магазином и квартирой, расположенной выше, непонятна, и с точки зрения Google и моего провайдера мой день увенчался посещением хэдшопа. Я понимаю это по рекламе, которую показывает мне Google, когда просматриваю прогноз погоды на завтра перед отходом ко сну.

Революция в использовании социальных данных идет полным ходом.

## Дашь на дашь

Схожие социальные данные ежедневно создают более чем миллиард людей. Социальные данные – это информация о вас, например о ваших перемещениях, поведении и интересах, а также об отношениях, связывающих вас с другими людьми, местами, товарами и даже идеологиями<sup>12</sup>. Некоторые из этих данных предаются огласке сознательно и добровольно, например, когда вы авторизовались в Google Maps и вводите свой маршрут; другие – не столь осознанно, а в качестве неотъемлемой составляющей пользования интернетом и мобильными устройствами. Понятно, что в некоторых случаях предоставление информации является необходимым условием получения услуги: Google не сможет проложить лучший маршрут, если вы не сообщите системе, где находитесь и куда хотите попасть. В других случаях вы сами рады поделиться информацией – например, когда лайкаете пост знакомого в Facebook или даете одобрительный отзыв о работе коллеги в LinkedIn просто потому, что хотите оказать им поддержку.

Социальные данные могут отличаться исключительной точностью, например указывать ваше местонахождение с точностью до метра, но часто бывают отрывочными и недостаточно полными. Например, пока я не зарегистрируюсь в приложении, которое считывает показания моего «умного» электросчетчика (допустим, по дороге в аэропорт, чтобы убедиться, что я действительно выключил весь свет в доме), энергосбытовая компания знает, что меня нет дома, но не более того. Этот касающийся меня показатель с равным успехом может оказаться и полезным, и бесполезным. Так, во время моего визита к знакомому в Сан-Франциско широта и долгота моего местонахождения были отражены точно, а предположения о том, что я делал тем вечером, оказались совершенно неверными. При всем своем правдоподобии вывод Google оказался всего лишь поверхностной интерпретацией. Отрывочные данные обычно оказываются недостаточными, сопряженными с риском ошибки, а иногда и умышленно сфальсифицированными<sup>13</sup>.

В целом же объем социальных данных (пассивных и активных, обязательных и произвольных, точных и приблизительных) растет в геометрической прогрессии: он удваивается каждые восемнадцать месяцев. Через пять лет объем социальных данных возрастет примерно десятикратно, или на порядок, а через десять лет он увеличится примерно в 100 раз. Другими

---

<sup>12</sup> Я преподавал курс под названием «Революция социальных данных» в Стэнфордском университете (с 2008 года) и в Калифорнийском университете в Беркли (с 2011 года), но концептуальным представлением «социальные данные» стал заниматься еще задолго до этого. На самых ранних порах к социальным данным относили всего лишь обнародованную человеком информацию, вроде отзывов в Amazon и постов в социальных сетях.

<sup>13</sup> Тем, кого интересуют подробности об отрывочных данных, рекомендую посмотреть видеозапись панельной дискуссии с моим участием на конференции DataEdge, которую проводил факультет информации Калифорнийского университета в Беркли в 2013 году. Она доступна на <http://www.catchtalk.tv/events/dataedge/videos/sketchy-data-panel-discussion-dataedge-2013>.

словами, сейчас за один день фиксируется столько же данных, сколько в течение всего 2000 года. А при сохранении существующих темпов роста в 2020 году мы будем создавать такой же объем данных менее чем за час.

Очень важно понимать, что «социальные данные» – отнюдь не просто очередное модное наукообразное словосочетание применительно к социальным медиа. Многие платформы социальных сетей создавались в целях широкого охвата массовой аудитории. Социальные данные становятся все более демократичными и доступными: информацией о себе, своей компании, своих успехах и своей точке зрения можно добровольно делиться в Twitter или в Facebook. Но люди оставляют намного большее количество глубоких цифровых следов на куда более обширной территории. Ваши поиски в Google, ваши покупки в Amazon, ваши звонки по скайпу, каждомomentное местонахождение вашего телефона – все эти и многие другие источники позволяют создать уникальный портрет вашей личности.

Далее, социальные данные – это не только вы сами. Характер ваших коммуникаций с родными, знакомыми и коллегами представляет собой информацию о прочности ваших связей с ними. Вы пополняете картину социальных данных и в ходе разовых контактов с совершенно незнакомыми людьми – так происходит, например, когда вы вводите тэги в Инстаграме или оставляете отзыв на какой-то товар. Создавая аккаунт на сайте аренды жилья Airbnb, вы подтверждаете свою личность не только официальными паспортными данными, но и профайлом в Facebook. Фиксация социальных данных предусмотрена и в домах с «умными» кондиционерами, и в автомобилях с навигационными системами, а также на рабочих местах с программными средствами коллективного пользования. Эти данные начинают аккумулироваться в учебных аудиториях и кабинетах врачей. По мере того как мобильные телефоны обрастают все большим количеством датчиков и приложений, отслеживающих ваше поведение дома, в магазине и на работе, становится все менее возможным контролировать распространение информации не только о своем обычном образе жизни, но и о своих сокровенных желаниях. Специалисты по обработке и анализу данных превращаются в детективов и художников, способных создавать все более и более точные образы личности на основе оставленных цифровых следов.

Эти цифровые следы исследуются и обрабатываются для того, чтобы получать представление о наших предпочтениях, выявлять тренды и делать прогнозы, в том числе и о возможных покупках. В качестве главного научного консультанта Amazon я вместе с Джеффом Безосом разрабатывал информационную стратегию компании, основанную на анализе интересов потребителя. Наши эксперименты были призваны установить, что в большей степени влияет на удовлетворенность покупателя продуктом – отзывы других потребителей или отзывы в прессе. Мы также хотели выяснить, что работает эффективнее – рекомендации на основе традиционных демографических характеристик или на основе истории просмотров каждого потенциального покупателя. В результате мы убедились, что эффективнее всего стимулирует продажи метод непосредственной коммуникации. Созданные нами для Amazon инструменты персонализации радикально изменили подходы людей к решению о покупке товара и стали стандартом электронной коммерции.

---

# ДОСТУП К ЛИЧНОЙ ИНФОРМАЦИИ ЧЕЛОВЕКА — РЕАЛЬНАЯ УГРОЗА ЕГО БЕЗОПАСНОСТИ

---

После ухода из Amazon я вел учебный курс «Революция в использовании социальных данных», который прослушали тысячи студентов – от старшекурсников и аспирантов Стэнфорда и Калифорнийского университета в Беркли до слушателей китайских бизнес-школ Фуданьского университета в Шанхае и Университета Синьхуа в Пекине.

Кроме этого я продолжаю руководить созданной мной в 2011 году Лабораторией социальных данных (Social Data Lab) – коллективом авторитетных специалистов в области обработки и анализа информации. На протяжении десятилетия работы с заказчиками, в числе которых крупнейшие корпорации (Alibaba, AT&T, Walmart и United Healthcare), авиакомпании, представители финансового сектора и сайты знакомств, я неизменно выступаю за предоставление права участия потребителей и пользователей, то есть обычных людей, в решении вопросов использования данных.

Проработать весь доступный в наши дни объем данных по какому-то вопросу, для того чтобы принять, как говорили раньше, «решение на основе полной информации», не под силу ни одному человеку. Но кто будет иметь возможность пользоваться инструментарием для обработки информации, пусть и в целях решения наших проблем? Будут ли сделанные на основе анализа данных выводы и прогнозы доступны любому желающему или они останутся в распоряжении немногочисленных могущественных организаций? Какую цену придется заплатить каждому из нас за бонусы от доступности своих социальных данных?

Я считаю, что значение социальных данных следует рассматривать не только в разрезе их доступности, но и в контексте результативности их использования, улучшения качества принимаемых нами решений. Некоторые решения принимаются нами ежедневно и неоднократно, а некоторые – один раз в жизни. Однако это не значит, что у созданных нами сегодня социальных данных короткий срок годности. Наши сегодняшние действия могут обусловить решение, которое нам предстоит сделать через несколько десятилетий. Лишь очень немногие умеют тщательно фиксировать все свои действия и анализировать их влияние на свою жизнь в краткосрочной или долгосрочной перспективе. Анализ социальных данных может позволить лучше оценить возможности и вероятности, притом что окончательное решение должно оставаться за вами.

Единственное, на что совершенно точно не способны эти технологии, – решить, какое будущее мы хотим для себя и как отдельно взятые личности, и как общество в целом. Законы, защищающие человека от дискриминации на работе или в сфере здравоохранения, могут прекратить действовать уже завтра, а в некоторых странах они отсутствуют и сейчас. Представьте себе, что вы решили поделиться своим беспокойством по поводу высокого уровня холесте-

рина с каким-нибудь медицинским приложением или сайтом, чтобы получить консультацию о режиме питания и физических нагрузках. Могут ли ваши опасения по поводу своего здоровья быть каким-то образом использованы против вас? Что, если законодательство позволит установить для вас лично максимально высокие цены на медицинское обслуживание, поскольку вы не перестали питаться полуфабрикатами и продолжаете валяться на диване и после того, как получили четкие рекомендации относительно образа жизни, который следует соблюдать при вашем состоянии здоровья? А если некий менеджер кадровой службы получит информацию о вас из интернета и решит, что ваш стиль жизни не подходит компании, и поэтому он не будет рассматривать вашу кандидатуру на имеющуюся вакансию? Это вполне реальные риски.

Если бы вы являлись единственным источником формирования и распространения данных о себе, то имели бы возможность пресечь появление информации, которую считаете нежелательной. Это было бы возможно, хотя и крайне хлопотно. Однако мир, в котором мы живем, устроен иначе. У вас нет возможности контролировать большую часть информации о себе. Это становится все более очевидным по мере того, как бизнес и власти все чаще прибегают к использованию социальных данных под предлогом повышения своей эффективности и ответственности.

В связи с такой доступностью социальных данных вопрос о том, как обращаться с ними наилучшим образом, касается каждого. Технологии быстро развиваются, а компании, которые собирают и анализируют наши данные, занимаются аккумулярованием и кодированием информации, а не разработкой правил ее использования. Многие принципиальные вопросы рассматриваются применительно к какой-то конкретной ситуации, если рассматриваются вообще. Нельзя оставлять на усмотрение информационно-технологических компаний решения относительно принципов, глубоко затрагивающих наше будущее.

Можно согласиться с тем, что все эти данные собирают, сводят, агрегируют и анализируют для того, чтобы мы имели возможность лучше понимать плюсы и минусы предлагаемого нам выбора. Мнение человека – основополагающий элемент при оценке альтернатив, которые предполагает любое важное решение. Информация не должна *руководить* нами. Она должна *расширять спектр наших возможностей*.

## Правила для эры постприватности

Понимание возрастающего значения информации в жизни человека стимулировало ряд инициатив, направленных на защиту интересов граждан. В 1970-х годах в США и Европе были разработаны примерно одинаковые принципы добросовестного использования информации. Частным лицам было предоставлено право знать, кто собирает о них сведения и как эти данные используются. Кроме того, граждане могли корректировать неточные данные о себе<sup>14</sup>. Для современных информационных потоков и средств аналитики эти защитные меры парадоксальным образом и избыточно жестки, и совершенно недостаточны одновременно.

Избыточная жесткость выражается в возможности отследить и сохранить любые данные о вас. Компания Amazon наверняка сможет доступно объяснить, как она использует полученную о вас информацию. Наверное, она даже сможет сделать это так, чтобы вы использовали это знание для принятия более обоснованных решений. Но ознакомление со всеми этими данными займет уйму времени. Многие ли захотят тратить время на причесывание всего информаци-

---

<sup>14</sup> Огромная разница между подходами к защите данных между США и Европой состоит, в том числе, и в правоприменении. В США регулирующие органы для защиты граждан от рисков незаконного использования личных данных создаются отдельно для каждой отрасли, тогда как в Европе существуют единые законодательные нормы. См. издание Аппарата Президента США *Big Data: Seizing Opportunities, Preserving Values*, (Совместный доклад рабочей группы по большим данным и защите неприкосновенности личной жизни и Совет по развитию науки и техники), Май 2014, pp. 17–18, [https://www.whitehouse.gov/sites/default/files/docs/big\\_data\\_privacy\\_report\\_may\\_1\\_2014.pdf](https://www.whitehouse.gov/sites/default/files/docs/big_data_privacy_report_may_1_2014.pdf).

онного массива? Действительно ли вам нужно понимать, каким образом в Amazon оценивают каждый параметр, или будет вполне достаточно итогового результата?<sup>15</sup>

В то же время эти защитные меры совершенно недостаточны, поскольку, даже если вы в состоянии контролировать каждый бит информации о себе, созданной и предоставленной лично вами, полной картины своих обращающихся личных данных вы не получите. Дело в том, что в эту картину попадает и та информация о вас, которую создают другие люди, например ваши родственники, друзья, коллеги и работодатели. Бизнесы, которые вы посещаете как онлайн, так и физически, тоже создают (а иногда и распространяют) данные о вас. То же относится и ко всем остальным, с кем вы взаимодействуете, – от разнообразных государственных и частных организаций до прохожих на улице. Кто поручится за достоверность этих данных? Поскольку в наши дни информация поступает из столь многочисленных источников, вопрос отнюдь не исчерпывается правом корректировать свои личные данные. Кроме того, даже достоверная информация может быть использована вам во вред.

На фоне масштабных количественных и качественных сдвигов в области создания, передачи и обработки информации становится совершенно очевидным, что права знать и права корректировать недостаточно. До настоящего времени попытки привести существующие правила в соответствие с требованиями времени были направлены почти исключительно на поддержание возможности индивидуального контроля над данными и сохранения тайны личной жизни<sup>16</sup>. К сожалению, такой подход порожден опытом прошлого века технологий. Стандарты контроля и конфиденциальности уже заставляют людей вступать в неравные отношения с информационно-технологическими компаниями. Если вы хотите повысить качество принятия решений с помощью информации, вам, скорее всего, придется согласиться с тем, что ваши данные будут собирать на условиях, которые вам продиктуют. И как только вы их примете, процесс сбора личных данных будет считаться соответствующим требованию законодательства о «контроле» вне зависимости от того, предоставляет ли вам компания – сборщик информации какие-либо варианты выбора и предохраняет ли от возможных нежелательных последствий для сохранения тайны личной жизни. Если вы хотите сохранить конфиденциальность своих личных данных, вам не следует давать согласия на их сбор и обработку. Правда, в результате вы лишитесь доступа к соответствующим информационным ресурсам и сервисам, а ценность ваших личных данных снизится. Вот и радуйтесь своему праву личного контроля.

Сегодня нам нужны стандарты, позволяющие оценить риски и пользу распространения и использования данных, а также позволяющие привлекать компании к ответственности. Два десятилетия сотрудничества с компаниями информационно-технологической отрасли привели меня к твердой убежденности в том, что соблюдение принципов прозрачности и свободы выбора – самое перспективное направление защиты от злоупотреблений в области обработки личных данных, позволяющее в то же время максимизировать выгоды от их использования.

Под *прозрачностью* понимается право людей знать, что происходит с их данными: что они собой представляют, куда они направляются и как способствуют получению результата пользователем. Наблюдает ли компания за вами из-за непрозрачной стороны зеркала или предоставляет окно, через которое можно видеть, что она делает с вашими данными, и, таким образом, иметь возможность судить, насколько ее интересы соответствуют вашим (и соответ-

---

<sup>15</sup> Для понимания алгоритма обычно бывает нужно увидеть, как он работает с реальными данными, в идеале – вашими собственными в сочетании с данными других людей (что дает больше возможностей для сравнения). В соответствии с действующим законодательством большинство компаний не имеют права предоставить вам информацию о других людях без их согласия. Это ограничивает возможности расшифровки алгоритмов вашей жизни.

<sup>16</sup> См., в частности, законопроект «О правах потребителя в области неприкосновенности частной жизни», внесенный Белым домом в феврале 2012 года, о котором говорится на стр.19–20 издания Аппарата Президента США Big Data: Seizing Opportunities, Preserving Values See.

ствуют ли вообще)<sup>17</sup>? Сколько личной информации нужно предоставить, чтобы получить нужный продукт или услугу? Исторически между частными лицами и организациями существовал серьезный информационный перекос в пользу организаций. Организации обладают не только большими возможностями для сбора информации о человеке, они могут также интерпретировать его данные, сравнивать с данными других. Баланс между тем, что вы предоставляете и что получаете взамен, должен быть совершенно прозрачен.

Оцените прозрачность в процессе покупок на Amazon по сравнению с традиционными отношениями между покупателем и розничной торговлей. Станет ли продавец в обычном магазине напоминать вам о том, что вы уже покупали этот товар, рискуя тем самым потерять в продажах? Если вы собираетесь купить на Amazon книгу, которую уже покупали на сайте, вас вежливо спросят: «Вы уверены? Вы уже покупали этот товар 17 декабря 2013 года». Если вы купили один трек из музыкального альбома, а теперь хотите купить весь альбом, Amazon «завершит покупку», то есть автоматически снизит цену на сумму, уплаченную за купленный трек. Компания отслеживает истории покупок и использует их для того, чтобы минимизировать поводы для огорчения своих покупателей. Схожим образом поступают и авиакомпании, когда информируют вас о том, что ваши накопленные мили скоро «сгорят», вместо того чтобы позволить этому случиться.

К сожалению, сегодня прозрачность – далеко не норма. Вспомните более чем типичный пример звонка в какой-нибудь центр обслуживания клиентов. Сразу же после соединения вы обязательно услышите предупреждение «Внимание! В целях контроля качества этот разговор может быть записан». Выбора у вас нет – если вы хотите поговорить с сотрудником компании, придется принять это условие. Допустим. Но почему эта запись будет доступна только компании? Что это за «цели контроля качества», когда только одна сторона имеет доступ к записи разговора? Принцип симметрии данных подразумевает, что и у вас – потребителя, который платит, – должен быть прямой доступ к этой записи.

Каждый раз, услышав, что мой телефонный разговор может записываться, я заявляю сотруднику клиентской службы, что я, пожалуй, тоже запишу этот диалог с целью контроля качества предоставляемых услуг. В большинстве случаев сотрудник вынужденно соглашается. Но случается, что на том конце вешают трубку. Конечно, я могу записать разговор и не спрашивая согласия сотрудника, – хотя, замечу, что кое-где это считается противозаконным. Тогда, если качество сервиса не будет соответствовать обещанному, я смогу обратиться к начальству с доказательствами на руках. Если и это не поможет, я могу выложить запись в сеть в надежде, что она станет вирусной и компании придется срочно исправлять ситуацию. Так произошло с Comcast<sup>18</sup>: один из клиентов пытался отменить подписку и раз за разом наткнулся на отказ, до тех пор пока не выложил запись разговоров в Twitter, где она попала в тренды<sup>19</sup>.

Закон не должен ограничивать возможности человека играть на равных с компаниями. Чтобы прозрачность стала новой нормой, надо, чтобы общедоступным оказался больший, а не меньший объем информации.

Но одной прозрачности недостаточно, нужна также и свобода выбора<sup>20</sup>. Под ней подразумевается право человека управлять своими данными. Легко ли вам разобраться в том, что представляют собой регламенты работы с личной информацией, которыми данная компания

---

<sup>17</sup> Спасибо Эстер Дайсон, предложившей эту метафору с зеркалом и окном.

<sup>18</sup> Comcast – американская компания, крупнейший в мире оператор кабельных и эфирных телесетей. – *Прим. пер.*

<sup>19</sup> Если вы упустили это в СМИ, то можете послушать душераздирающую запись звонка на SoundCloud: Block, Ryan, “Comcastic Service Disconnection (Recording Starts 10 Mins into Call)”, SoundCloud, July 14, 2014, <https://soundcloud.com/ryan-block-10/comcastic-service>.

<sup>20</sup> Мой друг Док Серлс, соавтор *The Cluetrain Manifesto* и автор *The Intention Economy*, давно ратует за право выбора в взаимодействии потребителей и компаний, и я благодарен ему за предложение использовать этот термин для обозначения всей полноты аспектов личного волеизъявления, воплощенных в этом праве.



руководствуется «по умолчанию»? Разрешается ли вам изменять свои данные по собственному усмотрению? Можете ли вы получать результаты в удобном для вас виде или же вас ненавязчиво склоняют (или насильно заставляют!) пользоваться определенными опциями, обычно теми, которые больше устраивают компанию? Можно ли поиграть с параметрами и попробовать разные сценарии, чтобы рассмотреть суженный или расширенный диапазон возможностей? Свобода выбора – личное право человека, реализуемое путем управления параметрами и алгоритмами, выявленными информационно-технологическими компаниями. Она предполагает возможность требовать от компаний предоставления информации *на ваших условиях*.

На фундаментальном уровне свобода действия подразумевает наличие у человека возможности создавать полезные именно ему данные. Amazon в полной мере воплотила идею нецензурируемых отзывов. Компании совершенно все равно, положительные это отзывы или отрицательные, пяти- или однозвездочные, написанные с целью заслужить одобрение окружающих или в рамках воплощения заветной мечты стать литературным критиком. Важно то, что они полезны другим пользователям, которые решают, что им купить. Например, отзывы могут показать, что покупатель жалеет о сделанном выборе даже в случае, когда он не возвращает товар продавцу. Такого рода данные помогают потребителю решить, насколько рекомендуемый товар может соответствовать его ожиданиям. Amazon предоставила потребителю большую свободу выбора.

Маркетологи любят поговорить о таргетировании, сегментации и конверсии. Не знаю, как вы, но лично я не хочу, чтобы меня таргетировали, сегментировали, конвертировали или нарежали тонкими ломтиками. Свободе выбора чужды подобные понятия. Не стоит ожидать, что однажды руководители всех компаний по своей собственной инициативе начнут руководствоваться принципами прозрачности и предоставления пользователю свободы выбора. И нам требуются четко очерченные границы прав, позволяющие сделать прозрачность и свободу выбора основой создания реального, практически полезного инструментария.

Если мы сможем заставить компании, занимающиеся сбором данных, согласиться с набором взвешенных правил и способов работы, это приведет к тому, что я называю «сменой значений» – полному пересмотру сложившейся практики в отношениях между частными лицами и учреждениями. Решение компании Amazon передать потребителям формирование большей части контента, связанного с товарами, – один из примеров такой смены значений, и революция в области социальных данных способна предоставить еще больше подобных возможностей. Одновременно с расширением инструментария, помогающего человеку принимать наиболее подходящие ему решения, падает эффективность традиционной рекламы и манипулирования потребителем. Времена, когда покупки осуществлялись под диктовку компаний, миновали. В скором будущем вы сможете диктовать компании, что она должна произвести для вас. Кое-где это уже возможно.

Смены значений – важный элемент картины мира с точки зрения физики. Часто они обозначают фазовые переходы, когда некая переменная во внешней среде вызывает резкие изменения свойств вещества, например переход воды из жидкого в газообразное состояние при достижении точки кипения. Воздействие, которое возрастающий объем информации оказывает на общество, можно сравнить с воздействием возрастающей температуры нагревания на физическую систему. При определенных условиях, то есть когда информационно-технологические компании предоставят пользователям прозрачные условия и свободу выбора, произойдет смена значений в пользу частного лица, то есть главным бенефициаром во взаимоотношениях человека и компании станет человек, а не компания или ее директор по маркетингу.

Заинтересованными лицами революции в использовании социальных данных являемся мы все. А тот, кто хочет извлечь из социальных данных пользу, должен делиться информацией о самом себе. Точка. Ценность социализации информации часто проявляется в виде расширения возможностей для принятия оптимальных решений в коммерческих переговорах, при

покупке товаров и услуг, получении кредитов, в поиске работы, при получении образовательных и медицинских услуг и в общественной жизни в целом. Цена и риски предоставления личных данных должны быть как минимум эквивалентны получаемым взамен благам. Существенна прозрачность того, что узнают о человеке компании и что они делают с этой информацией. То же относится и к возможности частного лица иметь определенный контроль над информационными продуктами и услугами. Без этого судить о том, насколько эквивалентным является происходящий обмен, невозможно.

## Баланс сил

Доступ к информации – одно из главных преимуществ. Те, у кого ее больше, почти всегда оказываются в выигрыше, как пресловутый продавец подержанных машин, втюхивающий развалюху несведущему покупателю. С удешевлением и повсеместным распространением средств коммуникации и обработки данных информации стало намного больше, соответственно намного возрос и риск существенных информационных перекосов, поскольку ни одному частному лицу не под силу справиться с обработкой всех имеющихся данных.

Большая часть создаваемой и распространяемой информации касается личной жизни: где мы живем, где работаем, где бываем, кого любим, а кого – нет, с кем проводим время, что едим, насколько регулярно занимаемся спортом, какие лекарства принимаем, какими бытовыми приборами пользуемся, что затрагивает наши чувства. Для компаний, собирающих и анализирующих данные, наша жизнь прозрачна. Иногда эти компании занимаются незаконной перепродажей информации, но чаще держат ее у себя для применения исключительно по собственному усмотрению. Необходимы более детальные правила использования личных данных, не допускающие возможности их изменения или обмена ими в какой-либо форме без разрешения их носителя. И те, кто создает данные, и те, кто их собирает и анализирует, должны обладать равной прозрачностью друг для друга и свободой выбора.

Для этого нам потребуется серьезно пересмотреть свое отношение к личным данным и самим себе. В первой главе я расскажу о некоторых способах, которые применяют информационно-технологические компании для анализа личных данных и превращения их в продукты и сервисы. В главе 2 речь пойдет о том, как цифровые следы нашей жизнедеятельности (история поиска, клики, просмотры, нажатия и свайпы) «затаптывают» иллюзии приватности, формируя представление о нашей личности и подавая сборщикам данных важные сигналы вне зависимости от того, хотим мы этого или нет. В главе 3 в центре моего внимания окажутся связи между людьми и то, как в цифровую эпоху социальные сети меняют представления о доверии. В главе 4 рассмотрим, насколько четкую картинку наших перемещений, настроений и уровня внимания позволяет получить анализ данных, полученных из объединенных сетей самых разнообразных датчиков (не только видеокамер).

Исходя из этого анализа, я формулирую шесть видов прав, соблюдение которых считаю необходимым условием того, чтобы в будущем данные о людях служили именно им самим. Право на доступ к данным и право на проверку компаний, собирающих и обрабатывающих данные, относятся к вопросу повышения прозрачности. Право на исправление данных, право на удаление данных, право на экспериментирование с данными и право на перенос данных в другие компании относятся к сфере повышения степени контроля человека над своими личными данными и обеспечения свободы его выбора. Распространение этих прав на личные данные и их использование будет иметь значительные последствия для всех нас. В заключительной главе, посвященной воплощению этих прав в реальность, мы рассмотрим, как это скажется на покупках, платежах, инвестициях, работе, учебе и управлении общественными ресурсами.

Мы находимся на переломной стадии оформления отношений между людьми, передающими свои данные, и компаниями, создающими на их основе продукты и услуги. Это не

просто изменения в правилах игры – сама игра стала качественно иной. Ее новые правила требуют пересмотра отношений между покупателями и продавцами, инвесторами и банками, работниками и работодателями, пациентами и врачами, учениками и учителями, гражданами и государством. Пора осознать, что в действительности означает использование данных, какие выгоды можно из этого извлекать и как контролировать последствия этого процесса, и занять по этим вопросам твердую позицию. Тогда мы сможем понять, насколько наши собственные интересы совпадают с интересами компаний, собирающих информацию о нас. Как и в подавляющем большинстве случаев, все изменения идут не от новых технологий и машин. Революция происходит в результате корректировки ожиданий и социальных норм в эпоху распространения новых технических средств.

Данные на людей могут стать данными для людей, если мы примем вызов и окажемся на высоте в решении этой задачи. Добро пожаловать в революцию.

# Глава 1

## Как научиться разбираться в данных

### Основной инструментарий цифрового гражданина

*В XVIII веке грамотным считался человек, способный прочитать вслух знакомые ему куски из Библии или катехизиса; в наши дни человека, не умеющего большего, сочли бы функционально неграмотным – неспособным усвоить тексты, которые необходимы для экономического выживания<sup>21</sup>.*

*Джордж Миллер*

***Как устроена обработка данных и какая именно информация о вас представляет ценность?***

«Данные для людей» – не просто лозунг. В повседневной жизни мы постоянно сталкиваемся с информационными продуктами и услугами в виде рейтингов и рекомендаций, созданных на основе социальных данных. На смену традиционным рекламным деятелям пришли специалисты по обработке и анализу данных, которые пропускают через просчитанные алгоритмы несметные количества цифровых следов миллиардов людей. Изменения в ментальности имеют даже более важное значение, чем лавинообразный рост количества получаемой нами информации. Чтобы сделаться полноценным участником революции в использовании социальных данных, необходимо отбросить старые «потребительские» установки по принципу «бери, что дают» и переключиться на новое мышление – мышление активного созидателя социальных данных. Происходят сдвиги в балансах сил между продавцами и покупателями, банкирами и заемщиками, работодателями и работниками, докторами и пациентами, учителями и учениками. Информация о людях и от людей может и должна стать информацией для людей.

На самом деле нет ничего важнее запроса на информацию для людей. Социальные данные – самое ценное сырье XXI века, новая нефть<sup>22</sup>. Такая аналогия представляется вполне уместной по целому ряду причин. На протяжении более чем ста лет нашу экономику и общественно-политическую жизнь во многом определяли нефть и развитие технологий ее добычи, хранения и переработки в продукцию, потребляемую каждым жителем планеты. Сегодня возможность переработки персональных данных в продукты и услуги привносит в жизнь человечества изменения, сопоставимые с эффектом промышленной революции.

Нефть не используется в ее первичном состоянии. Ее нужно переработать в автомобильное горючее, пластмассы и многие другие продукты нефтехимии. В свою очередь нефтепереработка дала толчок развитию техники индустриальной эпохи и сыграла важную роль в производстве подавляющего большинства видов физической продукции современной экономики. Схожим образом и первичные персональные данные сами по себе являются достаточно бесполезными. Данные приобретают ценность в результате переработки – агрегирования, анализа, сравнения, фильтрации и дистрибуции новых информационных продуктов и сервисов. В отли-

---

<sup>21</sup> Miller, George A., “The Challenge of Universal Literacy”, Science 241 (September 9, 1988), p. 1293, <http://science.sciencemag.org/content/241/4871/1293>.

<sup>22</sup> Я использую эту метафору в своей преподавательской работе уже много лет и дважды выступал на тему переработки первичных данных в 2011 году – в ООН и на конференции O’Reilly Strata Summit. Но в этом я не одинок. Среди тех, кто говорит о данных, как о новой нефти, – Клайв Хамби, участвовавший в создании клубной карты британской сети супермаркетов Tesco’s – одной из первых карт лояльности, учитывавшей все покупки в корзине покупателя. В ООН я выступал в рамках программы Global Pulse, запущенной по инициативе Генерального секретаря. Видеозапись доступна на <http://www.youtube.com/watch?v=lbmsDH8RJA4>.

чие от продукции нефтепереработки, продукты переработки данных становятся основой не индустриальной революции, но революции социальных данных.

К счастью, информация как ресурс радикальным образом отличается от нефти. Запасы нефти на планете конечны, и по мере истощения этого ресурса его эксплуатация обходится все дороже. Количество же информации, напротив, растет в геометрической прогрессии, а стоимость технологий ее передачи и обработки неуклонно снижается. По состоянию на конец 2015 года смартфонами пользовались более половины совершеннолетних граждан<sup>23</sup>. Среднестатистический американец проводит в разговорах по мобильному телефону примерно два часа в день<sup>24</sup>. По существующим оценкам, в течение дня люди прикасаются к своим телефонам от двухсот до трехсот раз – едва ли не чаще, чем большинство из нас прикасается к своей партнерше за месяц<sup>25</sup>. Информация, в отличие от нефти, не иссякнет никогда.

Использование нефти ограничено факторами ее дефицита и вещной формы, а использование информации практически не ограничено, учитывая ее изобилие и цифровую форму. Конкретный объем сырой нефти может быть использован для переработки в готовую продукцию лишь единственным владельцем, тогда как одним и тем же информационным массивом могут одновременно пользоваться и превращать его в разнообразные продукты многие. В основе наших законов и социальных норм лежит идея дефицитности информации. Так, например, отсутствие информационного поля породило страхование как способ защиты от неприятностей и материальных потерь, связанных с чрезвычайными жизненными обстоятельствами. Поскольку возможность точно рассчитать вероятность заболевания диабетом или ограбления для каждого конкретного человека отсутствовала, страховщики делили людей на группы риска и устанавливали для каждой из них среднюю ставку страховых взносов. По мере роста объема информации мы сможем прогнозировать риск для каждого конкретного человека и индивидуализировать суммы страховых взносов. Можно делать вид, что данных не существует, но можно признать факт их наличия и задуматься над тем, как это может изменить нашу жизнь. Какой мир нам хотелось бы создать при помощи этого нового вида ресурсов?

Богатый потенциал возможностей новых технологий можно раскрыть при наличии соответствующих инструментов. До изобретения Гутенбергом печатного станка книг было мало, а доставка новостей жителям отдаленных местностей обходилась дорого. Большая часть населения не извлекала никакой выгоды от долгих часов, потраченных на обучение чтению. Профессор психологии Принстонского университета Джордж Миллер писал о современных стандартах грамотности еще до изобретения интернета. Его беспокоило, что слишком многие ученики не достигают в чтении, математических и научных дисциплинах того уровня навыков, которые необходимы для получения работы в экономике «интеллектуальных услуг»<sup>26</sup>. Я считаю, что сегодня налицо другая, не менее насущная потребность в грамотности нового типа – информационной. Она подразумевает понимание процесса переработки информации, знание изменяемых и неизменяемых параметров, умение интерпретировать ошибки и понимать возможные последствия передачи информации о себе. Для мира, где большую часть наших решений будут направлять рекомендации, полученные на основе обработки социальных данных, такая грамотность является обязательной.

---

<sup>23</sup> “Planet of the Phones”, The Economist, February 28, 2015, <http://www.economist.com/news/leaders/21645180-smartphone-ubiquitous-addictive-and-transformative-planet-phones>; Rogowsky, Mark, “More Than Half of Us Have Smartphones, Giving Apple and Google Much to Smile About”, Forbes, June 6, 2013, <http://www.forbes.com/sites/markrogowsky/2013/06/06/more-than-half-of-us-have-smartphones-giving-apple-and-google-much-to-smile-about>.

<sup>24</sup> Lunden, Ingrid, “80 % of All Online Adults Now Own a Smartphone, Less Than 10 % Use Wearables”, TechCrunch, January 12, 2015, <http://techcrunch.com/2015/01/12/80-of-all-online-adults-now-own-a-smartphone-less-than-10-use-wearables>.

<sup>25</sup> Tecmark, “Smartphone Usage Statistics 2014: UK Survey of Smartphone Users”, October 8, 2014, <http://www.tecmark.co.uk/smartphone-usage-data-uk-2014>.

<sup>26</sup> Miller, George A., “The Challenge of Universal Literacy”, Science 241 (September 9, 1988), p. 1293.

## Процесс переработки данных

Неудивительно, что одним из первых «инфоперерабатывающих заводов» стало предприятие розничной торговли – компания Amazon. Преуспевающий магазин обязан знать, какие товары, интересующие потенциальных покупателей, должны быть в наличии, а для этого нужно отслеживать данные о товарообороте, ценах, рекламе и потребительских предпочтениях своей целевой аудитории.

Двести лет назад практически вся информация, нужная хозяину магазина, заключалась в данных об остатках товара на полках и денег в кассе. По окончании каждого торгового дня эти данные вписывали чернильной ручкой в гроссбух. При выборе покупок из примерно одинакового ассортимента в одной и той же ценовой категории покупатель руководствовался информацией о надежности, привлекательностью упаковки или же мнением друзей, соседей и родственников. Примерно 150 лет назад несколько компаний, самыми известными из которых были Montgomery Ward и Sears & Roebuck Company, порадовали жителей провинциальных американских городков каталогами для заказа товаров почтой. Эти инноваторы своего времени знали, что именно обычно заказывает конкретный покупатель и куда ему доставляют товары, и поэтому могли определять уровень спроса на отдельные виды продукции в разрезе регионов. Сто лет назад, для того чтобы прогнозировать спрос и оптимизировать товарные запасы, компании, занимавшиеся торговлей по каталогам, открывали и шоурумы, и обычные магазины, а также держали целые армии аналитиков, прочесывавших статистику продаж<sup>27</sup>. Спустя еще пятьдесят лет в розничной торговле опять произошли важные перемены. С появлением системы почтовых индексов в США рассылочным фирмам и их торговым точкам стало проще отслеживать особенности своих потребителей<sup>28</sup>. В течение двух следующих десятилетий компании смогли собрать подробную демографическую информацию о людях, проживающих в различных географических областях. А вошедшие в обиход американцев с середины 1960-х годов кредитные карточки позволили собирать данные о покупках конкретного потребителя. До наступления эпохи интернета это был предел детализации личных данных – где человек живет и сколько и где он тратит.

Основанная в 1969 году компания-брокер данных Acxiom и ряд других вдоль и поперек анализировали данные домохозяйств, распределяя частных лиц по потребительским нишам, в которых фигурировали, например, сегменты «Образцово-показательных граждан», «Барских домов», «Селян с дробовиками» и «Пригородных наседок». И это еще не самые худшие образцы социальной стереотипизации<sup>29</sup>. В качестве источников информации у этих брокеров были только официальная статистика и отчетность о покупках по каталогам<sup>30</sup>. Например,

---

<sup>27</sup> Madison, James H., “Changing Patterns of Urban Retailing: The 1920s”, *Business and Economic History*, vol. 5 (1976), p. 104, <http://www.thebhc.org/sites/default/files/beh/BEHprint/v005/p0102-p0111.pdf>.

<sup>28</sup> Clark, Anna, “The Tyranny of the ZIP Code”, *New Republic*, March 8, 2013, <https://newrepublic.com/article/112558/zip-code-history-how-they-define-us>.

<sup>29</sup> “Образцово-показательные граждане” – название одного из сегментов Acxiom. См. Hicken, Melanie, “What Type of Consumer Are You?”, *CNNMoney*, April 19, 2013, <http://money.cnn.com/2013/04/18/pf/consumer-type/>. “Барские дома” и “Селяне с дробовиками” – две категории программы «Potential Rating Index by Zip Markets» (PRIZM), которую разработала маркетинговая фирма Claritas, созданная в 1990-х годах. Сейчас Claritas – одно из подразделений Nielsen Company – компании, собирающей для производителей информацию о продажах их продукции в рознице и измеряющей телеаудиторию. См. Kotler, Philip, and Kevin Lane Keller, *Marketing Management 14* (Upper Saddle River, NJ: Prentice-Hall, 2012), p. 215. “Пригородные наседки” – категория базы данных розничной сети Best Buy, которая содержит информацию о 75 миллионах домохозяйств-покупателей и является одной из самых смелых попыток крупной розницы использовать данные для персонализации предложений. См. Kotler, Philip, and Kevin Lane Keller, *Marketing Management 14* (Upper Saddle River, NJ: Prentice-Hall, 2012), p. 71; и Zmuda, Natalie, “Best Buy Touts Data Project as Key to Turnaround”, *Advertising Age*, February 27, 2014, <http://adage.com/article/datadriven-marketing/buy-touts-data-project-key-turnaround/291897>.

<sup>30</sup> Tynan, Dan, “Acxiom Exposed: A Peek Inside One of the World’s Largest Data Brokers”, *IT World*, May 15, 2013, <http://>

узнать о том, сколько в данной местности имеется домов с бассейнами, можно было из кадастровой оценки недвижимости. Во времена, когда информации о потребителях было немного, маркетинговое сегментирование стало неслыханным прорывом. К началу нового тысячелетия годовая выручка Асxiом достигла почти миллиарда долларов<sup>31</sup>.

Желание этих брокеров распространить свою аналитику и на интернет-торговлю было вполне естественным. За год до моего прихода в Amazon я работал с командой специалистов Асxiом над возможностью включения цифрового компонента в их базы данных, основанные на почтовых индексах и адресах домохозяйств. Менеджеры Асxiом пытались найти способ привязки нужного адреса электронной почты к уже имеющимся в базе данным о домохозяйстве. И пока Асxiом рассматривала возможность совершения одного небольшого шага, Amazon и остальные были уже на старте гигантского рывка к изобилию социальных данных. Я очень хорошо помню, как за шесть лет до появления первого айфона пытался объяснить менеджерам, что данные из онлайн в скором будущем позволят компаниям знать о домохозяйствах значительно больше. Торговля получит возможность отслеживать каждый поисковый запрос, каждый клик и каждую покупку, обращать внимание на каждую недооформленную «корзину покупок». Имея в своем распоряжении такой объем информации, компании смогут по-настоящему индивидуализировать маркетинг своих товаров и услуг, то есть ориентировать его на сегмент, состоящий из одного человека<sup>32</sup>.

Из-за стремления торговать всем, что угодно, Amazon иногда называют «магазином всего», но, учитывая, насколько тщательно компания сохраняет каждый бит информации о своих клиентах и товарах, более правильным было бы называть ее «магазином, запоминающим все»<sup>33</sup>. В ассортименте предложений Amazon – сотни миллионов наименований, и поэтому она не может показать все, что в него входит. Пролистать весь ассортимент компании не получится в силу его масштаба. Компания не сможет показать вам что-то подходящее, пока вы не скажете ей, что именно вы ищете. Для того чтобы получить ранжированные результаты поиска, вам придется поделиться информацией. Варианта сохранить области своего интереса втайне от продавца у вас нет.

В 2002 году, когда я начал работать в Amazon, в числе прочих мы решали задачу перехода от анализа на уровне почтовых индексов к максимальному использованию всей информации о взаимодействии посетителей с сайтом. В итоге мы с командой определили пятьсот существенных признаков для каждого пользователя. А начиналась эта работа с того, что мы задались целым рядом вопросов, например: влияет ли расстояние между адресом доставки и ближайшим книжным магазином на то, как часто данный покупатель делает заказы в Amazon или на стоимость заказа? Можно ли прогнозировать покупательское поведение на основе типа его кредитной карты? Кто оставляет в Amazon больше денег за год – покупатели, делающие заказы в нескольких категориях, или те, кто заказывает только книги? Отличаются ли заказы какого-то конкретного покупателя, сделанные в утреннее время, от тех, которые он делает вечером? Результаты нашего анализа ложились в основу многих решений компании, например при выборе между затратами на рекламу и снижением цен на товары.

---

[www.itworld.com/article/2710610/it-management/acxiom-exposed – a-peek-inside-one-of-the-world-s-largest-data-brokers.html](http://www.itworld.com/article/2710610/it-management/acxiom-exposed-a-peek-inside-one-of-the-world-s-largest-data-brokers.html).

<sup>31</sup> Асxiом Corporation Annual Report 2000, June 26, 2000, p. 3, <http://www.getfilings.com/o0000733269-00-000012.html>.

<sup>32</sup> Меньше, чем через десять лет после того, как Дон Пепперс и Марта Роджерс предложили радикально новый подход к маркетингу в своей книге *The One to One Future*. См. Peppers, Don, and Martha Rogers, *The One to One Future: Building Relationships One Customer at a Time* (New York: Doubleday, 1993).

<sup>33</sup> Заголовок статьи репортера Bloomberg Businessweek Брэда Стоуна получил широкое распространение в качестве обозначения бизнес-идеи Джеффа Безоса. См. Stone, Brad, *The Everything Store* (New York: Little, Brown, 2013), p. 13. (Русское издание – Стоун, Брэд “*The Everything Store*. Джефф Безос и эра Amazon”. Азбука Аттикус, 2014). Тем не менее, «амазонцы» считают свою компанию в первую очередь аналитической компанией, и, кстати говоря, идея ее создания осенила Джеффа, когда он работал в хедж-фонде D. E. Shaw & Company, совершившем революцию в спекулятивных биржевых операциях за счет использования все новых и новых источников данных.



Этот анализ был полезен и для определения необходимого объема информации, которую посетитель должен предоставлять в ходе выбора покупок. Мы обнаружили, что на основе истории прошлых покупок можно с большей точностью прогнозировать вероятность покупок похожих видов продукции, а не одного конкретного продукта. Ассортиментные позиции связаны друг с другом по-разному, и просчитать эти связи можно различными способами. Судить о схожести товаров можно на основе сравнения спецификаций или анализа совпадений слов в описаниях, но самой важной информацией оказалось то, насколько часто две данные позиции просматривают или покупают вместе. Если можно было выявить тенденцию просмотра покупателями двух схожих позиций в течение одной сессии, их помечали как взаимозаменяемые. Когда покупатель рассматривал какую-то товарную позицию, ему предлагалось посмотреть на варианты ее заменителей («Какие другие товары покупают после просмотра этого?») и дополнений («С этим товаром часто покупают также»). Это делалось на основе анализа данных о прошлых запросах, просмотрах и покупках. Не менее полезными были и общие выводы о процессе принятия решений, которые можно было делать исходя из процентного соотношения просмотров и покупок конкретного товара.

Таким образом, система рекомендаций Amazon строилась на агрегированных данных просмотров и покупок. Кроме того, была создана платформа, позволяющая сторонним компаниям продавать свою продукцию на сайте с использованием складских мощностей Amazon, а это еще больше расширило область анализируемых данных. В отличие от рассылочных фирм с их «Пригородными насадками», «Селянами с дробовиками» и десятками прочих сегментов, Amazon могла прицельно обслуживать меняющиеся интересы и потребности каждого пользователя<sup>34</sup>.

Само по себе сохранение информации не являлось чем-то революционным. Отличительной особенностью Amazon стало то, что компания обрабатывала информацию таким образом, чтобы помочь покупателю определиться с покупкой исходя из его же собственных интересов, вкусов и конкретной ситуации. Но излишняя персонализация может и отпугнуть потребителя. Журналист из «Нью-Йорк таймс» Чарлз Дахигг приводит отличный пример: на основе истории покупок одной молодой девушки сеть магазинов Target выслала на ее домашний адрес рекламное предложение продукции для будущих мам. Ее отец пришел в ярость, однако спустя пару дней дочь сообщила ему, что беременна. Алгоритмы Target не ошиблись<sup>35</sup>.

---

<sup>34</sup> Такой подход к персонализации на основе каждого конкретного случая с сегментацией на уровне одной десятой родился в разговоре с моим частым сообщником в деле социальных данных Гэмом Диасом – руководителем и основателем MoData.

<sup>35</sup> Duhigg, Charles, “How Companies Learn Your Secrets”, New York Times Magazine, February 16, 2012, <http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html>.

**ЕДИНСТВЕННОЕ,  
НА ЧТО НЕ СПОСОБНЫ  
НОВЫЕ ТЕХНОЛОГИИ, —  
РЕШИТЬ, КАКОЕ БУДУЩЕЕ  
МЫ ХОТИМ ДЛЯ СЕБЯ  
И КАК ОТДЕЛЬНО ВЗЯТЫЕ  
ЛИЧНОСТИ, И КАК  
ОБЩЕСТВО В ЦЕЛОМ**

---

Amazon изменила принципы маркетинга, начав использовать всю информацию, созданную в процессе взаимодействия пользователей с сайтом. Кроме того, у покупателей появилась возможность создавать информацию в виде отзывов о товарах. Этот эксперимент полностью перевернул традиционные представления о маркетинге с его стремлением к полному контролю над брендовыми коммуникациями. Покупатели охотно делились своим опытом и зачастую в большей степени полагались на отзывы других потребителей, чем на описания производителя или рекламу продавца. Если товар получал множество низких оценок пользователей, то положительные отзывы о нем экспертов или сотрудников уже не имели большого значения. Кроме того, публикация отзывов позволила значительно расширить представление потребителей об ассортименте «МагазинаВсего» и предоставила им возможность знакомства со всем спектром мнений. Со временем Amazon вообще отказалась от редакционного персонала и перенаправила ресурсы на разработку алгоритмов, позволяющих показывать наиболее полезные отзывы пользователей на самом видном месте страницы товара. Средства, затраченные на технологии обработки информации, помогли улучшить обслуживание покупателей в большей степени, чем затраты на отбор и рецензирование ассортимента.

Методика обработки данных в Amazon изменила поведение миллиарда покупателей. В 2015 году почти половина покупок в США начиналась с поиска и просмотра товара на Amazon, вне зависимости от того, где в конечном счете приобретался товар<sup>36</sup>.

---

<sup>36</sup> Доля предварительного просмотра товаров в Amazon перед покупкой в рознице возросла с 30 процентов в 2012 году

Для того чтобы управлять автомобилем, необязательно понимать все тонкости устройства двигателя внутреннего сгорания. Точно так же нет никакой необходимости досконально разбираться в алгоритмах Amazon, для того чтобы найти то, что нужно или представляет интерес. Важнее понимать базовые принципы устройства системы и установить правила ее безопасного использования. По мере того как информации создается все больше и она передается все более широкому кругу пользователей, мы можем либо занять пассивную позицию и предоставить право определять правила ее использования другим (беззаботно нажимая кнопку «соглашаюсь» после просмотра по диагонали двадцати с лишним страниц текста), либо принять активное участие в установлении новых норм взаимодействия. Можно относиться к переработке социальных данных как к таинственному «черному ящику», а можно стать информационно грамотными людьми, находящими разумные способы заставить тех, кто получает и обрабатывает наши данные, давать взамен не меньшие ценности.

### **Чего стоят ваши данные?**

Уже сейчас при решении многих бытовых вопросов мы полагаемся на социальные данные – будь то выбор покупки на Amazon или места, где поужинать, и способа добраться до него. Социальные данные создаются в очень многих областях жизни, и постепенно мы попадаем во все большую зависимость от инфопереработчиков при принятии важнейших жизненных решений, в том числе выбора партнера для романтических отношений, места и условий работы, медикаментов и учебных заведений.

Во многих случаях истинное значение созданных нами данных становится понятным только в сравнении их с данными, созданными другими людьми. Поскольку объем социальных данных, доступных для инфопереработки, нарастает в геометрической прогрессии, сегодня можно надеяться на получение ответов на многие из вопросов, которые считались прежде неразрешимыми в принципе. Возможно даже, что это заставит задать ряд новых важных вопросов, которые раньше просто не приходили в голову.

Алгоритмы выявляют закономерности, которые люди не могут увидеть без помощи компьютеров. Эти закономерности могут помогать нам в принятии решений. Стоимость информации, предоставленной для переработки, определяется тем, насколько полезными окажутся полученные результаты для принятия нами решений – в коммерческих сделках, в приобретении товаров и услуг, в получении банковского кредита, в поиске работы, в получении медицинской помощи и образовательных услуг для себя и своих близких, а также в общественно-политической жизни.

Оценка степени полезности результатов деятельности компаний, работающих с данными, – это существенно иная постановка вопроса, нежели привычные рассуждения о том, как, когда и почему компании и государство собирают наш «цифровой выхлоп», то есть информацию, которую мы создаем изо дня в день. Некоторые считают, что объем собираемой информации слишком велик и что лучшим вариантом поведения для частного лица является поменьше рассказывать о себе или же требовать плату за создаваемые и предоставляемые личные данные. При такой сосредоточенности на входящем потоке информации мы упускаем из виду потенциальные выгоды, которые можно извлекать на выходе. Я считаю, что мы вправе требовать нечто намного более ценное, чем мелкая денежная подачка, за предоставление своих первич-

---

до примерно 50 в 2015-м. См. Ludwig, Sean, “Forrester: 30 % of Online Shoppers Research Amazon Before Buying”, VentureBeat, July 26, 2012, <http://venturebeat.com/2012/07/26/amazon-online-shoppers-research>; Mulpuru, Sucharita, and Brian K. Walker, “Why Amazon Matters Now More Than Ever”, Forrester Research, July 26, 2012, <https://www.forrester.com/Amazon/fulltext/-/E-RES76262>; и Cassidy, Mike, “Survey: Amazon Is Burying the Competition in Search”, BloomReach, October 6, 2015, <http://bloomreach.com/2015/10/survey-amazon-is-burying-the-competiton-in-search>.

ных данных. Мы должны настаивать на участии в управлении инфопереработкой на справедливых и понятных условиях, чтобы иметь возможность влиять на ее результаты.

Для начала давайте рассмотрим различия между первичной и переработанной информацией. Когда я ввожу в поисковую строку «Андреас Вайгенд», Google докладывает, что эти два слова обнаружены в «примерно в 122 000 результатов». Отсмотреть все эти страницы вручную невозможно: если на каждую тратить хотя бы пять секунд (феноменально высокая скорость просмотра), на это потребуется целая неделя, что совершенно нереально. Поэтому остается положиться на порядок выдачи результатов. Во главу списка Google может поставить самые свежие упоминания. Это будет прекрасно в случае, если я интересуюсь последними новостями о себе, но не так здорово, если я ищу видео семинара, который вел несколько лет назад. Другой вариант – посчитать, сколько раз мое имя упоминается на страничке, и ранжировать релевантность результатов в соответствии с наибольшим количеством таких упоминаний. Это может оказаться полезным, если я просматриваю статьи и хочу найти ту, где меня цитируют чаще, чем в остальных. Но представьте себе, что вместо своего имени я набрал в поисковой строке «айпад по дешевке» – это будет примерно 350 000 результатов, и над полезностью такой выдачи стоит задуматься. Специалисты по рекламным ловушкам наверняка поработали над страничками с популярными поисковыми запросами (что действительно так), и мне придется долго блуждать от результата к результату в поисках действительно полезной информации.

Чтобы повысить эффективность поиска, Google рассматривает полезность страницы с учетом многих характеристик, а не только наличия слов из запроса. Разработчики компании начинали с ранжирования релевантности страниц по принципу количества ссылок на них в других местах, что давало возможность судить об уровне внимания аудитории. Когда народ понял важность входящих ссылок для места странички в выдаче поисковика, появилась сфера деятельности под названием «поисковая оптимизация» с одиозными «фермами ссылок», или линкопомойками. Алгоритмы Google пришлось усовершенствовать, чтобы они могли отличать входящие ссылки реальных заинтересованных пользователей от созданных по заказу владельца сайта. Сегодня у Google помимо структуры ссылок сети есть накопленные за два десятилетия данные о том, на какие сайты, предложенные по поисковому запросу, заходили люди и сколько времени проводили на них, прежде чем вернуться к странице результатов поиска. Если на сайт заходят многие, но, лишь бегло взглянув на него, уходят искать что-то более интересное, релевантность страницы в поиске Google падает, и она опускается ниже в результатах поиска. Тем не менее высокое место странички в результатах поиска в Google не гарантирует достоверность представленной на ней информации, а является лишь свидетельством проявляемого к ней внимания.

Сколько поисковых запросов проходит через Google ежедневно? Сколько фотографий размещается в Facebook? Умение различать достоверные, недостоверные и невероятные данные – один из базовых навыков информационной грамотности. Точные цифры не столь важны: информационная грамотность означает умение видеть разницу между чем-то вполне приемлемым и явной ошибкой на порядок. В подобных оценках физики часто рассуждают с позиций порядковых величин, то есть десятикратных различий. Они скажут, что количество пользователей Google или Facebook составляет порядка миллиарда человек, поскольку оно точно больше 100 миллионов и меньше 10 миллиардов<sup>37</sup>. Далее они сделают допущение о том, что типичный пользователь делает в среднем 10 поисковых запросов в день, поскольку их точно больше одного, но меньше 100. При оценке количества фотографий в Facebook они будут исходить из показателя одно фото на пользователя в день, поскольку их точно больше, чем одно

---

<sup>37</sup> В 2015 году Facebook сообщила о 1,59 миллиарде посещений в месяц при среднем количестве ежедневных посетителей в 1.04 миллиарда. См. Mike, “Facebook Reports Soaring Revenue, Buoyed by Mobile Ads”, New York Times, January 27, 2016, <http://www.nytimes.com/2016/01/28/technology/facebook-earnings-zuckerberg.html>.

в месяц и меньше 10 в день. Таким образом, мы получаем порядковые оценки ежедневного количества поисковых запросов и размещаемых фотографий – 10 миллиардов и 1 миллиард соответственно. И это только два вида операций в области социальных данных<sup>38</sup>.

Осознав, что социальные данные ежедневно создаются во многих миллиардах других случаев, вы начинаете понимать, что ваши собственные первичные данные не имеют какой-то особой ценности в материальном смысле. Умилительное фото вашей собачки, которое вы запустили в Facebook, заинтересует от силы сотню человек, или 0,00001 процента пользователей сайта. Практически полезные закономерности и взаимосвязи можно выявить, только собрав и проанализировав данные нескольких миллионов человек. Отсутствие в их числе данных какого-то одного человека не повлияет на выводы, сделанные в результате переработки остального массива информации. Картина инфопереработчиков не исказится из-за пропуска данных одного человека из миллиарда.

Более того, входящая информация не всегда бывает столь же дискретной, как размещенное в Facebook фото. Отдельно взятый элемент данных похож на камушек или даже песчинку в океане – он обладает индивидуальными чертами и его трудно найти. Или же он может напоминать каплю чернил, которая растворяется в воде до состояния полной неотделимости. Информационная грамотность подразумевает также понимание того, как может быть удалена ваша информация: нужно ли для этого совершать конкретное действие, или же она сама собой растворится в общей массе данных всех пользователей. Выше я писал о том, что в Amazon просмотр товара увязывается с просмотром другого товара или покупкой. Клиент может удалить факт любой покупки из своей истории заказов, если не хочет, чтобы она там фигурировала. Но удалить факт просмотров из системы рекомендаций Amazon невозможно, поскольку они не привязаны к конкретному пользователю. В этом случае можно вновь провести параллель с нефтепереработкой: на определенном этапе становится невозможным выделить нефть, полученную на какой-то конкретной скважине.

Такое понимание соотношения количества и качества данных отчасти – но не полностью – обосновывает мою убежденность в том, что требовать плату за предоставление личной информации было бы ошибочным. Одним из главных пропагандистов идеи платного предоставления данных является концептолог Microsoft Research Джарон Ланье. Он страстно отстаивает эту точку зрения с момента публикации своей книги «Кому принадлежит будущее?» в 2013 году<sup>39,40</sup>. Один из его любимых примеров – работа сервиса «Google-переводчик». Почему, задается вопросом автор, все доходы от рекламы получает Google, а всем тем, кто помогает совершенствовать алгоритмы компании, исправляя и дополняя варианты переводов, не достается ничего? Каждое исправление и дополнение, предлагаемое пользователями, улучшает систему переводов Google, даже если они являются повторами. Система как раз обращает особое внимание на неоднократно предлагаемые варианты.

Помощники Ланье получают компенсацию за свою работу. Весьма вероятно также и то, что они извлекают выгоду и от использования «Google-переводчик». Но в последнем случае это не деньги, а доступ к постоянно совершенствующимся информационным продуктам и сервисам.

Теперь давайте посмотрим, какая информация создается в Facebook. Автором информации, которую представляет собой фото вашей собаки, совершенно точно являетесь вы. А как быть в случае группового фото с вечеринки по случаю дня рождения? Вы сделали и разместили эту фотографию, но коммерческая ценность этого поста в Facebook определяется тра-

---

<sup>38</sup> У Google нет регулярной отчетности о количестве поисков в ее сервисах. Тем не менее в ее отчете “Zeitgeist” за 2012 год говорилось о 3,3 миллиардах поисковых запросов ежедневно. См. <https://www.google.com/zeitgeist/2012/#the-world>.

<sup>39</sup> Who Owns the Future? В России книга не издавалась. – *Ред.*

<sup>40</sup> Lanier, Jaron, Who Owns the Future? (New York: Simon & Schuster, 2013), pp. 273–274.

фиком, который он создает, и уточнением данных о связях и интересах людей. Должны ли вы получить все средства, которые можно отнести на счет этого поста? Или вы должны разделить их со всеми, кто отмечен на этом фото? А как быть с комментариями, лайками и тэгами? Ведь они означают, что фото пробудило активность многих других людей, которая в свою очередь транслируется уже их друзьям. И эта информация о «цепочке» может быть намного более полезна с точки зрения возможностей обработки данных и извлечения из этого доходов. Ланье не обсуждает подобного рода связи, возможно, считая их недостаточно «креативным» контентом, за который стоит платить. Но такого рода цифровые следы составляют основную часть сырья инфопереработчиков, на результаты работы которых мы полагаемся в своей повседневной жизни.

Если заставить инфопереработчиков материально оценить вашу активность – все ваши поиски, просмотры, лайки и тэги во взаимосвязях со всеми, кто использует эти данные и дополняет их, можно с уверенностью предположить, что это они потребуют с вас плату за доступ к результатам поиска, рекомендациям и рейтингам. Разработка алгоритмов стоит денег, и подобный аналитический срез потребует создания специального инструментария для атрибуции и оценки каждого элемента данных, причем с учетом изменения его стоимости во времени.

Предложенная Ланье «микроплата» за предоставление данных – безнадёжное мероприятие не только в силу сложности и затратности решения проблемы атрибуции. Для начала давайте просто посмотрим на порядок цифр. Если бы Facebook решила поделиться всей своей прибылью за 2015 год, составившей 3,5 миллиарда долларов<sup>41</sup>, со своими пользователями (не выплачивая дивиденды акционерам), каждый из них получил бы примерно по 3 доллара 50 центов. Что для вас более важно – постоянно иметь в своем распоряжении платформу с неограниченными коммуникационными возможностями или одна лишняя чашка капучино в год? Если первое, то за ваши данные вам уже «заплатили».

Разбираемся дальше. Во многих случаях вам надо предоставить данные, чтобы получить услугу, например информацию о своем местоположении для приложения Uber. Решив, что впредь вы отказываетесь делиться своей информацией бесплатно, вы тем самым лишите себя возможности пользоваться бесплатными продуктами и услугами многих информационно-технологических компаний. Наконец, многие из продуктов этих компаний – от товарных рекомендаций до прогноза спроса на услуги такси – основаны исключительно на обработке первичных данных пользователей. И хотя конкретно ваши данные могут не иметь какого-то особого значения для конечного результата, сама просьба к потребителям продуктов и услуг предоставлять свою информацию является вполне уместной.

По этим причинам я полагаю, что, вместо того, чтобы настаивать на плате за свои первичные данные, следует потребовать предоставления более жесткого контроля над тем, как, когда и почему распространяется ваша информация, для чего она используется и что вы получаете в результате. Наиболее успешные инфопереработчики разъясняют, каким образом предоставляемые вами данные способствуют совершенствованию предлагаемых ими информационных продуктов. В обществе слишком много обсуждают ограничения, которые следует наложить на использование первичных данных организациями, но практически игнорируют вопрос о том, что инфопереработчики должны предлагать в качестве инструментов повышения прозрачности и расширения свободы выбора.

Переработка данных не превращает человека в товар в виде набора цифр, во всяком случае, это далеко не обязательно. Главное, что может почерпнуть из этой книги читатель, – это вывод о пользе обработки социальных данных для *его собственных* решений, а не только для рекламной кампании какой-нибудь мегакорпорации. Я считаю, что человека в равной степени

---

<sup>41</sup> Facebook Annual Report 2015, January 28, 2016, <https://investor.fb.com/financials/default.aspx>.

определяют и создаваемая им информация, и принимаемые им решения. И ценность ваших данных для вас состоит именно в этом.

## Исследование и использование

Процесс переработки данных подразумевает также компромисс между исследованием и использованием. Хочу перенести вас к рядам игровых автоматов в залитом неоновыми огнями Лас-Вегасе. В области искусственного интеллекта (то есть компьютерных программ, способных самообучаться на основе входящей информации) задача об «одноруком бандите» является в каком-то смысле коронной, примером дилеммы между поиском новых вариантов или использованием лучшего из известных<sup>42</sup>. Предположим, вы зашли в казино и узнали, что кто-то вроде бы выиграл целое состояние на одном из автоматов. Как вы поступите? Проведете остаток вечера у этого прославившегося автомата или же будете исследовать остальные в поисках еще более высоких шансов на джекпот? Разумеется, сбор данных о результативности всех автоматов потребует времени. Поскольку казино создаются ради прибыли, игры настроены так, чтобы клиенты в целом оставались в проигрыше. И в идеале, как учат теоретики вычислительных систем, вам придется в течение какого-то времени понаблюдать за работой автоматов, чтобы постараться обнаружить закономерность. И хотя статистик сможет порекомендовать, сколько времени нужно потратить на наблюдение за каждым из этих шумных устройств, у вас все равно останется выбор – или исследовать новые варианты, или использовать тот, который уже доказал свою результативность. Может показаться, что пример об «одноруком бандите» имеет мало общего с конечной продукцией инфопереработки, однако в списке рекомендаций для пользователей в качестве ключевой задачи значится как раз поддержание баланса между исследованием и использованием. Из этих соображений пользователь выбирает наиболее подходящую ему рекомендацию. И здесь аналогия с нефтью снова придется кстати. Геологи-нефтяники и инженеры оценивают альтернативу: стоит ли и дальше вкладывать значительные средства до полной выработки существующего месторождения или же переключиться на поиски новых, где добыча может обходиться не так дорого. Инфопереработчикам также приходится решать проблемы лучшего применения ресурсов для максимизации эффективности входящих и исходящих потоков данных. В том, что касается информации, главным параметром, который надлежит контролировать, является время пользователей.

Когда поисковик вроде Google выдает ответ на ваш запрос, перечень сайтов представляет не десятки одинаковых позиций, а предлагает различные варианты в рамках определенной степени релевантности предмету поиска. Иногда бывает понятно, что вам нужна информация о чем-то совершенно конкретном, например в случае, когда вы вводите в строку поиска словосочетание «Panthera onca». Но если вы введете в качестве предмета поиска просто «ягуар», то компьютер покажет вам не только то, что относится к представителю кошачьих, автомобилю или старой операционной системе компьютеров Mac<sup>43</sup>. Алгоритмы поисковой системы создают кластеры значений слова «ягуар» на основе количества слов на странице, связей между страницами и перемещениями пользователей между страницами и предлагают для изучения выборки из каждого такого кластера, чтобы обеспечить большую вероятность успеха вашего поиска.

Разновидностью задачи об «одноруком бандите» является задача «оптимального момента остановки», или «взыскательного ухажера». Впервые ее описал Мартин Гарднер в своей колонке «Математические игры» в журнале *Scientific American*. В его варианте на лист-

---

<sup>42</sup> Gittins, J. C., “Bandit Processes and Dynamic Allocation Indices”, *Journal of the Royal Statistical Society B (Methodological)* 41, no. 2 (1979), pp. 148–177, <http://www.jstor.org/stable/2985029>.

<sup>43</sup> Спасибо одному из заслуженных разработчиков Microsoft Яну О. Педерсену за его рассказ о том, как решалась проблема исследования/использования в Yahoo! (где он прежде работал главным научным сотрудником поисковой системы) на примере поискового запроса «ягуар».



ках бумаги пишутся любые числа, «от мельчайших долей единицы до астрономически больших, вроде единицы со ста нулями»<sup>44</sup>. Листочки перетасовываются, а затем перебираются один за другим до тех пор, пока вы не останавливаетесь на том, где, как вы считаете, значится самое большое число. Со временем листочки бумаги в мысленном эксперименте превратились в уха-жеров, идущих на свидание. Вы приходите на свидание, и вам надо решить: будете ли вы встречаться с остальными девушками или прекратите поиск, потому что *это та самая?* В реальной жизни выбор между исследованием и использованием может иметь критически важное значение.

Понятно, что пользователи приложений или сайтов знакомств решают задачу «взыскательного ухажера» в постоянном режиме. Самые первые сайты знакомств позволяли пользователям уточнять свои предпочтения по весу, росту или географическому положению потенциального партнера и соответственно этому ранжировали результаты поиска. Некий пользователь решил кликнуть фото потенциальной партнерши, которую мы назовем Сэм. Сайт не знал, что именно заставило его кликнуть фото Сэм. Может быть, то, что она была первой в списке? А может быть, то, что она брюнетка и носит очки? А может быть, это произошло потому, что она сфотографирована на фоне океана, а этого пользователя интересуют жительницы побережья или те, кто приезжает туда отдохнуть? Пользователя могло заинтересовать все, что угодно, но ему все равно нужно было решать – послать Сэм сообщение или продолжать разглядывать фото. В отличие от традиционной свахи, которая всеми силами стремится найти идеальную пару для каждого клиента, сайт знакомств предоставляет пользователю самому решать, хочет ли он увеличить число рекомендаций от сайта, видеть в подборке вариантов нечто схожее с предыдущими или нечто совершенно другое.

В основном инфопереработчики решают вопрос баланса между исследованием и использованием исходя из того, насколько подробно пользователь изучает рекомендации и возвращается ли он к ним, если возвращается вообще. Однако оптимальная настройка часто зависит от предпочтений пользователя в данный конкретный момент. Взыскательный ухажер может искать как вечный идеал, так и партнера на ближайший вечер, и инфопереработчику трудно определить характер конкретного поиска.

Принцип прозрачности подразумевает, что пользователи понимают, как обрабатываются их данные; свобода выбора диктует право пользователей на определенное влияние на этот процесс.

Стартап музыкальных рекомендаций MoodLogic, сооснователем которого я являлся<sup>45</sup>, предлагал своему пользователю определенную степень контроля над сочетанием использования и исследований, то есть между музыкой, которую он обычно слушает, и знакомством с чем-то новым. Мы анализировали коллекцию музыки в компьютере пользователя и создавали некую модель, позволяющую определить произведения, исполнителей, авторов, инструментальные составы, темпы и жанры, которые ему подойдут. Модель прогнозировала, насколько новое произведение может понравиться пользователю. Затем мы предлагали ему выбор из двух настроек. При выборе «безрискового» варианта система выдавала музыку примерно одного стиля, которая должна была, по нашему прогнозу, понравиться пользователю. «Зондирующая» настройка предлагала послушать музыку, которую, как мы считали, пользователь мог бы или полюбить, или возненавидеть. Выбор оставался за ним, но при этом аккумулировалась информация, которую мы могли использовать для совершенствования алгоритмов MoodLogic.

---

<sup>44</sup> Гарднер рассмотрел так называемую «задачу секретаря» в журнале Scientific American за февраль/март 1960 года. См. Gardner, Martin, Martin Gardner's New Mathematical Diversions (New York: Simon & Schuster, 1966), p. 35.

<sup>45</sup> Founded in 1998 by myself, Christian Pirkner, Elion Chin, and Tom Sulzer, MoodLogic, которую в 1998 году создали Кристиан Пиркнео, Элион Чин, Том Сулзер и я, была одной из первых систем музыкальных рекомендаций. На пике популярности на сайте присутствовали оценки более миллиона треков от примерно 50 000 пользователей. В 2006 году программное обеспечение и информационный массив компании были куплены All Media Guide – компанией группы.

Хотя количество информации безгранично, этого нельзя сказать о времени. Решения приходится принимать. Феномен социальных данных в том, что результаты процесса их переработки могут становиться новым входящим потоком.

### Работа над ошибками

Людам нравится считать свои решения обоснованными. Возможность перечислить все «за» и «против» («Стоит ли принять предложение работы в другом городе или согласиться с конкурентным предложением моего нынешнего работодателя?»), сравнить варианты и выбрать то, что лучше соответствует ситуации, текущим целям и кажется наименее рискованным, придает уверенности. В прошлом люди собирали информацию, разговаривая с родными, друзьями, коллегами и наставниками. Они принимали решения в мире «небольших данных».

Сегодня можно обратиться к рейтингам удовлетворенности работой портала Glassdoor, где на условиях анонимности оценивают условия и оплату труда<sup>46</sup>. Там собраны отзывы сотрудников о более чем 400 000 компаний и ежегодно поступает более полумиллиона новых комментариев. Например, по компании Amazon собрано 8000 отзывов о работе, 8000 – о собеседованиях при приеме на работу и 14 000 – о зарплате. Человек, рассматривающий возможность работы в компании, получает доступ к значительно большему количеству информации о ней, чем когда-либо прежде, однако времени на то, чтобы ознакомиться со всеми 8000 отзывов и сравнить их с условиями своего нынешнего места работы, у него нет. Какие из этих отзывов достоверны, какие соответствуют рассматриваемой позиции? А если кто-то неверно понял вопрос или случайно кликнул более низкую оценку, чем хотел?

Ошибки свойственны любой информации. Во времена небольших данных люди, собиравшие информацию, считали своим долгом досконально разобраться в ней и лично вычистить и исправить ошибки. И это было очень хорошо, поскольку от информации, полученной от небольшого количества людей, иногда зависели решения, затрагивающие всю общину или даже целый штат. Ошибка в количестве заявок на пособие по безработице, полученных за неделю в каком-то штате (например, опечатка вроде «254» вместо «2541»), могла привести к искажению данных по безработице, которые в свою очередь влияют на государственную экономическую политику. Долгосрочное исследование тенденций занятости, которое проводит Статистическое управление министерства труда США, охватывает выборку примерно в 10 000 человек, то есть основывается на порядке цифр, сопоставимых с количеством отзывов сотрудников Amazon на Glassdoor<sup>47</sup>.

Разумно предположить, что доля ошибочной информации не снижается пропорционально росту количества собранных данных. Если в наши дни мы имеем доступ к объему информации, в сто раз превышающему прежний, можно ожидать, что и количество ошибок в этом потоке увеличится в сто раз. Но теперь у нас нет возможности отслеживать и вычищать каждую ошибку в массиве информации.

Однако решение проблемы экспоненциального роста количества ошибочной информации содержится в самом факте экспоненциального роста объемов данных. Поскольку люди постоянно реагируют на продукцию инфопереработки созданием новых данных, алгоритмы

---

<sup>46</sup> Glassdoor создали в 2008 году Рич Бартон, Тим Бессе и Роберт Хоман, до этого работавшие в Expedia – портале бронирования туристических и гостиничных услуг. Бартон был также в числе создателей Zillow – инфопереработчика в сфере недвижимости.

<sup>47</sup> Общенациональный лонгитюдный опрос молодежи 1979 года, охватывающий американцев 1957–1964 года рождения, начинался с выборки в 11 000 гражданских лиц; общенациональный лонгитюдный опрос молодежи 1997 года, охватывающий американцев 1980–1984 года рождения, начинался с выборки в 8000 гражданских лиц. Более подробно – на <https://www.nlsinfo.org>.

могут учиться выявлять то, что может являться ошибкой ввода. Если в строке поиска ввести «Андреас Вайгенд», Google поинтересуется, не имелся ли в виду «Андреас Вайгенд».

Объединяя данные из многих источников, инфопереработка может выявлять наши ошибки во входящей информации. В июле 2012 года в моем смартфоне появился сервис под названием Google Now. Он сканировал мою электронную почту в поисках информации из моих электронных авиабилетов и сообщал о статусе рейсов, делая это даже раньше авиакомпаний. Казалось бы, все просто. Но этому сервису все же удалось удивить меня степенью продвинутости анализа данных. Как-то утром, когда я только собирался паковать свои чемоданы перед отъездом из Фрайбурга, приложение сообщило, что я должен выезжать в аэропорт немедленно. По моему графику до рейса оставалось еще несколько часов, а авиакомпании обычно не сдвигают регулярные перелеты вперед больше чем на пару минут. Тем не менее я доверял Google Now больше, чем своему календарю, и решил пошевелиться – возможно, сервис узнал об огромной пробке на дороге. Приехав в аэропорт, я понял, что неправильно ввел время рейса в календарь. Google Now проигнорировал данные, введенные вручную, и послал мне напоминание исходя из информации в моем электронном билете в Gmail. (А спустя три года Google Now автоматически вносит расписание рейсов в мой календарь сразу же после того, как электронные билеты попадают в мою почту.)

Мы уже свыклись с тем, что инфопереработчики указывают нам на подобные ошибки и исправляют их. Это полезные услуги. Вопрос в том, будем ли мы готовы принимать схожие корректировки в других областях своей жизни по мере того, как создаем и распространяем все больше личной информации.

Инфопереработке приходится также разбираться и в том, что является сигналом, а что шумом. На статистическом жаргоне сигналом называется значимая информация, а шумом – случайная и потому не имеющая значения. Сложность социальных данных в том, что различия между сигналом и шумом варьируются в зависимости от конкретного пользователя и конкретной ситуации. Когда ваш френд из Facebook ставит тэг на фото, где вы отсутствуете, что это – сигнал или шум? Когда как. Если он поставил тэг по ошибке, перепутав вас с Эндрю, который идет следующим в списке его друзей, то это шум, то есть статистический эквивалент статическим помехам в вашем радиоприемнике. А если он поставил тэг сознательно, желая ввести в курс запечатленных событий вас и ваших френдов, то это сигнал, хотя, возможно, и раздражающий. То есть, на языке статистиков, это не шум.

Обратная связь от пользователей играет важнейшую роль в совершенствовании алгоритмов обработки информации. Я не имею в виду необходимость заполнения опросников для потребителей или участия в фокус-группах. Поддержание и развитие постоянного диалога с пользователями позволяет инфообработчикам совершенствовать свои продукты и услуги и делать их более персонифицированными. Каждый сделанный вами выбор позволяет скорректировать иерархию вариантов. Но и вы сами тоже учитесь изменять формулировки своих поисковых запросов так, чтобы результаты в большей степени соответствовали ожиданиям. Не просто избегать опечаток, а правильно акцентировать свой интерес к различным аспектам темы или товарным категориям.

Тем не менее ваше взаимодействие с сайтом или приложением ограничивается предлагаемым набором опций. Я полагаю, что уточнение поисковых запросов стало бы намного более динамичным процессом, если бы пользователи имели возможность поиграть с вариантами инфопереработки примерно так же, как мы с коллегами делали это в музыкальных рекомендациях MoodLogic. По мере возрастания количества отзывов работников на Glassdoor порталу стоило бы придумать способы уточнения этой информации, чтобы сделать ее еще более полезной. Это может быть некая модель, позволяющая выделить наиболее интересные данному пользователю оценки не только по признакам названия позиции или ее географического положения, но и по другим характеристикам, которые предоставляют сайту (например, карьер-

ные цели или предпочтения по условиям труда). Но вне зависимости от количества входящей информации оценки все равно будут содержать элемент неопределенности.

Информационная грамотность означает понимание того, что любая рекомендация представляет собой величину вероятности и что любое решение есть компромисс между уровнем риска и размером извлекаемой выгоды, причем даже в тех случаях, когда на фоне больших объемов данных неопределенность кажется весьма незначительной. Инфопереработка не должна принимать решения за вас. Она должна предоставить вам возможность использовать намного большее количество информации, что позволит частично устранить риск ошибки.

С помощью инфопереработки мы получаем возможность использовать и анализировать обширные исторические данные, выявлять закономерности и делать прогнозы трендов, не все из которых оказываются правильными. Подобный подход к осмыслению информации и самих себя сильно отличается от привычного для подавляющего большинства людей.

### Превращение информации в решения

Факты! Факты! Факты! – нетерпеливо восклицал Холмс. – Я не могу лепить кирпичи без глины<sup>48</sup>.

*Артур Конан Дойль*

Когда в начале 1990-х годов я окончил докторантуру и стажировался в научно-исследовательском центре Хегох PARC в Пало Альто, мы с помощью суперкомпьютера анализировали закономерности дорожного движения. Одной из наших задач было прогнозирование времени в пути. Будучи физиками, мы рассматривали дорожное движение как поток и пытались определить условия его перехода из ламинарного состояния в турбулентное, то есть когда плавное движение превращается в прерывистое. По современным меркам, исходных данных было немного, и для создания моделей дорожного движения нам приходилось делать множество допущений.

Сейчас проблема расчетного времени прибытия в пункт назначения решается просто: практически в каждой машине есть кто-то с мобильным телефоном, оценивающим движение в режиме реального времени. В этой области работает, в частности, выделившаяся из Microsoft компания Inrix. В ней анализируют данные геолокации более чем 100 миллионов личных телефонов, чтобы определить направления движения автомобилей (и, что еще более важно, места, куда они не едут) и вывести тренды перемещения людей и товаров<sup>49</sup>. Данные для своего анализа Inrix получает от операторов мобильной связи, с базовыми станциями которых связываются эти 100 миллионов мобильных телефонов. Обработанную информацию у Inrix покупают Garmin, MapQuest, Ford, BMW и другие компании, которые хотят предоставлять водителям услуги картографического отображения местности и планирования маршрута. Кроме того, Inrix консультирует местные власти по вопросам городского планирования, в том числе строительству новых мостов, установке светофоров и местоположению новых государственных больниц и других учреждений.

Работа Inrix с данными о дорожном движении – наглядный пример того, насколько агрегированные показатели множества устройств могут быть полезнее для принятия решений, чем

---

<sup>48</sup> Одна из наиболее часто цитируемых фраз Шерлока Холмса. This is one of Sherlock Holmes's most frequently quoted lines of dialogue. См. Артур Конан Дойл, «Приключение в 'Лесных Буках'» – Doyle, Sir Arthur Conan, "The Adventure of the Copper Beeches", Strand Magazine (June 1892).

<sup>49</sup> Inrix – одна из многих коммерческих организаций, анализирующих данные о местоположении мобильных телефонов с целью изучения транспортных потоков. В анализе используются также данные Garmin и других специализированных GPS-устройств навигации и информирования о состоянии дорожного движения.

первичные данные одного-единственного человека<sup>50</sup>. Упреждающие системы, действующие на основе анализа социальных данных, будут консультировать нас по вопросам персональных проблем, финансов, рабочих ситуаций, медицинского обслуживания и во многих других областях, а возможно, и стимулировать к принятию определенных решений.

Кроме того, на этом примере особенно хорошо заметна ключевая роль интерпретации в процессе обработки данных. Обработанные данные могут представляться в трех видах: как описание, прогноз или инструкция. Описание характеризует нечто уже состоявшееся. Прогноз экстраполирует прошлое и настоящее на будущее в предположении, что система не будет подвергаться воздействиям или манипуляциям, способным повлиять на результат. Инструкция рекомендует, как действовать для получения желаемого результата исходя из анализа прошлых событий.

---

**ИНФОРМАЦИЯ**

**НЕ ДОЛЖНА**

**РУКОВОДИТЬ НАМИ.**

**ОНА ДОЛЖНА**

**РАСШИРЯТЬ СПЕКТР**

**НАШИХ ВОЗМОЖНОСТЕЙ**

---

В описательной статистике данные обобщаются, например, в виде кластеров однородных элементов информации. Такие данные могут формировать условия для принятия решений в виде набора критериев для сравнительной оценки конкретной ситуации. Если вам нужно узнать, как в данный момент обстоят дела с пробками на Манхэттене, вы можете посмотреть по данным геолокации с мобильных телефонов, насколько быстро движутся и где встают потоки автомобилей. Но даже в таком относительно простом вопросе присутствует элемент интерпретации. Скорее всего, вы увидите данные о большом скоплении автомобилей в районе небоскреба MetLife. Но, может быть, это следствие того, что MetLife находится рядом с вокзалом Гранд Централ, где всегда полно такси, ожидающих пассажиров, и самих пассажиров, садящихся в такси, и поэтому мобильные телефоны показывают как бы «остановившееся» движение? Если вам нужно узнать, насколько хорошо идет пред рождественская торговля в вашем

---

<sup>50</sup> Мой друг Джон Сквайр, генеральный директор DynamicAction, а в прошлом – главный стратег IBM Smarter Commerce, подчеркивает важность комбинирования данных из разных источников фразой «Ценность – в объединении».

магазине в этом году, вам нужно не только подытожить продажи, но и найти подходящую базу для сравнения. Сопоставление с вашими же данными за аналогичный период прошлого года не будет полностью корректным, поскольку в таком случае не будут учтены изменения, случившиеся в местной экономике. Вместо этого вы можете сравнить результаты своего магазина с результатами похожих магазинов по соседству.

Когда я работал в Amazon, мы исследовали динамику промежутков времени между просмотром товара и его покупкой. Некоторые статистические значения были явно ошибочными – они были отрицательными, а человек физически не может купить товар прежде, чем просмотрит его. Мы не знали, почему происходит такая ошибка, и просто не стали учитывать такие данные. При этом у нас накопилась масса данных, указывающих на то, что многие пользователи выжидают по восемь часов, прежде чем совершить покупку. Очень странно. И только потом мы сообразили, что, поскольку часть компьютеров Amazon настроена в тихоокеанском часовом поясе США, а часть – по Гринвичу, эта разница отражает различие во временных поясах применительно к кликам. Как часто случается, то, что сначала казалось шагом к интересным новым представлениям, в итоге объяснилось обычной ошибкой.

Интерпретация данных – итеративный процесс. Вот один из примеров этого. Одна авиакомпания решила провести целевую рекламную кампанию для потенциальных пассажиров бизнес-класса и поручила группе специалистов по работе с данными выявить владельцев смартфонов, регулярно прибывающих и убывающих через нью-йоркский аэропорт имени Дж. Ф. Кеннеди. Проблема состояла в том, что людьми, регулярно посещающими любой аэропорт, являются вовсе не бизнесмены, а сотрудники авиакомпаний и самого аэропорта. Аналитики убедились в этом по данным с телефонов, показывающим закономерности перемещений их владельцев. Часть постоянных посетителей ежедневно приезжала и уезжала по четкому сменному графику – это были работники аэропорта. Труднее оказалось выделить экипажи самолетов, базирующихся в Нью-Йорке, но и их удалось более или менее точно вычислить по сайтам и приложениям, которыми они пользовались через wi-fi аэропорта: поиск отеля или авторизация в Uber для заказа машины для них были редкостью, а вот в приложения для знакомств они заходили очень часто<sup>51</sup>.

Второй путь осмысления обработанной информации – прогнозный анализ, в котором данные используются для общих выводов относительно будущего, в том числе возможных вариантов поведения и развития событий. Например, градостроители использовали архивные данные о поминутном состоянии дорожного движения, которые собирала компания Inrix, для оценки возможного влияния различных событий (дорожной аварии, нового строительства или массового мероприятия) и подготовки более точных планов действий в особых ситуациях. Хедж-фонды пользовались данными Inrix о транспортных потоках к торговым центрам и пригородным гипермаркетам при принятии решений о покупке или продаже акций задолго до публикации квартальной отчетности этих предприятий. Анализ данных геолокации, собранных в «черную пятницу» 2012 года, позволил точно предсказать резкий скачок продаж на весь пред рождественский период.

Amazon также использует прогнозные модели для принятия деловых решений: например, сколько дополнительного персонала нужно будет нанять на склады и в доставку, чтобы справиться с возрастающим объемом заказов в период рождественских праздников. Это типичная задача из теории принятия решений: как сравнить потери от несвоевременности доставки заказов покупателям с потерями от избыточности мощностей службы доставки? В Amazon нагрузку на транспортные мощности анализируют с очень высокой степенью детализации, ежедневно и в разрезе каждого города. В 2013 году прогнозы компании не оправдались,

---

<sup>51</sup> Из беседы автора с Клаудией Перлич, главным научным сотрудником Dstillery, 25 января 2015 года. Клаудия – моя хорошая знакомая, она была моей студенткой в Колорадском университете в Боулдере и Нью-Йоркском университете.

так же как и прогнозы многих розничных сетей и компаний дистанционной торговли. Многие посылки пришли уже после Рождества, и покупатели были в ярости<sup>52</sup>. После анализа причин Amazon пересмотрела свою модель с целью обеспечения более высокой точности прогнозирования и гибкости в распределении ресурсов. В результате компания смогла предложить своим клиентам гарантированную бесплатную доставку к 24 декабря 2014 года для покупок, совершенных на два дня позже, чем в предыдущие годы<sup>53</sup>.

Поскольку многие из инфопереработчиков занимаются составлением списков рекомендованных покупок, следует помнить о том, что представленные рейтинги могут быть подготовлены без учета ваших интересов. Одной из первых инициатив в области больших данных была система бронирования авиабилетов Sabre Global Distribution System. Введенная в эксплуатацию в 1960 году изначально Sabre представляла собой систему для обслуживания авиакомпании American Airlines, которая вложила в ее разработку огромные деньги. В 1976 году систему Sabre начали устанавливать в офисах турагентов, а потому в нее встроили возможность приобретения билетов на рейсы других авиакомпаний<sup>54</sup>. Проанализировав закономерности процесса бронирования, в American Airlines обнаружили, что турагенты обычно выбирают рейсы, которые сразу появляются на экране системы, и в очень редких случаях обращают внимание на то, что находится за пределами первой страницы выдачи результатов поиска<sup>55</sup>. Авиакомпания немного повозилась с настройкой алгоритма таким образом, чтобы ее рейсы всегда оказывались первыми в списке. Потребители не знали, что представленные им «лучшие» варианты подобраны с таким искажением. А с учетом того, что турагенты работают на комиссии, им было не слишком интересно искать для своих клиентов варианты дешевле. Тем не менее две из конкурирующих с American Airlines авиакомпаний, New York Air и Continental, обнаружили, что их рейсы оказываются в самом низу списка, даже несмотря на открытие новых направлений и предложение скидок на билеты – что, по идее, должно было ставить их варианты на самый верх в выдаче результатов<sup>56</sup>. Потребовалось расследование конгресса<sup>57</sup>. В 1984 году такого рода искажения настроек были официально запрещены<sup>58</sup>.

Совершать подобные манипуляции становится намного труднее, когда пользователем обработанной информации является конечный потребитель, который с большей вероятностью обратит внимание на то, насколько предложения соответствуют его пожеланиям. Я участвовал в разработке системы рекомендаций для бангкокского сайта бронирования отелей Agoda. На первый взгляд могло показаться, что компании выгоднее ранжировать отели исходя из величины получаемой от рейтинга прибыли. Если отель готов платить Agoda более высокие комиссионные, то почему бы не поместить его на самый верх списка? Или все-таки рейтинг должен составляться на основе предпочтений туристов? Кто-то из клиентов, ориентируясь на рейтинг, составленный с учетом интересов Agoda, бронировал номер, но впоследствии сожалел об этом. Другие смотрели на начало списка, приходили к выводу о том, что у Agoda нет вариантов, соот-

---

<sup>52</sup> Finley, Klint, “Christmas Delivery Fiasco Shows Why Amazon Wants Its Own UPS”, Wired, December 30, 2013, [http://www.wired.com/2013/12/amazon\\_ups](http://www.wired.com/2013/12/amazon_ups).

<sup>53</sup> Kastrenakes, Jacob, “Amazon Guarantees Packages Ordered Through Friday Will Arrive Before Christmas”, The Verge, December 16, 2014, <http://www.theverge.com/2014/12/16/7401299/amazon-sets-dec-19th-cutoff-for-christmas-free-shipping>.

<sup>54</sup> Snyder, Brett, “Sabre Makes the Wrong Choice by Removing American Airlines”, CBS News Moneywatch, January 7, 2011, <http://www.cbsnews.com/news/sabre-makes-the-wrong-choice-by-removing-american-airlines>.

<sup>55</sup> American Airlines, “November Line of Sale Analysis”, memo to R. E. Murray from S. D. Nason, December 3, 1981.

<sup>56</sup> Tefft, Sheila, “Reservation Systems’ Bias a Sore Spot for Smaller Airlines”, Chicago Tribune, February 11, 1983, <http://archives.chicagotribune.com/1983/02/11/page/87/article/new-technology>.

<sup>57</sup> Whiteley, David, *An Introduction to Information Systems* (New York: Palgrave Macmillan, 2013), p. 109.

<sup>58</sup> С вступлением в силу Закона о либерализации авиаперевозок в 1978 году Совет по гражданской авиации ввел регламент антидискриминационных мер. See Pearlstein, Debra J., and Robert E. Poch et al., eds., *Antitrust Law Developments*, vol. 1 (Chicago: American Bar Association, 2002), p. 1428.

ветствующих их вкусам, и уходили к конкурентам. В долгосрочной перспективе более оптимальным для Agoda решением было увязывать свои интересы с интересами клиентов<sup>59</sup>.

Последний уровень работы с данными – предписывающая или инструктивная аналитика, которая на основе вашей информации подсказывает, каким образом следует изменить условия для достижения желаемого результата. Классический пример – анализ данных, примененный в ходе лунной экспедиции НАСА<sup>60</sup>. Для того чтобы доставить на поверхность Луны Нила Армстронга и американский флаг, НАСА приходилось непрерывно анализировать поток данных о положении лунного модуля в пространстве. Инженерам в центре управления нужно было не только обобщить данные (описание) и не только спрогнозировать, где и когда лунный модуль коснется поверхности (прогноз). Чтобы человек действительно попал на Луну, им нужно было определять предпочтительные действия в связи с постоянно изменяющимся положением модуля. Они оценивали, какое влияние оказывает каждое включение любого из реактивных двигателей модуля на траекторию его движения. После этого они прогнозировали, когда и как надолго следует включить его опять для успешного выполнения задачи.

Информационная грамотность подразумевает понимание того, что допущения – свойственная описаниям неопределенность – это неотъемлемый элемент прогноза, а для инструкций обязательно нужна обратная связь. Стоит ли инфообработчикам помещать вас в какой-либо маркетинговый сегмент на основе истории ваших поисковых запросов в Google? Можно ли объективно судить о кандидате на должность исключительно на основе анализа данных о его контактах в LinkedIn? Насколько обоснованными будут индивидуальные рекомендации по физическим нагрузкам, в основу которых положен анализ информации из Facebook о посещении этим человеком ресторанов?

## Эксперименты, эксперименты, эксперименты

Инфопереработчики не только описывают, прогнозируют и инструктируют – они еще и экспериментируют. Вполне возможно, что над вами экспериментируют каждый раз, когда вы покупаете бестселлеры в Amazon, подбираете себе мокасины в Zappos или ищете пару на Match.com. Эксперименты нужны, чтобы совершенствовать продукты и услуги инфопереработки при помощи так называемого А/В-тестирования.

В науке причинно-следственная связь устанавливается экспериментальным путем: реакция на изменение одной независимой переменной в экспериментальной группе сравнивается с реакцией контрольной группы, для которой эта переменная остается неизменной. А/В-эксперименты, как правило, начинаются с вопроса. Например: «Какими зонтиками, красными или синими, мне нужно торговать, чтобы максимизировать их продажи?» Этот вопрос кажется очень простым, но из него вытекает масса сложностей в проведении удачного А/В-эксперимента. Продавец зонтиков может попытаться найти правильное решение, поставив свой прилавок на некой точке и продавая только синие зонтики в первый день и только красные – на второй. Он может даже проводить этот эксперимент два понедельника подряд, когда работающие в этом районе вроде бы должны быть более склонны забывать зонтики в суматохе перед выходом из дому. Но, определяя место для торговой точки и день недели для эксперимента, он не принимает в расчет одну из самых важных переменных, определяющих потребность в любом зонтике, красном или синем, а именно – идет ли дождь.

Инфопереработчикам приходится учитывать намного больше переменных, чем нашему торговцу зонтиками. В Amazon все, что касается внешнего вида страниц, от размера строки поиска до места размещения диалогового окна, от опций оформления и оплаты до части опи-

---

<sup>59</sup> Я консультировал Agoda в 2004–2007 гг. В ноябре 2007 года компания была поглощена Priceline.

<sup>60</sup> В технике предписывающая аналитика называется «теорией управления».



сания товара, доступной без второго клика, решается после проведения А/В-экспериментов. Широко известна история про то, как Google проводила А/В-эксперименты для определения оттенка синего цвета для рекламных ссылок. Источники в Google утверждают, что в результате выбора одного из пятидесяти возможных вариантов ежегодная выручка от рекламы возросла на 200 миллионов долларов<sup>61</sup>.

Описательный анализ дает возможность выявлять «естественные эксперименты» – ситуации, когда можно проследить последствия изменения некоего условия, произошедшего случайно или по ошибке (например, когда при внедрении программного обеспечения обнаруживается баг). Веб-разработчики французского сайта Amazon каким-то образом умудрились забыть добавить услугу доставки в стоимость оформления заказа. Резкий рост заказов, последовавший в результате этой ошибки, дал Amazon представление о том, насколько бесплатная доставка способствует увеличению продаж.

В основе научного метода лежит прогноз: ученый создает предполагающую нечто модель, проводит эксперименты и выясняет, насколько их результаты соответствуют предположению. Если они не соответствуют, ученый вносит изменения в модель и повторяет процесс тестирования.

В области социальных данных меня больше всего интересуют эксперименты с элементами прескрипции, в которых пользователь получает возможность изменить некий параметр и увидеть, как это повлияет на результаты. Обработка данных о пробке на дороге позволяет предупредить водителей, сообщить им о том, насколько увеличится время в пути, и предложить альтернативные варианты маршрута. Если большинство водителей изберет какой-то один альтернативный маршрут, то на нем может возникнуть еще одна дорожная пробка. Для подобных ситуаций может быть предложен набор различных вариантов объезда и информация о том, какая часть водителей уже выбрала определенный маршрут, чтобы позволить принять решение поехать другой дорогой. Эти же данные можно использовать и для того, чтобы в целях оптимизации транспортного потока изменить частоту смены сигналов светофоров.

Один из лучших умов в области А/В-экспериментов – мой бывший коллега Рон Кохави, покинувший Amazon в 2005 году, чтобы возглавить работу по созданию экспериментально-аналитической группы в Microsoft. Чтобы создать базовые практики удачного онлайн-экспериментирования, Ронни и его команда провели сотни экспериментов примерно на двадцати интернет-сайтах (в том числе на MSN.com и Bing). Исходя из своего опыта Ронни утверждает: «Получить цифры легко; получить цифры, заслуживающие доверия, – трудно»<sup>62</sup>. Я полностью с этим согласен. Более того, то же самое можно сказать и о наиболее фундаментальных аспектах обработки данных: создавать рекомендации просто, оценивать рекомендации – трудно.

При проведении А/В-экспериментов на интернет-сайтах очень многое может пойти не так. Для начала: от 15 до 30 процентов просмотров страниц некоторых сайтов приходится на долю поисковых роботов, и эти визиты надо идентифицировать и отделить от посещений людей (если, конечно, обработку данных не проводят с целью оптимизации роботов).

Существует также искушение разделять пользователей на экспериментальные и контрольные группы по каким-то признакам, а не произвольно. Однако, каким бы разумным это ни выглядело на первый взгляд, большинство неслучайных способов выборки искажают результаты экспериментов и загрязняют аналитику. Например, если пользователь часто удаляет со своего компьютера cookie-файлы, его могут отнести к одной группе во время первого визита на сайт и к другой во время следующего. В некоторых экспериментах отнесение к той или

---

<sup>61</sup> Hern, Alex, “Why Google Has 200M Reasons to Put Engineers over Designers”, Guardian, February 5, 2014, <http://www.theguardian.com/technology/2014/feb/05/why-google-engineers-designers>.

<sup>62</sup> Kohavi, Ron, Roger Longbotham, and Toby Walker, “Online Experiments: Practical Lessons”, IEEE Computer 43, no. 9 (September 2010), pp. 82–85, <http://www.computer.org/csdl/mags/co/2010/09/mco2010090082-abs.html>.

иной группе коррелирует с сайтом, на котором был пользователь, прежде чем попасть на экспериментальную или контрольную страницу. Действительно ли люди более склонны кликать рекламу зонтиков на сайте WeatherChannel, где постоянно предупреждают о грядущих ураганах? Если выборки делались не по принципу случайности, результаты окажутся искаженными.

Кроме того, ученые стараются учитывать переменные, которые могут влиять на поведение пользователя, но не включены в эксперимент. Результаты эксперимента могут быть искажены багом, который появился в версии программного обеспечения, предложенной одной группе, но отсутствует во всех остальных. Проблемой могут быть и особенности работы программного обеспечения на различных платформах. Люди, использующие для доступа в сеть айфоны, и люди, использующие для этого телефоны на Андроиде, не являются равномерно распределенными и не зависимыми друг от друга группами населения. По результатам эксперимента может казаться, что на сайт чаще заходят пользователи с айфонами, но на самом деле различались не клиентские базы, а программное обеспечение – частота обновления страниц по умолчанию в айфонах выше. Придумывать такого рода версии и расследовать их – ежедневная работа детективов мира данных.

Компании экспериментировали на потребителе, предлагая новые товары или упаковки, задолго до появления интернета. Новизна состоит в возможности экспериментировать в режиме реального времени и моментально получать обратную связь, которая может быть использована для совершенствования продуктов и услуг, в том числе информационных. В прошлом цикл «идея – результат» измерялся месяцами. Теперь же, в мире с выходом в интернет, его продолжительность сократилась до минут. Это в корне отличается от временных параметров медицинских исследований, где эффекты от изменения рецептуры лекарства могут проявляться через недели, месяцы, годы, а то и десятилетия.

По мере дальнейшей интеграции социальных данных в процессы решения проблем и принятия решений обычными людьми инфопереработка будет создавать продукты и услуги в таких важнейших областях человеческой жизни, как здравоохранение и образование. Нам как обществу нужно будет решать, какие эксперименты с социальными данными желательны и каким результатам мы можем доверять. Когда достаточно данных, собранных за час или за день, а когда лучше проводить эксперимент в течение более длительного периода? Например, в области образования очевидных ответов на этот вопрос нет. Для того чтобы результаты А/В-эксперимента способствовали совершенствованию процесса преподавания, необходимо прежде всего сформулировать задачи, ради которых проводится сбор и анализ данных. Как я уже говорил, давать рекомендации просто, а оценивать их – трудно.

Этот процесс не должен отпугивать. Эксперименты с социальными данными пошли на пользу всем: нам стали доступны услуги и продукты, которые еще два десятка лет назад было трудно себе представить. Сегодня они уже считаются едва ли не такими же предметами первой необходимости, как водопровод и электричество<sup>63</sup>. Инноваций в использовании социальных данных будет еще больше, их сдерживают лишь рамки бюджетов, социальных шаблонов и творческих способностей. Чтобы обратить результаты деятельности по обработке больших социальных данных на пользу себе, следует принять факт того, что над нами экспериментируют, и доходчиво убедить аналитиков в том, что эти эксперименты должны помогать нашим решениям. Мы не должны оставаться в неведении.

Поэтому, прежде чем перейти к вопросу о новых правилах работы с данными, которые я считаю необходимыми для регулирования деятельности инфопереработчиков, нам следует поближе познакомиться с тремя категориями социальных данных – нашими кликами, нашими контактами и нашим контекстом. Нам предстоит убедиться в том, что использование

---

<sup>63</sup> Döpfner, Mathias, “An Open Letter to Eric Schmidt: Why We Fear Google”, Frankfurter Allgemeine Zeitung, April 17, 2014, <http://www.faz.net/aktuell/feuilleton/debatten/mathias-doepfner-s-open-letter-to-eric-schmidt-12900860.html>

этих источников первичной информации противоречит многим существующим социальным нормам, в том числе и глубоко личного и эмоционального характера. Как идентифицируется личность? До какой степени иллюзорны представления о приватности? Что означает понятие «друг»? Как наше окружение влияет на нас, а мы на него? Возможно, для вас станет сюрпризом, что ответы на эти вопросы содержатся в истории ваших поисков в Google, вашем взаимодействии в Facebook и в сенсорных датчиках вашего мобильного телефона.

## Глава 2

### Знаки и значения

# Противостояние между цифровым частным и цифровым честным

*...чтобы суметь быть тем, что он есть, он должен верить, что он есть нечто большее<sup>64</sup>.*

*Роберт Музиль*

#### ***Вы – это созданная вами информация?***

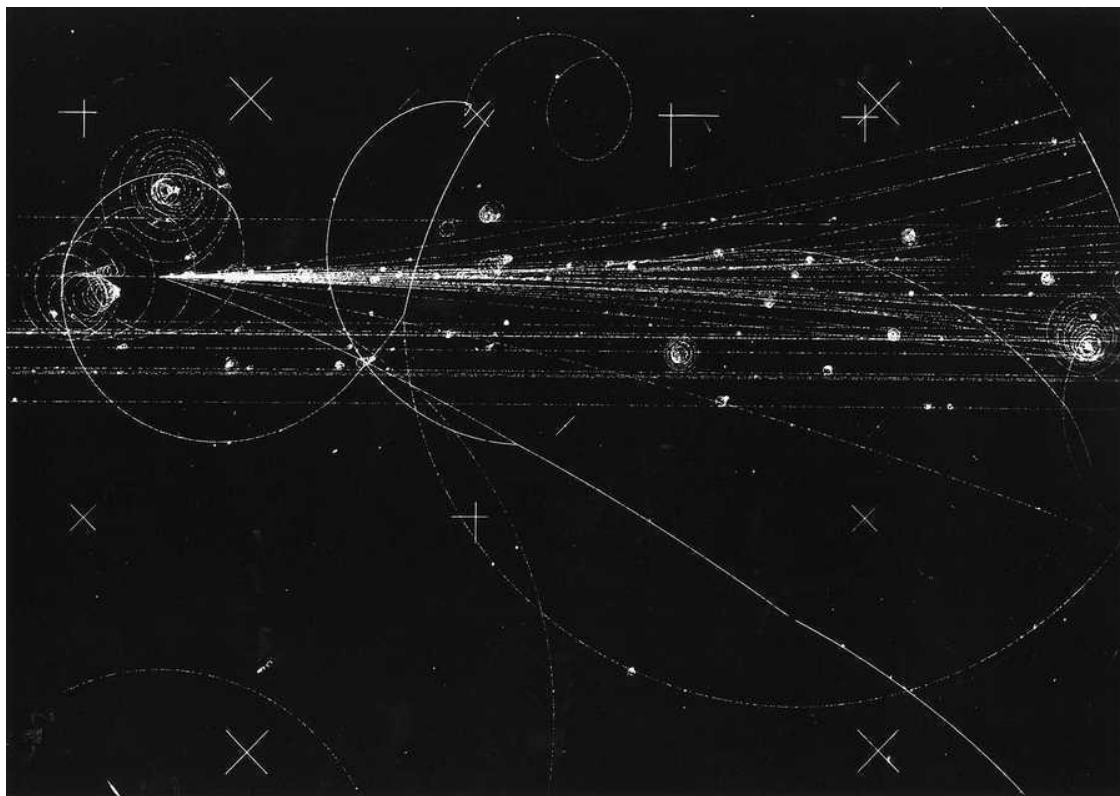
Как и многие из тех, кто в наши дни экспериментирует с социальными данными, по образованию я – физик. Если вдуматься, в этом нет ничего особенно удивительного: цифровые следы, которые мы оставляем в интернете или при использовании мобильных телефонов, очень похожи на «хвосты» и импульсы, которые улавливает детектор частиц. Опыт экспериментальной работы в физике элементарных частиц – действительно идеальная тренировочная база для того, чтобы в дальнейшем заниматься тем же в электронной коммерции.

В физике высоких энергий наблюдать за частицами как таковыми невозможно. Можно лишь наблюдать за их взаимодействием с созданным вами детектором. Анализируя это взаимодействие, физики приходят к выводам о свойствах частиц. Студентом-старшекурсником я работал в лаборатории физики элементарных частиц ЦЕРН под Женевой<sup>65</sup>. Я занимался данными, полученными в опытах с пузырьковой камерой, в том числе изучением траектории и радиуса микроскопических пузырьков, образующихся при попадании частицы в камеру и ее взаимодействии с почти кипящей жидкостью. Пробег и размер пузырьков использовались для расчета электрического заряда и массы частицы. Этот принцип косвенного наблюдения применяется во всех экспериментах физики элементарных частиц: никто никогда не увидит бозон Хиггса, но полученные косвенные данные дают подавляющему большинству физиков основание считать, что эта частица существует.

---

<sup>64</sup> Цитируется по Hochschild, Jennifer L., “How Ideas Affect Actions”, in Robert Goodin and Charles Tilly, eds., Oxford Handbook of Contextual Political Analysis (Oxford: Oxford University Press, 2006), pp. 284–296.

<sup>65</sup> CERN – Conseil Européen pour la Recherche Nucléaire, Европейский центр ядерных исследований. – *Ред.*



Следы протона, создающего двадцать шесть заряженных частиц в эксперименте с пузырьковой камерой в fermilab.

Подобно элементарным частицам, люди обладают свойствами, в которых можно убедиться, лишь наблюдая, как они взаимодействуют с другими людьми и оценивают их. Это несколько напоминает главного персонажа романа Роберта Музиля «Человек без свойств»<sup>66</sup> – математика, черты характера которого определяет окружающий мир<sup>67</sup>. Оставленные нами следы говорят о нас очень многое, что ставит под вопрос саму возможность сохранения тайны личной жизни. Эти следы также позволяют инфопереработчикам наблюдать за нашим поведением и интерпретировать его, а затем прогнозировать наши действия и области интереса, причем вне зависимости от того, известно нам об этом или нет.

Новые социальные платформы, где мы активно делимся созданной нами личной информацией, предоставляют невиданные ранее возможности для самовыражения. Мы можем тщательно подбирать атрибуты для демонстрации своей персоны, причем не только внешние вроде прически или стиля одежды, как это происходило раньше в рамках так называемого показного потребления. Теперь происходит то, что я называю показной коммуникацией: мы можем с легкостью изменять характерные признаки своего цифрового образа, от фото в профайле до псевдонимов, используемых на разных платформах. На заре интернета нам казалось, что анонимность предоставляет поразительную новую свободу, однако смена образов не проходит даром. Желание оградить себя от риска злонамеренного использования подлинных личных данных вполне естественно, но в то же время сохранение анонимности сокращает возможности использовать эти данные с выгодой для себя.

<sup>66</sup> На русском языке роман издавался несколько раз. Первое издание – издательский дом «Ладомир», 1994 г. В 2008 г. вышел в издательстве «Эксмо». – *Ред.*

<sup>67</sup> Название романа Музиля *Der Mann ohne Eigenschaften* в буквальном переводе на английский звучало бы как «Человек без своего собственного». Подозреваю, что английский издатель был обеспокоен тем, чтобы читатели не решили, что у главного героя полностью отсутствует имущество.

Для многих людей инстинктивной реакцией на возможность использования их данных против них самих является требование усилить меры защиты личной информации, поскольку личная информация должна быть неприкосновенна. Однако многие из подобных мер на самом деле лишь ограничивают возможность возложить ответственность на конкретного человека. Дефицит ответственности идет вразрез с теми принципами прозрачности и свободы выбора, которые призваны помогать нам извлекать выгоду из своей личной информации. В мире социальных данных старые правила уже не работают. Экономика повсеместного создания и распространения информации требует от нас создания новых норм и идеалов.

Но прежде чем заглянуть в будущее, давайте исследуем историю такой категории, как неприкосновенность личной информации. Оказывается, само это понятие появилось относительно недавно как следствие развития технологий. И заложенные в нем принципы и стандарты неспособны обеспечить нашу безопасность в эпоху открытости социальных данных. Не надо бороться за неприкосновенность личной информации только потому, что сто лет назад подобная борьба считалась действенным способом решения проблемы.

### **Краткая история недолгой истории приватности**

Большую часть истории человечества люди жили у всех на виду. Мы жили не только с самыми близкими, но и со всей семьей, собиравшейся вокруг домашнего очага и знавшей всю нашу подноготную. Такое «добрососедство» означало постоянное пристальное наблюдение за нами. Нарушение общественных норм влекло за собой безжалостное осуждение и сплетни, а тех, кто действительно пошел наперекор правилам, в лучшем случае изгоняли.

Вполне возможно, что первым технологическим средством обеспечения тайны личной жизни стали камины<sup>68</sup>. В XVII веке они получили широкое распространение в домах европейцев, что позволило многим семьям разделить дома на отдельные комнаты со стенами и дверями, отгораживавшими людей от любопытных глаз родственников. Примерно тогда же целая серия инноваций в сельском хозяйстве изменила способы заработка на жизнь<sup>69</sup>. К середине XVIII века темпы роста производства продуктов питания опережали темпы роста населения. Многие люди переезжали в города, где появлялись первые фабрики эпохи промышленной революции.

Жители городов вели обособленный образ жизни и держали двери закрытыми. У этой новообретенной приватности были издержки: конструкция каминов, обычно достаточно глубоких, чтобы ставить на огонь кухонные котлы, была крайне неэффективной, и в небольших помещениях скапливался дым. Такова была цена обособленности жилья. В 1740-х годах условия наконец начали улучшаться. Не кто иной, как Бенджамин Франклин, предложил конструкцию так называемого Пенсильванского камина, который быстрее обогревал помещение и имел дымоотведение в дымоход<sup>70</sup>. Наконец-то и обычные люди получили возможность захо-

---

<sup>68</sup> Заслуга моего знакомства с представлением о камине как о технологическом средстве обеспечения тайны личной жизни принадлежит блистательному Джону Тэйсому – основателю BlinkBox Music и Reuters Venture Capital. Возможность сохранять определенную приватность при доступе к услугам инфопереработчиков всегда была одной из приоритетных задач в работе Джона. См. Taysom, John, “How Much Privacy Do We Need?”, presentation at the Alan Turing Institute Financial Summit, British Library, London, October 14, 2015.

<sup>69</sup> Это особенно справедливо в отношении Англии, где земли переходили из общинной в частную собственность в процессе консолидации мелких земельных наделов. В числе других сельскохозяйственных инноваций того времени, способствовавших переменам, были севооборот, селекция в животноводстве, более эффективный железный плуг, более эффективная мелиорация и развитие сети каналов. См. Overton, Mark, *Agricultural Revolution in England: The Transformation of the Agrarian Economy 1500–1850* (Cambridge: Cambridge University Press, 1996).

<sup>70</sup> «Пенсильванский камин» Бенджамина Франклина был на самом деле печью, а не частью здания, но конструкция его дымохода была, бесспорно, революционной. См. Butler, Orville R., “Smoke Gets in Your Eye: The Development of the House Chimney”, n.d., <http://www.ultimatehistoryproject.com/chimneys.html>.

пывать свои двери, не рискуя задохнуться угарным газом. Дом стал святой святых, местом, где люди могли рассчитывать на некое подобие приватности и покоя. Франклин и остальные «отцы-основатели» пошли еще дальше: Четвертая поправка в Конституцию США законодательно закрепила право на неприкосновенность жилища. Она запретила обыск и арест по месту жительства без наличия законных оснований.

Домашняя жизнь становилась все более частной и изолированной, чего нельзя было сказать о политической. В самых ранних демократических опытах человечества голосование являлось публичным действием. Идея заключалась в том, чтобы поощрять свободу волеизъявления обычных людей. Как рассказывает профессор истории Гарвардского университета Джилл Лепор, в первые годы существования Соединенных Штатов голосование происходило прилюдно путем поднятия рук или построения у одной из стен помещения. В штате Айова, например, до сих пор именно так организовано голосование на президентских кокусах (caucus, кокус – собрание членов или сторонников политической партии или политического движения). По мнению Лепора, «тайное голосование считалось чем-то трусливым, нечистоплотным и недостойным», компрометирующим то, что многие философы считали одним из основных элементов демократической системы – прямое и непосредственное информационное взаимодействие<sup>71</sup>. Например, в 1850-х годах английский философ Джон Стюарт Милль утверждал, что тайное голосование подвержено «эгоистическим» интересам<sup>72</sup> и что «нормой должна быть не келейность, а гласность»<sup>73</sup>. Благородный человек должен голосовать исходя из общественных, а не личных интересов, а лучшим способом обеспечить это является ответственный выбор, сделанный в обстановке открытости, иными словами – прозрачный<sup>74</sup>.

В те времена, когда право голоса имели только белые владельцы недвижимости, лучшей из имеющихся технологий было использование избирательных бюллетеней, и она считалась элитарной. Бумажная форма требовала от избирателей грамотности, которой могли похвастаться не все обладатели собственности. Однако со временем бумага, как более надежный способ регистрации волеизъявления, одержала верх над взмахами рук и перемещениями тел. Сначала избиратели должны были являться на выборы с готовыми бюллетенями, заполнив их заранее самостоятельно или попросив кого-то сделать это за себя. Это упрощало процесс голосования. Однако тем, кто мог позволить себе напечатать бюллетени для других, ничто не мешало манипулировать выбором голосующих. Известны случаи использования очевидно корыстных методов ведения избирательных кампаний, когда политическая партия впечатывала в бюллетени для голосования только своих кандидатов и раздавала такие бланки и своим сторонникам, и прохожим на улицах. Бумажный бюллетень вошел в обиход не из-за возможности обеспечить конфиденциальность, а потому, что его можно было хранить. А то, что хранится, можно пересчитывать<sup>75</sup>.

Первый бюллетень для тайного голосования, напечатанный властями, был использован на выборах в австралийском городке Виктория в 1856 году. Спустя еще поколение система при-

---

<sup>71</sup> Этот рассказ о тайном голосовании в большой степени основывается на увлекательной статье Лепор: Lepore, Jill, “Rock, Paper, Scissors: How We Used to Vote”, *The New Yorker*, October 13, 2008, [http://www.newyorker.com/reporting/2008/10/13/081013fa\\_fact\\_lepore](http://www.newyorker.com/reporting/2008/10/13/081013fa_fact_lepore).

<sup>72</sup> Сохранившаяся переписка Милля свидетельствует о том, что самое позднее в 1853 году он уже придерживался прямо противоположных взглядов. См. Buchstein, Hubertus, “Public Voting and Political Modernization”, in John Elster, ed., *Secrecy and Publicity in Votes and Debates* (Cambridge: Cambridge University Press, 2015), pp. 29, 30.

<sup>73</sup> Mill, John Stuart, “Thoughts on Parliamentary Reform”, in *Dissertations and Discussions: Political, Philosophical, and Historical*, vol. 4 (New York: Henry Holt, 1873), pp. 36–37.

<sup>74</sup> Buchstein, “Public Voting and Political Modernization”, p. 31.

<sup>75</sup> Я с изумлением узнал, что, получив патент на механический аппарат для голосования с коленчатым рычагом в 1869 году, Томас Эдисон обнаружил, что спроса на него нет. Политики хотели слышать свою аудиторию и, как можно с уверенностью предполагать, влиять на нее. Рычажный аппарат Эдисона приобрел сторонников лишь в конце 1950-х годов. См. Stephey, M. J., “A Brief History of Ballots in America”, *Time*, November 3, 2008, <http://content.time.com/time/politics/article/0,8599,1855857,00.html>.

жилась в Британии, а американские города и штаты начали переходить на нее в конце 1880-х годов. Реформы прошли с ограниченным успехом. Явка американцев на выборы больше никогда не достигала типичных для середины XIX века значений 80 %<sup>76</sup>, возможно, потому, что неучастие в выборах сейчас не влечет за собой серьезных социальных последствий.

Примерно в то же время, когда приобретало популярность тайное голосование, пара бостонских адвокатов принялась обосновывать необходимость некоего нового «права на неприкосновенность личной жизни». Считается, что впервые это словосочетание употребили бывшие партнеры по юридической фирме Сэмюэл Уоррен и Луис Брэндис в 1890 году в опубликованной в журнале *Harvard Law Review* статье, гневно осуждавшей эскалацию вторжения в личную жизнь граждан. Виновники? «Новейшие изобретения и методы ведения бизнеса», включая фотографии и газеты, распространяющие сплетни в погоне за тиражами<sup>77</sup>. Как и многие другие изобретения, это новое «право на неприкосновенность личной жизни» появилось в целях решения собственной проблемы: семья Уоррена незадолго до этого стала мишенью нелестных и бестактных комментариев в колонках светской хроники<sup>78</sup>. Дело явно происходило не в те времена, когда в Facebook размещается по миллиарду фотографий ежедневно.

Увы, в горячем стремлении оградить своих жен и дочерей от чувства неловкости перед окружающими доблестные адвокаты не увидели различия между желанием определять собственный образ во внешнем мире и правом контроля над тем, что говорят о тебе другие люди. В развитом демократическом обществе никто не ограничен в праве высказывать свои мысли и чувства окружающим, даже подозреваемые в нарушении закона. Если вы поделитесь с кем-то своей тайной, вполне возможно, что о ней узнают и другие. Пресечь подобную болтливость на законодательном уровне невозможно, но справиться с ней могли бы социокультурные нормы, особенно если сохранение тайны происходит в интересах «широкой публики». В технических кругах принято считать, что целью коммуникации является передача информации. Но внутреннее чутьё подсказало создателю Facebook Марку Цукербергу, что информации предназначено быть поводом для коммуникации между людьми.

Около ста лет назад Брэндис вошел в состав Верховного суда США, и тема права на неприкосновенность частной жизни стала одной из главных в его деятельности на этом посту. Брэндис стал увязывать право на приватность с личными свободами, ярыми поборниками которых являются американцы. Взять, к примеру, судебный спор «Майер против штата Небраска» о том, насколько законным является запрет преподавания немецкого языка в этом штате. Дело рассматривалось вскоре после окончания Первой мировой войны, когда антигерманские настроения были все еще сильны. Тем не менее в своем вердикте Верховный суд указал, что любое лицо имеет право «заниматься любыми обычными видами деятельности, приобретать полезные знания, вступать в брак, устраивать жилище и обзаводиться детьми, поклоняться Богу в соответствии со своими убеждениями и в целом пользоваться всеми правами, издавна и общепризнанно считающимися неотъемлемой частью законного стремления свободного человека к счастью» вне зависимости от места проживания<sup>79</sup>. С точки зрения

---

<sup>76</sup> Разумеется, на протяжении многих десятилетий некоторые категории американцев не допускались к голосованию на выборах на основе так называемых «экзаменов избирателей» – проверок на грамотность, которые местные избирательные комиссии специально устраивали предназначенным для отсева преимущественно черным жителям Юга.

<sup>77</sup> Warren, Samuel D., and Louis D. Brandeis, “The Right to Privacy”, *Harvard Law Review* 4, no. 5 (December 15, 1890), [http://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/Privacy\\_brand\\_warr2.html](http://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/Privacy_brand_warr2.html). NOTES TO CHAPTER XXX 243

<sup>78</sup> Glancy, Dorothy J., “The Invention of the Right to Privacy”, *Arizona Law Review* 21, no. 1 (Spring 1979), pp. 9–10, <http://digitalcommons.law.scu.edu/facpubs/317>.

<sup>79</sup> Само это решение означало, что учитель имел право преподавать немецкий, но прецедент использовался впоследствии в качестве основания «права на личную жизнь» во многих других ситуациях – в диапазоне от решения супружеской пары прибегнуть к аборту (дело «Американская федерация планирования семьи против Кейси») до решения однополрой пары на секс по взаимному согласию (дело «Лоуренс против штата Техас»). См. *Meyer v. Nebraska*, 262 US Supreme Court 390 (1923), p. 399, <https://supreme.justia.com/cases/federal/us/262/390/case.html>.



закона вторжение в частную жизнь было приравнено к посягательству на права и свободы граждан.

То, что мы считаем исключительно своим собственным делом, представлялось надежно защищенным от излишне любопытных глаз и злых языков. Или же нам просто хотелось так думать. Отклонения, подобные маккартизму, считались «отдельными аномалиями», когда вмешательство в политические убеждения было оправдано коммунистической угрозой. Но с развитием средств обнаружения информации и коммуникаций в интернете сложившееся понимание приватности стало претерпевать драматические изменения.

### От стен к окнам

В 1996 году, когда Ларри Пейдж и Сергей Брин изучали проблему поиска в сети данных с использованием структуры перекрестных ссылок, они опирались на открытую информацию. Все страницы, просматриваемые Google, были открыты для всеобщего доступа. Кто-то создавал страницу и выкладывал ее в интернет, чтобы остальные могли прочитать написанное, а кое-кто, чтобы добавить ссылку на нее.

На разработку алгоритмов и создание собственной сети серверов требовалось немало денег, и Ларри с Сергеем сообразили, что идеальным способом заработать их будет продажа рекламы, соответствующей поисковым запросам пользователей. Рекламодатели «покупали» ключевые слова, фразы и категории, которые, как они считали, соответствуют интересам их потенциальных потребителей. Отдача была моментальной. Посещаемость сайтов рекламодателей, разместивших рекламу в Google, в том числе и на страницах с контентом, имеющим отношение к их продукции, показала четырехкратный рост по сравнению со средними показателями<sup>80</sup>. Данные о поисковых запросах пользователей стали ценным товаром, поскольку давали возможность отслеживать предметы их интересов.

В апреле 2004 года Google запустила Gmail и получила еще один источник информации об объектах внимания своей аудитории. Сервис анализирует содержание писем пользователя, чтобы определить, какую рекламу вывести на его веб-интерфейс. До этого момента подавляющее большинство считало, что электронная почта – нечто вроде обычного письма, запечатанного в конверт, чье содержание предназначено только получателю. Борцы за приватность утверждали, что пользователи Gmail разглашают содержание всей своей глубоко личной переписки и делают его достоянием компании Google. Сегодня это самый распространенный в мире сервис электронной почты, которым ежемесячно пользуются более миллиарда человек<sup>81</sup>. Большинству из них известно, что компьютеры Google «читают» их переписку, но взамен они имеют бесплатную электронную почту. Эта сделка, включающая и показ индивидуализированной рекламы, заключается добровольно и сознательно.

Амбиции компании Google идут намного дальше рекламы в интернете. В 2013 году был выпущен прототип устройства Google Glass – очков со встроенными сенсорами, позволяющими записывать и передавать происходящее вокруг с ракурса их пользователя. Критики выражали озабоченность тем, что Glass будут использовать для обнародования частных разговоров без согласия их участников. Но Glass – отнюдь не единственное технологическое средство, которое может быть использовано в подобных целях. Для этого прекрасно подойдет и карманный диктофон, и миниатюрная видеокамера. А можно просто вынуть из кармана мобильный телефон, который в числе прочего оборудован такими же сенсорами. Носимые

---

<sup>80</sup> Пресс-релиз Google, “Google’s Targeted Keyword Ad Program Shows Strong Momentum with Advertisers”, August 16, 2000, <http://googlepress.blogspot.co.uk/2000/08/googles-targeted-keyword-ad-program.html>.

<sup>81</sup> Miller, Ross, “Gmail Now Has 1 Billion Monthly Active Users”, The Verge, February 1, 2016, <http://www.theverge.com/2016/2/1/10889492/gmail-1-billion-google-alphabet>.

устройства, к которым относится Google Glass, не надо ниоткуда доставать, чтобы запечатлеть момент, – они находятся в той же степени готовности, что и мы сами.

---

## **ЗАИНТЕРЕСОВАННЫМИ**

## **ЛИЦАМИ РЕВОЛЮЦИИ**

## **В ИСПОЛЬЗОВАНИИ**

## **СОЦИАЛЬНЫХ ДАННЫХ**

## **ЯВЛЯЕМСЯ МЫ ВСЕ**

---

За несколько месяцев до начала работы Gmail в Гарварде запустился скромный сайт под названием Facemash. Сейчас эта история стала уже легендарной: старшекурсник Гарвардского университета Марк Цукерберг написал программу для скрейпинга фотографий из онлайн-справочников девяти общежитий, чтобы студенты могли выбирать, какая из двух случайно выбранных «сексапильнее»<sup>82</sup>. Сайт пользовался бешеной популярностью среди однокашников Марка и одновременно вызвал нешуточный скандал. Университет счел, что публикация фото без разрешения является нарушением авторских прав и неприкосновенности частной жизни, и выдал Марку по полной. Еще раз подчеркну: фото были использованы в новом формате коммуникаций ради удовлетворения свойственного людям стремления посудачить. Это привело бы судью Брэндиса в ужас, но тем не менее десятилетие спустя Facebook стал системой коммуникации по умолчанию для немалой части населения планеты<sup>83</sup>. По состоянию на 2016 год на Facebook присутствовал каждый четвертый житель планеты, причем более миллиарда человек ежемесячно заходили туда со своих мобильных телефонов<sup>84</sup>. Марк расширил границы социальных норм, и вместе с ним на неизведанные территории цифровой идентичности устремился огромный людской поток.

Разумеется, по мере своего роста компания Facebook поняла, что она, как и Google, может зарабатывать на рекламе в интернете. Содержание постов создавало даже больше возможностей для адресной рекламы, чем электронные письма. Люди указывали свое семейное положение, уровень образования, политические взгляды и религиозные убеждения; создавали

---

<sup>82</sup> Идея была не нова – впервые сравнение фотографий стало популярным на сайте “Hot or Not”, который выпускники Калифорнийского университета в Беркли Джеймс Хонг и Джим Янг запустили в октябре 2000 года.

<sup>83</sup> Распространенность Facebook частично ограничивается его блокировкой в ряде стран, в частности в КНР, где он почти полностью заблокирован как минимум с 2009 года. См. Chen, George, “China to Lift Ban on Facebook – But Only Within Shanghai Free-Trade Zone”, South China Morning Post, September 24, 2013, <http://www.scmp.com/news/china/article/1316598/exclusive-china-lift-ban-facebook-only-within-shanghai-free-trade-zone>.

<sup>84</sup> Цифры из отчета Facebook за август 2016 года – <http://newsroom.fb.com/company-info>.

списки любимых фильмов, телесериалов, книг и музыкальных произведений; отчитывались о своих путешествиях; делились мнениями об огромном количестве брендов и рекламных кампаний. Они выкладывали фото – свои, своих детей, любимых собак и кошек. Все это предназначалось аудитории друзей и родственников. Летом 2008 года я оказался в главном офисе Facebook как раз в тот день, когда компания начала размещать на сайте персонализированную рекламу<sup>85</sup>. На рекламных объявлениях нового типа присутствовала кнопка обратной связи. Если пользователю не нравилось то, что ему показывают, его просили объяснить почему. Чтение поступающих отзывов стало откровением. Люди жаловались не на то, что рекламщики используют слишком много личных данных, которые они разместили в Facebook, а на то, что они используют их недостаточно. Типичный пример: «В моем профайле черным по белому написано, что я мужчина, который интересуется мужчинами. С какой стати я получаю рекламу, предлагающую «знакомства с женщинами за пятьдесят»?» Пользователи просили рекламировать то, что может им действительно понадобиться.

В 2016 году Facebook стукнуло тринадцать. Скоро мы увидим поколение, о детстве которого во всех подробностях рассказали в Facebook родители, бабушки и дедушки еще задолго до того, как детишки смогли официально открыть собственный аккаунт. В прошлом у выпускника средней школы была на руках куча документов: свидетельство о рождении, карта вакцинации, диплом и табель успеваемости. У подавляющего большинства были и водительские права<sup>86</sup>. У некоторых были рекомендательные письма от работодателей или религиозных наставников, а возможно, и паспорт. Сравните тех подростков с нынешними детьми, которые подходят к пубертатному возрасту с полным набором открытой информации о себе, созданной родителями, дедушками и бабушками, тетями, дядями, старшими братьями и сестрами и друзьями семьи. Можно найти снимки УЗИ, сделанные до рождения, комментарии о трудностях воспитания малыша, молитвы о поправке здоровья и любые подробности относительно внешности, навыков и увлечений. Почему Facebook все еще запрещает заводить аккаунты тем, кому не исполнилось тринадцати? Гораздо разумнее было бы открывать страницу каждому появившемуся на свет младенцу<sup>87</sup>. Это сможет обеспечить каждому уникальный общепринятый идентификатор, который можно использовать или не использовать по собственному усмотрению. Кроме того, социальные данные будут привязаны к учетной записи, что позволит людям самостоятельно определять их содержание по достижении возраста, в котором законодательство будет это разрешать.

Мы проделали путь от открытого очага, подразумевающего публичность человеческого существования с минимальными возможностями для частной жизни, до законодательного признания права на тайну личной жизни в стенах своих спален, а также конфиденциальности в кабинках для голосования. По мере вплетения интернета в ткань общественной жизни мы не огорчились тому, что наша частная жизнь будет «предана гласности» в обмен на возможность бесплатного и прямого контакта с родными, близкими и совершенно незнакомыми людьми.

---

<sup>85</sup> Я посещал одного из своих бывших студентов, ставшего одним из первых специалистов по обработке и анализу данных компании.

<sup>86</sup> В разных странах удостоверением личности могут служить разные документы. Сегодня в США водительские права есть примерно у половины восемнадцатилетних, и это резкое падение по сравнению с ситуацией десятилетней давности, когда они были у двух третей их тогдашних сверстников. См. Halsey, Ashley III, “Fewer Teens Get Driver’s Licenses”, Washington Post, July 31, 2013, [http://www.washingtonpost.com/local/trafficandcommuting/fewer-teens-get-drivers-licenses/2013/07/31/60a32aae-f9c7-11e2-a369-d1954abcb7e3\\_story.html](http://www.washingtonpost.com/local/trafficandcommuting/fewer-teens-get-drivers-licenses/2013/07/31/60a32aae-f9c7-11e2-a369-d1954abcb7e3_story.html). В 2012 году паспорт был у трети граждан США, а в 1989-м – у ничтожных трех процентов. См. Bender, Andrew, “Record Number of Americans Now Hold Passport”, Forbes, January 30, 2012, <http://www.forbes.com/sites/andrewbender/2012/01/30/record-number-of-americans-now-hold-passports>.

<sup>87</sup> Возможно, Facebook выбрала возраст 13 лет, чтобы компании было проще соответствовать требованиям Закона о защите конфиденциальности детей в интернете (COPPA) 1998 года, требующего, чтобы коммерческие сайты, предназначенные для лиц младше тринадцати, получали согласие их родителей или опекунов на сбор личных данных. См. <https://www.ftc.gov/enforcement/rules/rulemaking-regulatory-reform-proceedings/childrens-online-privacy-protection-rule>.

Период с момента зарождения понятия приватности до отказа от нее занял всего пару веков – не более чем мимолетное мгновение истории человечества.

деревенские сплетни  
*отсутствие частной жизни*

камины и миграция в города (1600-е годы)  
*социальная обособленность и появление понятия частной жизни*

четвертая поправка к Конституции США (1792 год) и переход к тайному голосованию (1856–1896 годы)  
*приватность приходит в политику*

«Право на неприкосновенность частной жизни» (1890 год)  
*тайна частной жизни узаконена*

Google, Facebook и иже с ними (наши дни)  
*тайна частной жизни – иллюзия, и нам нравится делиться информацией о себе*

Мы заботливо оберегали свою частную жизнь на протяжении последних ста лет, но настало время признать, что ее тайна – всего лишь иллюзия. Мы хотим иметь возможность привлекать к себе внимание, ощущать себя частью чего-то большего и контактировать с людьми. Идея, осенившая судью Брэндиса, была прекрасна, но это была идея его эпохи, когда информация была скудной, общественная жизнь ограничивалась рамками местного сообщества, а коммуникации стоили дорого. Тогда было просто предотвратить публикацию своей фотографии в случае, если она вам не нравилась. Сегодня это не так. Кроме того, для демократии анонимность не является настройкой по умолчанию. Лучше создать новые нормы, учитывающие реалии настоящего и возможности будущего, чем идеализировать приватность и надеяться на то, что законы прошлого будут защищать нас и впредь. Чтобы поставить информацию на службу людям, нужны прозрачность и свобода выбора.

Вместо того чтобы тратить силы на разграничение личного и публичного и выстраивать стены, ограждающие информацию (или ограждающие от нее), давайте сосредоточимся на возможности быть самими собой. Это позволит нам в полной мере использовать инфопереработку и достичь баланса между потенциально негативными и потенциально позитивными последствиями распространения социальных данных.

### **Все в интернете знают, что ты собака**

В том, что касается социальных данных, тайны частной жизни не существует. Ее больше нет. Подпись под знаменитой карикатурой из журнала «Нью-Йоркер» гласила: «В интернете никто не знает, что ты собака»<sup>88</sup>. С 1993 года, когда была опубликована эта карикатура, изменилось очень многое. Сегодня правильнее было бы сказать: «Весь интернет знает, что ты собака. Ошейник у тебя синий. Ты гоняешь кошек. А твои хозяева уехали отдыхать». Это потому, что ты сам сообщил эту информацию инфопереработчикам, чтобы иметь возможность общаться

---

<sup>88</sup> Газета «Нью-Йорк Таймс» пишет, что эта карикатура, опубликованная 5 июля 1993 года, воспроизводилась в других СМИ больше, чем любая другая за всю историю существования журнала. См. Fleishman, Glenn, “Cartoon Captures Spirit of the Internet”, New York Times, December 14, 2000, <http://www.nytimes.com/2000/12/14/technology/cartoon-captures-spirit-of-the-internet.html>.

с друзьями и получать персональные рекомендации. Это твоя плата за возможность ознакомиться с рекламой кормов Purina Puppy Chow. А все думали, что в интернете можно оставаться анонимным.

Но информация была привязана к «удостоверению личности» задолго до появления Facebook. В середине 1990-х годов теоретик информатики Латания Суини решила выяснить, насколько в действительности анонимна «анонимная» база данных о состоянии здоровья<sup>89</sup>. Штат Массачусетс решил, что предоставление научному сообществу информации о посещениях госслужащими врачей в клиниках соответствует государственным интересам. Не будучи совсем дураками, чиновники понимали, что передавать эти данные с указанием имен людей нельзя, и удалили ряд идентификаторов – фамилии, адреса и номера карточек социального страхования. Но поскольку статистика предназначалась для целей совершенствования системы здравоохранения, кое-какие данные о пациентах в ней оставили – пол, дату рождения и почтовый индекс. Сопоставив эти три элемента информации с другой базой данных – списками избирателей, зарегистрированных по городскому округу Кембриджа, официально доступных любому человеку за плату в двадцать долларов, Суини смогла вычислить медицинскую карту губернатора штата. После чего доктор Суини «сделала театральный жест, направив копию медицинской карты губернатора со всеми диагнозами и выписанными рецептами прямо ему в офис»<sup>90</sup>.

По оценке Суини, зная пол, возраст и почтовый индекс человека, можно идентифицировать 87 процентов населения США<sup>91</sup>. Более поздние исследования понизили этот показатель до примерно 63 процентов – тоже поразительно много, учитывая, что это можно сделать без использования намного более точных сведений, которыми люди постоянно делятся в Facebook и на других сайтах, где собирают социальные данные<sup>92</sup>. Почему для идентификации личности человека нужно столь мало информации, поясняет порядковый расчет. В США примерно 40 000 активных почтовых индексов и около 300 миллионов человек населения, из чего следует, что на один почтовый индекс приходится примерно по 7000 жителей, которые приблизительно поровну делятся на мужчин и женщин<sup>93</sup>. Если предположить, что количество новорожденных равномерно распределяется по количеству дней в году, то получается, что «привязанными» к одному почтовому индексу будут по десять мужчин или женщин с одинаковым днем рождения.

Теперь посмотрим на социальные данные, которыми обычно располагает инфопереработчик. Представление о том, что человека нельзя идентифицировать по его цифровому следу, рассыпалось в пух и прах после того, как два крупнейших инфопереработчика поделились «обезличенными» социальными данными с учеными. Сначала интернет-провайдер AOL предоставил для исследовательских целей историю поиска 658 000 пользователей за трехмесячный период. Однако по чьей-то оплошности эти данные оказались в сети, и двум журналистам из газеты «Нью-Йорк таймс» удалось установить личности нескольких человек по их поис-

---

<sup>89</sup> Суини провела этот эксперимент, будучи студенткой магистратуры МТИ. Сейчас она преподает государственное управление и технологии в Гарвардском университете и является директором его Лаборатории защиты информации

<sup>90</sup> Ohm, Paul, “Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization”, *UCLA Law Review* 57, no. 6 (August 2010), p. 1720, <http://www.uclalawreview.org/broken-promises-of-privacy-responding-to-the-surprising-failure-of-anonymization-2>.

<sup>91</sup> Sweeney, Latanya, Uniqueness of Simple Demographics in the U. S. Population, Laboratory for International Data Privacy working paper LIDAP-WP4–2000, <http://dataprivacylab.org/projects/identifiability/index.html>.

<sup>92</sup> Golle, Philippe, “Revisiting the Uniqueness of Simple Demographics in the U. S. Population”, *Proceedings of the 5th ACM Workshop on Privacy in the Electronic Society* (New York: Association for Computing Machinery, 2006), pp. 77–80, <http://dl.acm.org/citation.cfm?id=1179615>.

<sup>93</sup> US Post Office FAQ, <http://faq.usps.com>. Если бы были присвоены все 90 000 возможных номеров (10000–99999), процент возможности точной идентификации людей был бы еще выше. Другая причина невозможности более высокого процента возможности точной идентификации состоит в неравномерном распределении населения США по почтовым индексам.

ковым запросам<sup>94</sup>. Это оказалось довольно просто, поскольку люди любят искать информацию о самих себе или своих родственниках или прокладывать маршруты от своего домашнего адреса. Затем интернет-видеосервис Netflix устроил конкурс на самый точный прогноз оценки, которую зритель поставит фильму, на основе анализа предыдущих оценок других пользователей. Для построения алгоритмов участникам нужны были данные, и компания предоставила «100 миллионов оценок, поставленных 480 000 клиентами, с датой каждой оценки»<sup>95</sup>. Имена клиентов не раскрывались, но двум ученым из Университета штата Техас в Остине, Арвинду Нараяну и Виталию Шматикову, удалось деанонимизировать людей из базы данных путем сопоставления обезличенной информации с рецензиями, опубликованными на сайте IMDb.com<sup>96</sup>. В чем, собственно, проблема, если эти рецензии уже были достоянием гласности? А в том, что клиенты Netflix не выкладывают отзывы о всех фильмах, которые смотрят, и некоторые из репертуара «тайно любимых» ими фильмов были весьма показательными. По крайней мере, так утверждала истица, чье имя не разглашается, подавшая на Netflix в суд. У нее возникли опасения, что теперь каждый из 50 000 ученых, получивших доступ к базе данных конкурса, знает, что она лесбиянка.

Даже если вы спокойно отнесетесь к тому, что список просмотренных вами фильмов выложат на всеобщее обозрение, вам вряд ли понравится, если будет обнародована вся история ваших поисковых запросов в интернете. Если вы не отличаетесь от подавляющего большинства людей, то чаще всего вводите в Google Maps свой домашний адрес. Ваше место жительства, посещаемые места, покупки, люди, которыми вы интересуетесь, и проблемы, которые вас беспокоят, относятся к наиболее интимным подробностям жизни. Поисковые запросы отражают и то, что в данный момент волнует общество, и Google предлагает получить представление об этом с помощью обработанной информации на Google Trends. Многие считают, что в Trends преобладают новости, но там можно узнать и о том, что в последние пару лет возрос интерес людей к таким проблемам, как интернет-травля и трансгендер. В то же время поисков по словам «приватность» и «транссексуал» стало меньше<sup>97</sup>.

Теперь представьте, что у вас есть возможность видеть чей-то поиск в режиме реального времени. В 1990-х я навещал своего приятеля из Стэнфордского университета, который работал в стартапе в области поисковых систем. Я мог следить за поступающим потоком запросов. Один из них привлек мое внимание: кто-то только что искал «как совершить самоубийство»<sup>98</sup>. Что делать в таком случае? Отследить пользователя по его IP-адресу через сервис-провайдера и позвонить на «горячую линию» предотвращения самоубийств? А не будет ли это вторжением в личную жизнь? Может быть, сначала попробовать внимательно изучить историю поиска этого пользователя, чтобы попробовать понять его мотивацию и более точно оценить вероятность события, которое сразу же приходит в голову при виде такого запроса? А может быть, это писатель, собирающий материал, у которого и в мыслях нет причинить себе вред? Но затем появляется следующий запрос этого человека – «мост Золотые Ворота», где покончили с жиз-

<sup>94</sup> Barbaro, Michael, and Tom Zeller, Jr., “A Face Is Exposed for AOL Searcher No. 4417749”, New York Times, August 9, 2006, <http://www.nytimes.com/2006/08/09/technology/09aol.html>.

<sup>95</sup> Singel, Ryan, “Netflix Spilled Your Brokeback Mountain Secret, Lawsuit Claims”, Wired, December 17, 2009, <http://www.wired.com/2009/12/netflix-privacy-lawsuit>.

<sup>96</sup> Narayan, Arvind, and Vitaly Shmatikov, “Robust De-Anonymization of Large Sparse Datasets”, paper presented at the 2008 IEEE Symposium on Security and Privacy, Oakland, CA, May 18–21, 2008, pp. 111–125, <http://dl.acm.org/citation.cfm?id=1398064>.

<sup>97</sup> Судья по Google Trends, “большие данные” не были на слуху у общественности до 2011 года.

<sup>98</sup> Что касается темы самоубийства, метаанализ, проведенный учеными Оксфордского университета, показал, что «существенно более половины (59 процентов) опрошенных молодых людей сказали, что вели поиск в интернете по слову ‘самоубийство’». См. Daine, Kate, Keith Hawton, Vinod Singaravelu, Anne Stewart, Sue Simkin, and Paul Montgomery, “The Power of the Web: A Systematic Review of Studies of the Influence of the Internet on Self-Harm and Suicide in Young People”, PLoS One 8, no. 10 (October 30, 2013), <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0077555>.

нию более 1600 человек<sup>99</sup>. И после этого вы спокойно отвернетесь от монитора и вернетесь к своей работе по улучшению качества поиска, забыв о том, что человек в опасности? Простого ответа на подобные вопросы не существует.

Схожим образом подробности о вас, а иногда и о ваших близких сообщают ваши покупки в интернете. Чтобы доставить вам заказ, Amazon нужны данные вашей кредитной карточки, в том числе ваши имя и адрес. Сообщить правильный адрес в ваших интересах, иначе посылка до вас не дойдет. В то же время история заказов может вводить в заблуждение, если в ней указаны вещи, которые покупались для кого-то еще. В Amazon можно пометить приобретаемый товар значком «в подарок», и тогда он не будет учитываться в рекомендациях для вас<sup>100</sup>. Алгоритмы обработки данных могут отделять то, что вы указали как покупку не для себя, от всех других ваших заказов. Когда вы покупаете блузку в подарок женщине, то, выбирая размер, сообщаете информацию о ее комплектации. Если это происходит за одну-две недели до Дня матери, а фамилия получательницы совпадает с вашей, алгоритмы Amazon могут сделать вывод о ваших родственных связях. Возможно, что ближе к следующему Дню матери Amazon порадует вас рекомендациями подарков по этому поводу.

Страничка Your Amazon предоставляет пользователям определенную прозрачность и свободу выбора. Возможность просматривать свою первичную информацию, в том числе историю покупок, позволяет контролировать данные, которые становятся основой для выработки персональных рекомендаций. В историю покупок можно включать и вещи, купленные в других местах, причем даже приобретения, сделанные много лет назад. В 2014 году похожий подход был применен в Facebook: «Журнал действий» представляет собой список запросов друзей, лайков, историй и фото с вашими тэгами, приглашений на мероприятия и многого другого. При желании можно удалять из истории отдельные элементы данных. А поскольку ваша цифровая личность в Facebook используется для персонификации рекламы, удаление части данных позволяет влиять на то, какие предложения вам будут присылать<sup>101</sup>.

Удаление одного, двух или двадцати лайков из истории действий вряд ли изменит общий характер представления о вас. Исследования Дэвида Стиллиуэлла из Психометрического центра Кембриджского университета показали, что действия в Facebook действительно достаточно точно отражают черты личности пользователя. Несколько тысяч пользователей Facebook прошли тест на «Большую пятерку» индивидуальных различий (открытость опыту, сознательность, экстравертность, доброжелательность и нейротизм), а затем Стиллиуэлл предложил другой группе участников опыта оценить черты характера этих пользователей исходя из их профайлов. Обе оценки оказались удивительно схожими. Обычно образ человека в Facebook довольно точно отражает действительность – люди остаются самими собой, даже когда тщательно редактируют свои профайлы в социальных медиа<sup>102</sup>. Если совершенно незнакомые

---

<sup>99</sup> Эту цифру назвал сержант Калифорнийской дорожной полиции Кевин Бриггс, выступая с рассказом о патрулировании моста Золотые Ворота на TED Talk. См. Briggs, Kevin, “The Bridge Between Suicide and Life”, TED Talk, March 21, 2014, [https://www.ted.com/talks/kevin\\_briggs\\_the\\_bridge\\_between\\_suicide\\_and\\_life](https://www.ted.com/talks/kevin_briggs_the_bridge_between_suicide_and_life).

<sup>100</sup> Вы можете сделать это на своей страничке “My Amazon”: <https://www.amazon.com/gp/yourstore/iyr>.

<sup>101</sup> Facebook называет это «удалением», но эта информация не удаляется совсем. Происходит это потому, что найти и вычистить все случаи ее появления на серверах значительно дороже, чем сохранять их, изменив способ пользования. Кроме того, в главе 3 мы убедимся, что Facebook постоянно экспериментирует, чтобы усовершенствовать взаимодействие с пользователями и повысить их активность. В одном из исследований рассматривалась самоцензура – случаи, когда люди начинали писать апдейты и комментарии, но в конечном итоге решали не публиковать их. Отмененные апдейты и комментарии нельзя удалить, поскольку их не публиковали, и тем не менее в Facebook изучают их контекст – всевозможные «как и где». См. Das, Sauvik, and Adam Kramer, “Self-Censorship on Facebook”, Proceedings of the 7th International AAAI Conference on Weblogs and Social Media, Cambridge, MA, July 8–11, 2013 (Palo Alto: AAAI Press, 2013), <https://www.aaai.org/ocs/index.php/ICWSM/ICWSM13/paper/viewFile/6093/6350>.

<sup>102</sup> Bachrach, Yoram, Michal Kosinski, Thore Graepel, Pushmeet Kohli and David Stillwell, “Personality and Patterns of Facebook Usage”, Proceedings of the 4th Annual ACM Conference on Web Sciences, Evanston, IL, June 22–24, 2012 (New York: Association for Computing Machinery, 2012), pp. 24–32, <http://dl.acm.org/citation.cfm?id=2380722>.

люди могут оценить основные черты вашего характера по ленте событий на вашей странице в Facebook, то алгоритмы тем более в состоянии это сделать. Чтобы иметь возможность держать друзей и знакомых в курсе своей жизни, придется мириться с выводами, к которым придут эти алго-ритмы.

В 2013 году Стиллуэлл, его коллега Майкл Косински и группа исследователей из Microsoft Research решили выяснить, насколько точно можно судить об интеллектуальном уровне, этнической принадлежности, политических взглядах, сексуальной ориентации и наличии наркозависимости по поведению человека в Facebook. Для этого они создали приложение YouAreWhatYouLike. По утверждению авторов, в 88 процентах случаев их модель «проводила точные различия между мужчинами гетеросексуальной и гомосексуальной ориентации» исключительно на основе лайков, причем не обязательно в темах, связанных с политикой или правами человека<sup>103</sup>. По данным исследования, достоверными признаками мужской гомосексуальности были, в частности, лайки на «Косметику MAC» и мюзикл «Злая», а среди явных признаков мужской гетеросексуальности была реакция на Wu Tang Clan<sup>104105</sup>. Проверая кандидатов на позицию, работодатели используют тесты на ай-кью и личностные качества. Вполне возможно, что в один прекрасный день вас попросят установить специальное приложение, чтобы оценить справедливость ваших утверждений о своей высокой организованности или стрессоустойчивости<sup>106</sup>.

Данные о личных качествах могут формироваться и без активного участия человека. Один из таких примеров – несметное число фотографий, выложенных в сеть. Появление ваших изображений в интернете – вне вашего контроля, а права на них – и подавно. Если вас случайно сфотографировали во время какого-то мероприятия, идентификация вашей личности всего лишь вопрос времени. В лаборатории искусственного интеллекта Facebook, которую возглавляет Ян Лекун, разработана система DeepFace, которая может определять идентичность лиц на фотографиях с очень высокой точностью<sup>107</sup>. Система пока не может самостоятельно определять имя человека на изображении, но если фото подписано, алгоритм присвоит эту подпись всем остальным фото с похожими лицами. Создается и другая программа, которая будет анализировать место действия, то есть сможет различать, сфотографированы ли вы в людном баре или на пустынном холме. В зависимости от того, где вас фотографируют чаще, система отнесет вас либо к любителям потусоваться, либо к одиноким странникам.

---

<sup>103</sup> Kosinski, Michal, David J. Stillwell and Thore Graepel, “Private Traits and Attributes Are Predictable from Digital Records of Human Behavior”, *Proceedings of the National Academy of Sciences USA* 110, no. 15 (April 9, 2013), p. 5802, <http://www.pnas.org/content/early/2013/03/06/1218772110>. Одна из трудностей, возникающих при оценке состоятельности подобных научных трудов, вызвана статистикой свойств личности. Так, если модель построена исходя из предпосылки гетеросексуальной ориентации 100 процентов анализируемых мужчин, она будет точной в 90 процентов случаев, поскольку примерно 10 процентов пользователей мужского пола указывают на наличие у себя «интереса к мужчинам». Тем не менее, выводы исследования на основе приложения YouAreWhatYouLike представляют интерес, особенно в силу того, что такие индивидуальные особенности, как, например, уровень интеллектуального развития, выявляются другими методами.

<sup>104</sup> Хип-хоп-группа. – *Пед*.

<sup>105</sup> Kosinski, Stillwell, and Graepel, “Private Traits and Attributes Are Predictable from Digital Records of Human Behavior”, p. 5804.

<sup>106</sup> Для доступа к лайкам в Facebook исследователи использовали программный интерфейс приложения (API); им было труднее идентифицировать тех, кто использовал настройки приватности для ограничения доступа к своим лайкам, даже несмотря на их согласие участвовать в оценке индивидуальных особенностей. См. <http://appliedmagicsauce.com>. В интервью Косински говорил: «Это может приносить огромную пользу в деле подбора персонала»; Adams, Stephen, “‘Like’ Curly Fries on Facebook? Then You’re Clever”, *Telegraph*, March 12, 2013, <http://www.telegraph.co.uk/technology/news/9923070/Like-curly-fries-on-Facebook-Then-youre-clever.html>.

<sup>107</sup> Simonite, Tom, “Facebook’s New AI Research Group Reports a Major Improvement in Face-Processing Software”, *MIT Technology Review*, March 17, 2014, <http://www.technologyreview.com/news/525586/facebook-creates-software-that-matches-faces-almost-as-well-as-you-do>; Taigman, Yaniv, Ming Yang, Marc’Aurelio Ranzato, and Lior Wolf, “DeepFace: Closing the Gap to Human-Level Performance in Face Verification”, paper presented at the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, June 24–27, 2014, pp. 1701–1708, [https://www.cs.toronto.edu/~ranzato/publications/taigman\\_cvpr14.pdf](https://www.cs.toronto.edu/~ranzato/publications/taigman_cvpr14.pdf).



Научный сотрудник Microsoft Research Синтия Дворк с коллегами доказали, что сам факт существования баз данных подразумевает информационную открытость любого человека. Базы данных существуют для того, чтобы предоставлять ответы, и можно сформировать такую последовательность вопросов, утвердительным ответам на которые будет соответствовать единственный человек в базе. Обычно Синтия демонстрирует это на таком примере: сначала она спрашивает, сколько человек с признаками серповидноклеточной анемии значится в медицинской базе данных сотрудников Microsoft. Затем уточняет, сколько из них мужчин с вьющимися волосами в должности старшего научного сотрудника. Поскольку Синтия – единственный в Microsoft старший научный сотрудник – женщина с вьющимися волосами и признаками серповидноклеточной анемии, разница между ответами на два ее вопроса точно указывает на нее<sup>108</sup>.

Люди предоставляют данные для переработки, чтобы получать результаты, помогающие в принятии решений. В базах данных, похожих на базу из примера Синтии Дворк, собирается относительно специфическая информация ограниченного объема. Это так называемые малые данные. Они не сопоставимы с умом непостижимым количеством «следов», которые накапливают современные центры обработки «больших данных». Чтобы получить от инфопереработчика нечто действительно полезное, надо предоставить ему точные исходные данные, например о ваших интересах и предпочтениях. Если вы не готовы поделиться этой информацией, придется удовлетвориться рекомендациями для среднестатистического гражданина, то есть тем, что пользуется популярностью или подходит большинству обывателей. Если вы предоставите неверные исходные данные, то, скорее всего, получите совершенно бесполезные результаты на выходе. Альтернатива выглядит так – незначительный выигрыш в приватности оборачивается проигрышем в полезности.

### **Что за псевдонимом?**

Решение предоставлять или не предоставлять личную информацию влечет за собой последствия. В одной ситуации раскрытие своих идентификационных данных может оказаться рискованным или вредным; в другой ситуации то же самое может произойти, если они не предоставлены. Цифровые следы, оставленные нами, делают анонимность практически невозможной.

Тем не менее использование реальных имен на социальных платформах начало становиться нормой только с появлением Facebook. До этого обычно использовались псевдонимы. Отчасти это было обусловлено чисто техническими причинами. Некоторые имена распространены настолько широко, что в случае использования настоящих имен различать пользователей было бы невозможно; кроме того, некоторые сайты не принимали имена с большим количеством букв. В то же время были люди, осознанно не желавшие раскрывать имя, опасаясь хищений личных данных или неприятностей в связи со своими высказываниями, не совпадающими с общепринятым мнением. Так или иначе, но при желании можно было создавать разные имена пользователей или даже несколько для каждого сервиса или интернет-форума. В результате в первые десятилетия существования интернет предоставлял неслыханные ранее возможности для фрагментирования собственной персоны. А различные псевдонимы позволяли человеку исследовать новые способы взаимодействия с окружающими.

Исторически человека идентифицировали по ряду простых признаков, вроде имени, даты рождения, роста, цвета глаз, национальности и места жительства. Эта базовая инфор-

---

<sup>108</sup> Синтия использует этот пример в своей лекции под названием «Я – в базе данных (Но никто об этом не знает)»; I'm in the Database (But Nobody Knows), Dean's Lecture, University of California – Berkeley School of Information, February 4, 2015, <http://www.ischool.berkeley.edu/newsandevents/events/deanslectures/20150204>.

мация использовалась для подтверждения того, что человек действительно тот, кем он представляется. Возможность подтвердить личность – необходимое условие выполнения многих законов и правил. Веками для доказательства права на посещение территорий использовались паспорта<sup>109</sup>, а чеки и гарантийные письма служили подтверждением того, что в каком-то далеком банке у нас есть средства, достаточные для оплаты покупки<sup>110</sup>. Возраст или гражданство предоставляют определенные права и обязанности по отношению к обществу, например избирательное право или право употреблять алкоголь в общественных местах, или обязанность платить налоги, или нести воинскую повинность. Мы приучены к тому, что в огромном количестве жизненных ситуаций необходимо предъявить официальное удостоверение личности или сообщить его номер, ввести пароль или ответить на ряд вопросов, диапазон которых – от количества часов, проводимых в авиаперелетах, до домашних животных, которые были у нас в детстве.

Многие из оставляемых вами цифровых следов создаются через взаимодействие с физическими устройствами, и в этом взаимодействии есть немало характерных особенностей, позволяющих вас идентифицировать. Поскольку для выхода в интернет все чаще используются мобильные телефоны и планшеты, многие инфопереработчики вкладывают значительные ресурсы в исследования возможностей идентификации личности на основе устойчивых особенностей поведения при использовании разных устройств. Самым простым способом является требование регистрации пользователя, однако существуют более тонкие признаки, например установленные шрифты. Кроме того, многим людям свойственно постоянно делать одни и те же опечатки. Эта особенность также может быть установлена.

Оставляет след и само физическое взаимодействие с устройством. Сооснователь израильской компании BioCatch Ури Ривнер считает, что цифровые «отпечатки пальцев», то есть манера пользователя работать с компьютером, планшетом или мобильным телефоном, являются «одним из способов опознания личности»<sup>111</sup>. BioCatch создает свою коллекцию данных, заставляя не догадывающихся об этом пользователей выполнять действия, подтверждающие их личности. Компанию не интересует, что именно вы ищете, зато ее интересует, как вы это делаете. Вы бодро ударяете по тачскрину или мягко пошлепываете по нему? Насколько сильно дрожит ваша рука с мобильным телефоном? Где именно на экране вы указываете прокрутку вверх или вниз? Как быстро вы двигаете мышкой? Предпочитаете открывать новые вкладки через ссылки или перемещаетесь между несколькими вкладками? В числе заказчиков BioCatch – банки, которым требуются дополнительные способы аутентификации клиентов<sup>112</sup>. Анализ данных в режиме реального времени может быть полезен для идентификации личности и в других случаях, например, когда документов недостаточно или нет при себе в данный момент.

Детские игровые сайты и приложения вызывают целый ряд вопросов – от безопасности пользователя до адекватности контента. Самый простой пример – сайт, предлагающий игры для пользователей в возрасте от шести до шестнадцати лет, должен обеспечить правильное предложение для игроков соответствующего возраста. При этом разработчиков игр не слиш-

---

<sup>109</sup> Формат паспортов стран был стандартизирован лишь вскоре после Первой мировой войны, но сама идея государственного документа, удостоверяющего личность путешественника, значительно старше – слово «паспорт» появилось в английском языке около 1540 года. См. Benedictus, Leo, “A Brief History of the Passport: From a Royal Letter to a Microchip”, *Guardian*, November 17, 2006, <http://www.theguardian.com/travel/2006/nov/17/travelnews>.

<sup>110</sup> Подлинность чеков требовалось удостоверить – на ранних этапах это делалось путем сличения подписи на чеке с оригиналом из банковского досье. См. Quinn, Stephen, and William Roberds, “The Evolution of the Check as a Means of Payment: A Historical Survey”, *Economic Review* 93, no. 4 (December 2008), [https://www.frbatlanta.org/-/media/Documents/research/publications/economic-review/2008/vol93no4\\_quinn\\_roberds.pdf](https://www.frbatlanta.org/-/media/Documents/research/publications/economic-review/2008/vol93no4_quinn_roberds.pdf).

<sup>111</sup> Цитируется по: Leber, Jessica, “Forget Passwords: This Startup Wants to Authenticate Your Mind”, *Fast Company Exist*, July 24, 2014, <http://www.fastcoexist.com/3033383/forget-passwords-this-startup-wants-to-authenticate-your-mind>.

<sup>112</sup> O’Hear, Steve, “Pre-Crime Startup BioCatch Authenticates Users via Touch and Your Phone’s Accelerometer”, *TechCrunch*, July 7, 2015, <http://techcrunch.com/2015/07/07/pre-crime-startup-biocatch-authenticates-users-via-touch-and-your-phones-accelerometer>.

ком заботит, что восьмилетний ребенок может начать игру, предназначенную для подростков, – считается, что слишком трудная игра ребенка быстро разочарует, а слишком простая так же быстро ему наскучит. Ориентироваться на введенные регистрационные данные сайт не может, поскольку ребенок имеет возможность зайти на сайт с компьютера, которым пользуется кто-то из других членов семьи. Поэтому для оценки возраста пользователя сайты анализируют его взаимодействие с игрой. Часто для обеспечения безопасности в чатах игроков им оставляют только возможность использовать набор готовых фраз – это позволяет избежать риска неумышленного предоставления ребенком домашнего адреса или другой важной информации взрослому, изображающему ребенка. Оказалось, что дети старшего возраста используют другие, чем малыши, готовые фразы. Кроме того, системы игровых сайтов могут определять возраст ребенка с точностью до трех-шести месяцев по характеру движений мышью, которые тесно коррелируют с уровнем развития мелкой моторики, различной в разных возрастах<sup>113</sup>.

Обмануть алгоритмы машинного осмысления, рассматривающие подобные неявные следы, намного сложнее, чем подделать явные отличительные признаки. Если вы находитесь в больнице и человек в белом халате и со стетоскопом на шее просит вас раздеться, вы, скорее всего, решите, что это врач. Но в то же время хорошо известно, что люди по тем или иным причинам порой выдают себя за других. В январе 2015 года в одном из медицинских центров штата Флорида полиция задержала семнадцатилетнего молодого человека, в течение месяца выдававшего себя за доктора. Чтобы проходить через охрану на входе, ему было достаточно белого халата и стетоскопа.

Исторически псевдонимы служили одним из средств обеспечения свободы слова. Статьи из цикла «Федералист»<sup>114</sup>, первая из которых увидела свет в 1787 году, содержали решительную отповедь резкой критике только что увидевшего свет проекта Конституции США. Они были подписаны псевдонимом Публий, скрывавшим авторство Александра Гамильтона, Джеймса Мэдисона и Джона Джея. Лишь немногие из участников той дискуссии выступали под своими настоящими именами<sup>115</sup>. Позже англичанка Мэри Энн Эванс использовала псевдоним Джордж Элиот, чтобы не стать жертвой шаблонного снисходительного отношения к женщинам-писательницам, характерного для XIX века. В своем (тоже анонимном) эссе она отзывалась о написанных дамами романах как о «глупых, поверхностных, банальных, ханжеских или поучающих»<sup>116</sup>. Ей хотелось, чтобы ее творчество воспринималось всерьез, и она опасалась, что это будет невозможно, если читатели увидят на обложке женское имя.

Иногда причиной для использования псевдонима является не столько стремление к свободе самовыражения, сколько желание порвать с прошлым. В 1947 году человек по имени Ганс Фаллада (настоящее имя – Рудольф Дитцен) написал роман «Каждый умирает в одиночку» – историю немецкой супружеской четы, молчаливо сопротивляющейся нацистскому режиму. Сюжет книги был частично основан на реальных событиях: советский атташе по культуре поручил Фалладе ознакомиться с архивами гестапо и создать великое антифашистское произведение<sup>117</sup>. Фалладу не сильно заботило, что его хорошо известное писательское имя будет связано

<sup>113</sup> Рост моторных навыков ребенка останавливается примерно в тринадцатилетнем возрасте (беседа автора с Лэйном Меррифилдом – генеральным директором FreshGrade и сооснователем Club Penguin, 21 января 2016 года). После приобретения Club Penguin компанией Disney в 2007 году Лэйн стал исполнительным вице-президентом интернет-подразделения этой медийной компании, а сейчас занимается над образовательным технологическим стартапом FreshGrade.

<sup>114</sup> Цикл из 85 статей в поддержку ратификации Конституции США, печатавшихся с октября 1787 года по август 1788-го в нью-йоркских газетах The Independent Journal и The New York Packet. – *Ред.*

<sup>115</sup> О первых попытках компьютерного анализа словесных моделей в статьях см. Mosteller, Frederick, and David Wallace, *Inference and Disputed Authorship: The Federalist* (Reading, MA: Addison-Wesley, 1964).

<sup>116</sup> Anonymus, “Silly Novels by Lady Novelists”, *Westminster Review*, new series, vol. 10 (October 1856), p. 442.

<sup>117</sup> Wilkes, Geoff, “Afterword”, in *Alone in Berlin* [Английское название романа *Jeder stirbt für sich allein* (“Каждый умирает в одиночку”)] (London: Penguin, 2009), pp. 578–579.

с политическим заказом. Псевдоним он взял за много лет до этого: похоже, он хотел вырвать свою литературную репутацию из кольца личных проблем<sup>118</sup>.

У этих знаменитых псевдонимов есть одна общая черта: обладатели хотели сделать их реальными именами с собственной репутацией. Все статьи Публия были написаны в поддержку ратификации Конституции. Все произведения Элиота и Фаллады были опубликованы под этими именами. Авторы хотели, чтобы их творчество идентифицировалось именно так.

На заре интернета возможность использования нескольких псевдонимов казалась превосходным вариантом. К сожалению, с этим была некая проблема: создать новый псевдоним легко, но можно ли быть уверенным в том, что он не связан с человеком, которого вытурили с сайта за неделю до этого? Сайт может потребовать, чтобы с каждым псевдонимом регистрировался и адрес электронной почты, но создать новый адрес тоже не проблема. Некоторые сайты отреагировали усложнением регистрационных форм, заполнение которых требовало несколько более серьезных усилий при создании аккаунта, но это не остановило мошенников, которые стали нанимать для заполнения анкет людей или использовать специальные программы – боты. Искоренить такими способами то, что экономист Эрик Фридман и информатик Пол Резник назвали «социальными издержками дешевых псевдонимов», невозможно<sup>119</sup>.

В каких случаях целесообразно использовать псевдоним, а в каких – реальное имя? Это зависит от ситуации. Когда взаимное доверие требуется с самых первых шагов взаимодействия, например, когда речь идет о банковских счетах или о выпуске кредитной карточки, политика «реального имени» полностью оправдана, поскольку включает в информационное поле и историю поведения клиента в прошлом. Напротив, использование псевдонима подразумевает необходимость выстраивания репутации с нуля.

Когда я работал в Amazon, мы занимались исследованием на тему, какие отзывы пользователи считают более полезными – опубликованные под настоящими именами или под псевдонимами<sup>120</sup>. Выяснилось, что отзывы зарегистрированных пользователей Amazon, опубликованные под псевдонимом, несколько реже считают «бесполезными». Мы также убедились в том, что пользователи придают большее значение отзывам, опубликованным с указанием имени автора. Даже если пользователь Amazon изменяет свое имя на сайте, то соответственно изменяются и все его подписи под отзывами как текущими, так и сделанными в прошлом, поэтому история отзывов данного человека полностью сохраняется. Идентификация реального человека с историей его отзывов всегда совпадает, в отношении комментариев под псевдонимом этого утверждать нельзя. Можно было бы настаивать на том, чтобы отзывы публиковались под реальными именами, тем более что реальность имен клиентов Amazon подтверждается данными их кредитных карточек. Но оказалось, что наиболее важным фактором является указание на то, что данный пользователь действительно купил этот товар. Люди действительно больше верят мнению реального человека, но более убедительным аргументом для них служит факт того, что товар был куплен, а не имя автора. С учетом этого в Amazon пересмотрели методику расчета средней оценки товаров, придав больший вес отзывам о действительно куп-

---

<sup>118</sup> Фаллада не только совершил попытку самоубийства, что само по себе было сочтено постыдным, он договорился о совместном самоубийстве с другим человеком, которого ему удалось тем самым погубить. См. Oltermann, Philip, “The Cow, the Shoe, Then You”, *London Review of Books* 34, no. 5 (March 8, 2012), p. 27.

<sup>119</sup> Основополагающая научная работа по этой теме – Friedman, Eric J., and Paul Resnick, “The Social Cost of Cheap Pseudonyms”, *Journal of Economics and Management Strategy* 10, no. 2 (Summer 2001), pp. 173–199, <http://onlinelibrary.wiley.com/doi/10.1111/j.1430-9134.2001.00173.x/abstract>.

<sup>120</sup> В Amazon обсуждались три варианта: практически анонимные отзывы, когда создать новое имя пользователя не составляет никакого труда; отслеживаемые отзывы под псевдонимом, когда пользователь может выбрать любое имя, но оно будет привязано к его аккаунту, удостоверенному действующей кредитной картой; или отзывы под настоящим именем, привязанным к действующей кредитной карте, но с возможностью использования первого инициала автора вместо имени в случае, если он не хочет показывать свой пол.

ленных товарах<sup>121</sup>. (Кроме того, Amazon судилась с некоторыми компаниями, которые предположительно платили потребителям за «пятерочные» отзывы о своей продукции<sup>122</sup>.)

У анонимности есть и недостатки. Задумайтесь о разнице между заполнением «отзыва о качестве обслуживания» в заведении, которое вы регулярно посещаете, и онлайн-опросом. Хотя отзыв на бумаге и выглядит анонимным, многие не заполняют его, и не по причине собственной лени. Люди знают, что их смогут идентифицировать по ряду признаков – почерку, языку изложения, поставленным вопросам и времени, когда была сдана карточка с отзывом. Они могут опасаться последствий, которые способны повлечь нелестные комментарии. Кроме того, анонимный комментарий по определению является «игрой в одни ворота». Диалог между сторонами, а следовательно, и возможности прояснить смысл замечания или намерения человека, или некое сотрудничество, не предусмотрены. Это позволяет получателям обратной связи игнорировать мнение под предлогом незначительности, отклонения от общего вектора, специфичности случая или отсутствия поводов для изменений. От анонимного отзыва можно отмахнуться, сочтя его корыстным или злонамеренным.

---

## СОЦИАЛЬНЫЕ ДАННЫЕ —

### САМОЕ ЦЕННОЕ СЫРЬЕ

#### XXI ВЕКА,

### НОВАЯ НЕФТЬ

---

Сетевым дискуссионным форумам вроде Reddit пришлось разбираться с этими недостатками анонимности с помощью машинного обучения. На Reddit имена пользователей могут быть и постоянными, применяемыми во взаимодействии с сообществом, и одноразовыми – для единственного комментария или голосования. Каждый псевдоним получает полную свободу самовыражения, что стимулирует людей примерять разнообразные личности, выступая на платформе. Пользователей никогда не просят привязывать аккаунты к адресам электронной почты или реальным именам: создатели сайта не хотят взывать к ответственности пользователей за свои действия такими способами. Ответственность регулируется иначе. Интересный пост привлекает внимание, с ним соглашаются или спорят, добавляют комментарии или голосуют «за» и «против» уже опубликованных. Если пост или комментарий получает достаточно много голосов «против», он постепенно исчезает со страницы, уходит в самый низ рэнкинга и попадает в категорию «оценка ниже уровня», хотя при желании его можно найти и просмотреть.

---

<sup>121</sup> Rubin, Ben Fox, “Amazon Looks to Improve Customer-Reviews System with Machine Learning”, CNet, June 19, 2015, <http://www.cnet.com/news/amazon-updates-customer-reviews-with-new-machine-learning-platform>.

<sup>122</sup> Rubin, Ben Fox, “Amazon Sues Alleged Reviews-for-Pay Sites”, CNet, April 9, 2015, <http://www.cnet.com/news/amazon-sues-alleged-reviews-for-pay-sites>. Оказалось, что некоторые компании отправляли заказной почтой пустую коробку или конверт, чтобы получить статус «Подтвержденной покупки».

реть. Сайт не выводит из обсуждения комментарии под псевдонимом, а позволяет пользователям поддерживать диалог по существу и решать, какие комментарии заслуживают их внимания, а какие нет.

Для Reddit в первую очередь важно то, что огромному числу разных людей действительно интересно содержание дискуссий на «горячих», «набирающих обороты» и «скандальных» страницах сайта, а не «официальное» авторство. Дискуссии, попадающие в список двадцати пяти топовых тем на этих страницах, обычно вызывают пристальное внимание всего интернета. Чтобы снизить уровень «фальсификаций при голосовании», когда одни и те же люди выступают под разными именами не для того, чтобы высказаться, а для того, чтобы «заплюсовать» свои посты и «заминусовать» чужие, в Reddit не стали тратить кучу денег и времени на модераторов, следящих за соблюдением порядка, а положились на машинное обучение. Когда одновременно появляется множество активных псевдонимов, идущих с одного IP-адреса или пишущих в одинаковом стиле, их «объединяют в одно производство» как «соучастников». Голоса, поданные «соучастниками», имеют меньший вес, а иногда просто игнорируются.

### Честные сигналы

В 2016 году более 100 миллионов человек обратились к приложениям или интернет-сайтам в поисках знакомств для легких связей, любовных встреч или долговременных отношений. Требуется найти человека, которому нужно то, что предлагается, и обладающего тем, что нужно, и – само трудное – проявляющего ответный интерес.

В том, что касается знакомств, некоторые люди время от времени бывают кое в чем правдивы. Уровень правдивости варьируется в зависимости от характера человека и от ситуации. Иногда люди просто экспериментируют, чтобы понять, чего они хотят на самом деле. Разговоры – это одно. То, что человек делает, – другое. Сигналы, которые следуют из фактических поступков людей, социологи назвали «честными сигналами».

Дизайн пользовательского интерфейса приложения знакомств и разработка алгоритмов рекомендаций – особенно трудная задача, поскольку пользователь может «официально» считать привлекательными одни черты, а характер его работы с сайтом свидетельствовать о совершенно других предпочтениях. Один из создателей сайта OkCupid Кристиан Руддер доказал, что пользователи могут не полностью осознавать свои расовые и этнические предпочтения или не желать в них признаваться<sup>123</sup>. Но простой подсчет кликов и запросов о контакте быстро выявляет такие предпочтения.

Это несколько напоминает старую проблему с оценками кинофильмов. Когда Netflix просит пользователей высказать мнение о фильмах, получивших высокую оценку критиков, вроде игрового «Гражданина Кейна» или документального «Черного плавника», огромная часть людей выставляет им «пятерки», будучи уверены, что они обязаны оценить их именно так: ведь «все» сказали, что это отличное кино. Netflix может рекомендовать клиенту фильмы на основе его оценок, но это имеет смысл только в случае, если это честные оценки. Но люди должны убедиться в том, что правдивые ответы – в их собственных интересах. В Netflix пришли к выводу, что самым честным сигналом об интересе к определенному жанру кино является продолжительность стримминга видео до момента выхода из него. Другими словами, для

---

<sup>123</sup> Rudder, Christian, “How Your Race Affects the Messages You Get”, OkTrends blog, October 5, 2009, <http://blog.okcupid.com/index.php/your-race-affects-whether-people-write-you-back>; and “Race and Attraction, 2009–2014”, OkTrends blog, September 10, 2014, <http://blog.okcupid.com/index.php/race-attraction-2009-2014>. Еще больше интересных подробностей для тех, кто знакомится шаблонным образом, содержится в книге Руддера *Dataclysm: Who We Are (When We Think No One’s Looking)* (New York: Crown, 2014).

составления рекомендаций данные о просмотрах полезнее, чем данные об оценках фильмов<sup>124</sup>. Профессор социальной психологии Мичиганского университета Ричард Нисбетт утверждает, что люди часто не понимают когнитивные процессы, определяющие их поступки и решения. Наши способности к самопознанию и самонаблюдению достаточно ограничены<sup>125</sup>.

В реальности некоторые критерии предпочтений выглядят весьма неопределенно. В подавляющем большинстве случаев люди не отказывают в знакомстве тем, кто на несколько месяцев старше, чем их воображаемый идеал. Тем не менее, консультируя несколько сайтов знакомств, я обнаружил, что очень многие склонны указывать свой возраст «некруглой» цифрой в конце какого-то десятка, например двадцать девять, а не тридцать лет<sup>126</sup>. Солгали ли эти люди по поводу своего возраста в момент создания профайла? Или же они скорректировали свои годы по результатам взаимодействия с приложением, выяснив, что «староваты» для пользователей, которые их заинтересовали по результатам поиска? И я задумался о том, как может повлиять на поведение пользователей возможность увидеть историю редактирования.

В ряде случаев мотив уточняющего редактирования понятен и безобиден. Например, после нескольких знакомств человек может решить уточнить раздел своих интересов, поняв, что излишне напирает на свое мастерство скалолаза и не договаривает о том, как ему нравится ходить на концерты. Аналогичным образом он может скорректировать описание объекта поиска. Исправления же более кардинальные, например частые смены статуса с «одинокого» на «состоящего в отношениях», могут вызвать неодобрение.

Вообразим себе сценарий, при котором пользователи получают возможность видеть не только историю редактирования профайла, но и историю общения на сайте. Обычная проблема приложений знакомств для людей традиционной ориентации в том, что женщин часто засыпают сотнями сообщений, в то время как некоторые мужчины вообще не получают их. Чтобы стимулировать более симметричную коммуникацию, в приложениях знакомств попробовали ограничить число сообщений, которые один человек может направить за определенный период времени. Но поскольку аудитория этих приложений ежедневно меняется (пользователи выбывают и прибывают), это верный способ вызвать ее недовольство. А вдруг девушка вашей мечты активирует свой профайл на следующий день после того, как вы исчерпали свой месячный лимит, а к моменту его возобновления снова исчезнет? Вы же не знаете причины, по которым алгоритм не показывает ее вам, – то ли потому, что вы не отреагировали на ее появление сразу, то ли она уже встречается с кем-то еще. Использование механизмов прозрачности, позволяющих выявить честные сигналы в поведении пользователя, – лучший подход по сравнению с навязыванием жестких рамок. Например, в каждом профайле может быть видно, сколько сообщений отправил и получил этот человек за последние сутки, неделю и месяц, а также средние частоту и скорость откликов. Это позволит более обоснованно решать, с кем контактировать.

Подобная информация уже используется в некоторых приложениях знакомств. Приложение знакомств для геев Jack's предоставляет информацию о частоте откликов на входящие

---

<sup>124</sup> Занимаясь стартапом музыкальных рекомендаций MoodLogic, я часто говорил, что упускать из поля зрения функцию «skip» («пропустить») – ужасная ошибка: негативная обратная связь от пользователя, пропускающего трек, была невероятно полезна для совершенствования алгоритма рекомендаций. Равным образом, как и длинный клик, когда слушают трек целиком, хотя в этом случае уровень интереса нуждался в интерпретации на более длительной основе, исходя из количества повторных прослушиваний и оценок.

<sup>125</sup> Nisbett, Richard E., and Timothy D. Wilson, “Telling More Than We Can Know: Verbal Reports on Mental Processes”, *Psychological Review* 84, no. 3 (March 1977), pp. 231–259, <http://psycnet.apa.org/psycinfo/1978-00295-001>.

<sup>126</sup> На эту удивительную гистограмму с пиком в возрасте 29 лет и провалом в 30 я впервые обратил внимание, работая с сайтом Gay.com. (Людей в возрасте 28 лет было значительно меньше, чем в возрасте 29, но больше, чем 30-летних). С той поры я видел ее на нескольких других сайтах, в том числе на китайском сайте знакомств Baihe, стоимость которого оценивалась в 2015 году в 250 миллионов долларов. Сфера знакомств в Китае невероятно чувствительна к возрасту, и при написании алгоритмов подбора пар пользователей следует учитывать жесткие рамки. См. Hufford, Austen, “Chinese Dating Site Jiayuan Agrees to Be Bought by Baihe”, *Wall Street Journal*, December 7, 2015, <http://www.wsj.com/articles/chinese-dating-site-jiayuan-agrees-to-be-bought-by-baihe-1449501088>.

сообщения и описательную статистику (возрастной диапазон, этническая принадлежность, телосложение и т. п.) по людям, которыми действительно интересовался данный пользователь (а не только то, что он описывает в своем профайле в качестве интересующих его черт). Такая прозрачность позволяет пользователям лучше понимать не только имеющиеся варианты, но и свои шансы. Если интересующий вас персонаж отвечает только на 12 процентов сообщений, то, наверное, лучше потратить время на знакомство с кем-то еще, тем более что 64 процента интересующих его пользователей сообщают о своей «накачанности», а это совсем не про вас. Для получения большей достоверности о вкусах пользователей Jack's опирается не на статистику просмотров или входящих и исходящих сообщений, а исключительно на анализ списка «Избранных», а также специальной функции, позволяющей пользователям выражать свою заинтересованность в знакомстве. Эта опция направляет информацию в базу данных только в случае взаимного интереса.

Самая богатая и более персонализированная по сравнению с этими прямыми сигналами информация, которой обладает любой сайт знакомств, содержится в истории просмотров профайлов каждым пользователем. Однако интерпретировать мотивы этих просмотров достаточно сложно. Работая с сайтом Match.com, я обнаружил пользователя, который заблокировал доступ к себе огромному количеству черных женщин. Скорее всего, это был расист, правильно? Неправильно! Посмотрев на установки его фильтра и его просмотры, мы убедились в том, что все обстоит ровно наоборот: он интересовался именно черными женщинами, причем только теми, кто описывал себя «фигуристыми». Для экономии времени и сил он блокировал пользователей-женщин, попытки познакомиться с которыми оказались неудачными. Информационным детективам приходится решать и такие забавные задачки. А способность обнаруживать интересные случаи и рассказывать о них – важная составляющая умения понимать информацию.

Чтобы правильно трактовать данные, надо встать на точку зрения пользователя. Здесь очень важна обстановка. По ходу дня наши желания меняются, или же это происходит по ходу ночи. Когда я консультировал сингапурский сайт знакомств Fridae, мы заметили, что типы профайлов, просматриваемых в два часа дня пятницы, отличаются от просматриваемых в два часа ночи субботы. В этой связи научной команде Fridae предстояло решить, как использовать эту информацию в рейтингах, представленных на сайте.

Сайты знакомств все чаще предлагают пользователям возможность раскрыть свою «истинную сущность», стимулируя их добавлять в профайл кнопки ссылок на аккаунты в Facebook, Instagram или Twitter. Но это не значит, что случаи предосудительного поведения пользователей прекратились. Себастиан Боэр, один из специалистов по обработке данных мобильной платформы знакомств Skout, написал алгоритм фильтрации неприемлемых сообщений, под неформальным названием «гадоистребитель»<sup>127</sup>. Что считалось неприемлемым? Это определялось по кликам и взаимодействию пользователей. Если огромное количество пользователей блокировали кого-то определенного, то с большой долей вероятности можно было судить о том, что он – неприятный тип. А тот, кто постоянно слал безответные сообщения конкретному пользователю, явно был неприятен последнему. Со временем алгоритм научился распознавать содержание сообщений, которые обычно остаются без ответа или приводят к блокировке отправителя. Обычно в них присутствовал негатив. Очевидным признаком могли бы служить неприличные выражения, однако определение неприемлемости – значительно более тонкое дело. То, что кажется отвратительным одному человеку, может выглядеть привлекательным для другого. При выявлении закономерностей в блокировках «гадоистребитель» бло-

---

<sup>127</sup> Я был членом совета директоров Skout с 2007 по 2012 год. В июне 2016 года компания приобрела MeetMe. См. Yeung, Ken, “MeetMe Acquires Mobile Flirting App Skout for \$55 Million in Cash and Stock”, VentureBeat, June 27, 2016, <http://venturebeat.com/2016/06/27/meetme-acquires-mobile-flirting-app-skout-for-55-million-in-cash-and-stock>.



кировал доставку сообщений. Кроме этого алгоритм пресекал избыточную отправку безответных сообщений какому-то конкретному пользователю. Задачей «гадоистребителя» было поддержание позитивной среды для большинства пользователей.

В начале этой главы я писал о том, как мое образование и опыт физика помогли мне в создании, проведении и анализе результатов экспериментов с социальными данными. Многие из этих экспериментов подразумевают наблюдение за тем, как изменения в порядке переработки данных влияют на изменения в поведении людей. Если пользователь приложения знакомств увидит, что объект его интереса редко отвечает на сообщения, будет ли он тратить время на написание душевного послания или переключит внимание на кого-то с большей вероятностью отклика? Что скорее заставит негодя прекратить рассылку своих писем – просьба администратора или нулевая реакция пользователей? Где проходит тонкая грань экспериментов с идентичностью, переступив которую человек начинает выглядеть в глазах окружающих мошенником? Большая прозрачность поведения пользователей позволяет людям самостоятельно решать, насколько представленный в профайле персонаж соответствует их идеалу родственной души.

### **Призыв к открытости**

Когда дело касается приватности и открытости, люди всегда требуют первого для себя, а второго – для всех остальных<sup>128</sup>.

*Дэвид Брин*

Чем чаще люди заходят в интернет при помощи мобильных устройств, тем более важную роль в формировании идентичности играет еще один псевдоним – ваш телефонный номер. Когда домашние телефоны только появились, соединение осуществлял оператор, который представлял звонящего и спрашивал у абонента, будет ли тот разговаривать. С изобретением дискового импульсного набора и автоматических телефонных станций люди стали звонить друг другу напрямую и, чтобы выяснить, кто звонит, приходилось снимать трубку, подтверждая тем самым возможность разговора. Однако, в связи с тем что стоимость телефонной связи оставалась очень высокой, нежелательные звонки людям домой случались очень редко. Затем, когда цены начали падать, получил распространение ставший финансово целесообразным телефонный маркетинг. Еще позднее, в 1990-х, примерно в одно время с изобретением интернета появилась система тонального набора с возможностью определения номера входящего звонка.

Изначально идея о том, что телефонный номер, а возможно, и имя звонящего станут автоматически сообщаться адресату звонка, встретила определенный отпор. Однако сейчас ситуация прямо обратная: люди не склонны снимать трубку, когда им звонят со скрытого номера. Такие звонки отправляются прямиком на автоответчик. Чтобы принять звонок, люди должны знать, кто им звонит. Притом что вам как будто спокойнее, если ваш номер неизвестен, вам так же спокойнее, когда вы знаете номера тех, кто вам звонит. Но коммуникация эффективнее, когда она симметрична, то есть когда обе стороны могут идентифицировать друг друга.

Создатель интернет-справочника телефонных номеров *Whiterpages* Алекс Альгард считает, что телефонным коммуникациям можно придать большую прозрачность, которая пойдет на благо всем пользователям. Благодаря огромной базе данных телефонных номеров сервис *Niya* (ранее известный как *Whiterpages Caller ID*) идентифицирует все входящие звонки вне зависимости от настроек звонящего или телефонной книги вашего мобильного. В эпоху все

---

<sup>128</sup> Brin, David, “Questions I Am Frequently Asked About (Part V): Transparency, Privacy and the Information Age”, *Contrary Brin*, April 10, 2013, <http://davidbrin.blogspot.co.uk/2013/04/questions-i-am-frequently-asked-about.html>. Reprinted with permission of the author.

увеличивающегося объема телефонного «спама» это особенно полезная услуга. Niya помещает телефонный номер в какую-то категорию, например «телемаркетинг», извлекая информацию из источников в интернете и анализируя закономерности звонков с него своим подписчикам. Нам как обществу предстоит решить, обладают ли обе стороны телефонного разговора правом идентификации собеседника. При положительном ответе на этот вопрос нужно будет решить, что позволительно делать с такой информацией.

Это довольно запутанный вопрос, поскольку удостоверенная личность – необходимое, но недостаточное условие для возникновения взаимного доверия. Рассекречивание звонящей персоны является, скорее, инструментом, позволяющим призвать человека к ответственности в случае ненадлежащего поведения. Я знаю пару человек, запостивших в Facebook скриншоты того, что они считали неприемлемым поведением на сайте знакомств. В одном случае это был ответ «Спасибо, мне не интересно», в другом имело место намеренное оскорбление. Получатели вполне могли ограничиться нажатием кнопки «заблокировать», но решили поделиться примерами неправильного поведения со своими знакомыми.

На что можно рассчитывать с точки зрения приватности в коммуникациях в наши дни? Вышеупомянутые получатели нежелательных сообщений на сайте знакомств посчитали, что обнародовать эти частные сообщения было в интересах сообщества их друзей. Скриншоты служили предупреждением для других людей, которые могут пользоваться этим сайтом знакомств. Поскольку на скриншотах не были затушеваны ни имена, ни фото виновников, они стали известны всем друзьям получателей в Facebook. Схожим образом, в случае если ваш босс несправедливо обругает вас в электронном письме, вы можете переслать это сообщение своим знакомым или выложить в интернете. С точки зрения закона это может быть сочтено нарушением, поскольку письмо было конфиденциальным и предназначалось только для служебного пользования. Но в то же время распространение такого письма послужит интересам общества, поскольку потенциальные сотрудники будут лучше представлять себе условия работы в компании.

Реакция на публикацию частной переписки в некоторой степени обусловлена тем, насколько мы доверяем тому, кто это делает. Любой скриншот очень легко подделать. На форумах вроде Reddit, где основная масса пользователей анонимна, верифицировать личность автора поста маловероятно, не говоря уже о самих сообщениях. Авторы постов в Facebook обычно входят в число наших знакомых (или знакомых знакомых). А поскольку аккаунты взламывают очень редко, желание человека запостить скриншот сдерживается мыслью о том, что друзья будут считать его способным предать огласке частную переписку. При этом мы не обязаны считать скриншот настоящим. Может быть, его подделали с целью дискредитации.

После того как скриншот, настоящий или фальшивый, размещен, он становится обычным элементом информации. Любой, кому он попадается на глаза, может распространять его дальше. Что, если пост так возмутит или развеселит одного из фейсбучных френдов, что он решит расшерить скриншот своим френдам? А френд этого френда решит его твитнуть? Вы не успеете моргнуть глазом, как программа распознавания лиц идентифицирует человека и подпишет фото его именем. К тому моменту все уже будет происходить вне изначального контекста, включая и личность автора поста, но эта информация о человеке станет доступна любому интересующемуся.

Как человек сможет защищать свой образ в интернете в будущем? Один из вариантов был предложен Европейским судом, который в мае 2014 года вынес решение в пользу «права на забвение». Один испанец устал от того, что потенциальные работодатели и домовладельцы обнаруживают старую газетную статью о том, как у него конфисковали дом за налоговую задолженность, тем более что он уже рассчитался по своим долгам<sup>129</sup>. Он не требовал удалить статью

---

<sup>129</sup> “The Man Who Sued Google to Be Forgotten”, Reuters, May 30, 2014, <http://www.newsweek.com/man-who-sued-google>

о конфискации. Он просто хотел, чтобы эту страницу не показывали людям, которые гуглят его имя. Суд постановил, что люди имеют право на удаление из результатов поиска тех страниц, которые, по их мнению, наносят им ущерб. В первый же день после вступления в силу решения Европейского суда Google получила более 12 000 запросов на удаление, а в течение первого года к ним добавились еще более 275 000<sup>130</sup>.

Google обнародовала некоторые из запросов на удаление ссылок, полученных после судебной победы испанца. Итальянка попросила, чтобы в результатах поиска по ее имени не появлялась статья более чем десятилетней давности об убийстве ее мужа. Латвийский активист, пострадавший во время акции протеста, просил исключить из результатов поиска по его имени статью об этой акции. «Осужденный за преступление небольшой тяжести» более десяти лет назад немецкий учитель тоже просил изъять из результатов поиска сообщение о своем судебном процессе. В Google решили, что право на забвение в отношении отдельного лица важнее, чем «общественная значимость предмета»<sup>131</sup>. Все эти запросы восхитительны сами по себе, но должны ли определять общественную значимость предмета алгоритмы Google (или ее юристы)?

В 1890-х годах, ратуя за «право на неприкосновенность частной жизни», звезды юриспруденции Сэмюел Уоррен и Луис Брэндис особенно акцентировали фундаментальное право собственности частного лица на свою личность<sup>132</sup>. Кто захочет, чтобы без его ведома на всеобщее обозрение попало компрометирующее фото? Идея заключалась в том, чтобы силой закона заставить людей относиться друг к другу по-человечески. Законодательно закрепленное «право на неприкосновенность частной жизни» должно было охранять достоинство, а не способствовать развитию свободного общества<sup>133</sup>. В то время считалось, что неограниченные свободы ведут к тирании масс. Свобода была плохой штукой.

---

be-forgotten-252854.

<sup>130</sup> По состоянию на август 2016 года Google получила 520 000 запросов на удаление, что соответствовало 1,6 миллионам URL-адресов. See Google, “Transparency Report: European Privacy in Search”, August 8, 2016, <https://www.google.com/transparencereport/removals/europeprivacy>.

<sup>131</sup> Эти запросы – из числа двадцати трех, особо выделенных Google как имеющих очевидную причину для удовлетворения или отказа. См. Google, “Transparency Report: European Privacy in Search”, August 1, 2016, <http://www.google.com/transparencereport/removals/europeprivacy>.

<sup>132</sup> Хотя ведущим философом того времени был Иммануил Кант, теорию «свободы, основанной на самостоятельности, воле и личности человека» в германском праве развивал (под влиянием Канта) Фридрих Карл фон Савиньи. См. See Eberle, Edward J., “The German Idea of Freedom”, *Oregon Review of International Law* 10, no. 1 (2008), p. 16, [http://docs.rwu.edu/law\\_fac\\_fs/56/](http://docs.rwu.edu/law_fac_fs/56/).

<sup>133</sup> Bloustein, Edward J., “Privacy as an Aspect of Human Dignity: An Answer to Dean Prosser”, *New York University Law Review* 39, no. 962 (December 1964), pp. 962–1007, <http://heinonline.org/HOL/LandingPage?collection=journals&handle=hein.journals/nylr39&div=71>

## **Конец ознакомительного фрагмента.**

Текст предоставлен ООО «ЛитРес».

Прочитайте эту книгу целиком, [купив полную легальную версию](#) на ЛитРес.

Безопасно оплатить книгу можно банковской картой Visa, MasterCard, Maestro, со счета мобильного телефона, с платежного терминала, в салоне МТС или Связной, через PayPal, WebMoney, Яндекс.Деньги, QIWI Кошелек, бонусными картами или другим удобным Вам способом.