

Psycho-Informatics: Big Data shaping modern psychometrics	Психоинформатика: большие данные формируют современную психометрию
<p>Alexander Markowetz ^a, Konrad Błazzkiewicz ^a, Christian Montag ^b, Christina Switala ^c, Thomas E. Schlaepfer ^{c, †}</p> <p>^a <i>Department of Computer Science, University of Bonn, Germany</i> ^b <i>Department of Psychology, University of Bonn, Germany</i> ^c <i>Department of Psychiatry and Psychotherapy, University of Bonn, Germany</i></p>	
a b s t r a c t	Аннотация
<p>For the first time in history, it is possible to study human behavior on great scale and in fine detail simultaneously. Online services and ubiquitous computational devices, such as smartphones and modern cars, record our everyday activity. The resulting Big Data offers unprecedented opportunities for tracking and analyzing behavior. This paper hypothesizes the applicability and impact of Big Data technologies in the context of psychometrics both for research and clinical applications. It first outlines the state of the art, including the severe shortcomings with respect to quality and quantity of the resulting data. It then presents a technological vision, comprised of (i) numerous data sources such as mobile devices and sensors, (ii) a central data store, and (iii) an analytical platform, employing techniques from data mining and machine learning. To further illustrate the dramatic benefits of the proposed methodologies, the paper then outlines two current projects, logging and analyzing smartphone usage. One such study attempts to thereby quantify severity of major depression dynamically; the other investigates (mobile) Internet Addiction. Finally, the paper addresses some of the ethical issues inherent to Big Data technologies. In summary, the proposed approach is about to induce the single biggest methodological shift since the beginning of psychology or psychiatry. The resulting range of applications will dramatically shape the daily</p>	<p>Впервые в истории можно изучать поведение человека в больших масштабах и в мельчайших деталях одновременно. Онлайн-сервисы и вездесущие вычислительные устройства, такие как смартфоны и современные автомобили, фиксируют нашу повседневную деятельность. Получающиеся большие данные предлагают беспрецедентные возможности для отслеживания и анализа поведения. В этой статье выдвигается гипотеза о применимости и влиянии технологий больших данных в контексте психометрии как для исследований, так и для клинических применений. В нем в общих чертах описывается уровень техники, в том числе серьезные недостатки в отношении качества и количества полученных данных. Затем он представляет технологическое видение, состоящее из (i) многочисленных источников данных, таких как мобильные устройства и датчики, (ii) центрального хранилища данных и (iii) аналитической платформы, использующей методы интеллектуального анализа данных и машинного обучения. Чтобы дополнительно проиллюстрировать драматические преимущества предлагаемых методологий, в документе затем намечены два текущих проекта: ведение журнала и анализ использования смартфона. В одном из таких исследований делается попытка динамически количественно оценить тяжесть тяжелой депрессии; другой исследует (мобильную) интернет-зависимость. Наконец, в документе рассматриваются некоторые этические проблемы,</p>

<p>routines of researches and medical practitioners alike. Indeed, transferring techniques from computer science to psychiatry and psychology is about to establish Psycho-Informatics, an entire research direction of its own.</p>	<p>присущие технологиям больших данных. Таким образом, предлагаемый подход собирается вызвать самый большой методологический сдвиг с начала психологии или психиатрии. Получившийся диапазон применений кардинально определит повседневную рутину исследований и практикующих врачей. Действительно, перевод методов из компьютерной науки в психиатрию и психологию вот-вот создаст психоинформатику, целое направление исследований.</p>
<p>Introduction</p>	<p>Вступление</p>
<p>Reliable measurements of emotion, cognition and behavior are of equal and central importance to psychiatry and psychology. Despite the crucial role of these parameters, the basic methodology has remained essentially unchanged for the better half of a century. To this day, researchers rely on clinicians' observations and self-rated psychometric tests. Requiring specially trained experts for conducting interviews or observations, these methods found widespread use in research, but failed to penetrate clinical practice. And, despite the highly skilled evaluators (and associated cost), the gathered data remains rather unsatisfactory, both in terms of quantity <i>and</i> quality. As outlined below, these shortcomings are inherent to the methodologies themselves, and cannot be overcome by enlarging financial budgets. At the same time, there appears to be no immediate successor. On the one side, novel and scientifically interesting methods like neuroimaging and genetics represent promising approaches to evaluate the course of psychiatric treatment [1]. On the other side, these approaches are still in an early stage of development, costly and often do not provide reliable diagnostics for individual patients.</p>	<p>Надежные измерения эмоций, познания и поведения одинаково важны для психиатрии и психологии. Несмотря на решающую роль этих параметров, базовая методология практически не изменилась в течение большей половины столетия. По сей день исследователи полагаются на наблюдения клиницистов и психометрические тесты с самооценкой. Требуя специально обученных экспертов для проведения интервью или наблюдений, эти методы нашли широкое применение в исследованиях, но не смогли проникнуть в клиническую практику. И, несмотря на высококвалифицированных оценщиков (и связанные с этим расходы), собранные данные остаются довольно неудовлетворительными, как с точки зрения количества, так и качества. Как указано ниже, эти недостатки присущи самим методологиям и не могут быть преодолены путем увеличения финансовых бюджетов. В то же время, похоже, нет непосредственного преемника. С одной стороны, новые и научно интересные методы, такие как нейровизуализация и генетика, представляют многообещающие подходы к оценке курса психиатрического лечения [1]. С другой стороны, эти подходы все еще находятся на ранней стадии разработки, являются дорогостоящими и часто не обеспечивают надежной диагностики для отдельных пациентов.</p>
<p>The temporal granularity at which traditional methods collect data commonly is too coarse to reveal fine-granular patterns. The most important drawbacks are of entirely practical nature: around-the-clock shadowing is neither affordable nor acceptable to a participant. Instead,</p>	<p>Временная гранулярность, при которой традиционные методы обычно собирают данные, является слишком грубой, чтобы выявить мелкие гранулярные паттерны. Наиболее важные недостатки носят исключительно практический характер: круглосуточное теневое</p>

<p>researchers employ questionnaires at fixed intervals, essentially relying on the participants' self-report. This method naturally imposes a bound on the temporal granularity at which participants can be interviewed: weekly, monthly, or at an even coarser level. Furthermore, it is not possible to execute the identical psychometric test multiple times over the course of a single day. Memory and training effects would limit the reliability of the ratings. Next, holding interviews at high frequency would be prohibitively expensive, because such interviews have to be conducted by a trained professional. Depending on the level of training and employment status, a single interview quickly costs several hundred Euros. In addition, the necessary appointments impose too great of a burden on the participant, especially when the content of the interviews only relies to negative aspects of life such as psychopathological disorders. Including travel, a single interview can consume the better half of a day, a burden that only be imposed infrequently (in particular with participants pursuing a professional career). Self-reports in form of diaries do not provide a viable solution either. This method, too, quickly meets a limit of how much time commitment can be expected from a participant. In sum, the constraints (i) reduce the temporal granularity at which data can be gathered, and (ii) pose tremendous problems for longitudinal studies with respect to the amount and completeness of data gathered over an extended time range.</p>	<p>копирование не является ни доступным, ни приемлемым для участника. Вместо этого исследователи используют вопросники с фиксированными интервалами, в основном полагаясь на самоотчеты участников. Этот метод, естественно, накладывает ограничение на временную гранулярность, при которой участники могут брать интервью: еженедельно, ежемесячно или даже на более грубом уровне. Кроме того, невозможно выполнить идентичный психометрический тест несколько раз в течение одного дня. Память и тренировочные эффекты ограничивают надежность рейтингов. Далее, проведение интервью с высокой частотой было бы чрезмерно дорогостоящим, потому что такие интервью должны проводиться обученным профессионалом. В зависимости от уровня подготовки и статуса занятости одно интервью быстро стоит несколько сотен евро. Кроме того, необходимые встречи накладывают на участника слишком большое бремя, особенно когда содержание интервью касается только негативных аспектов жизни, таких как психопатологические расстройства. Включая поездки, одно собеседование может занять лучшие полдня, но это бремя, которое накладывается нечасто (в частности, на участников, занимающихся профессиональной карьерой). Самостоятельные отчеты в форме дневников также не обеспечивают жизнеспособного решения. Этот метод также быстро достигает предела того, сколько времени можно ожидать от участника. В итоге, ограничения (i) уменьшают временную гранулярность, при которой могут быть собраны данные, и (ii) создают огромные проблемы для продольных исследований в отношении количества и полноты данных, собранных в течение расширенного временного диапазона.</p>
<p>Unfortunately, data collected by the traditional means is also strongly biased. Most notably, it is commonly faulty and distorted, due to poor recollection of the variable of interest. This holds especially for coarse intervals of reporting, and especially for questions regarding interaction with digital devices. Very few people could accurately report how often they have checked their email over the past 10 days (which would be an interest variable to study Internet use/addiction). Additionally, reports</p>	<p>К сожалению, данные, собранные традиционными способами, также сильно предвзяты. В частности, он обычно ошибочен и искажен из-за плохой памяти переменной, представляющей интерес. Это особенно актуально для грубых интервалов отчетности, особенно для вопросов, касающихся взаимодействия с цифровыми устройствами. Очень немногие люди могут точно сообщить, как часто они проверяли свою электронную почту за последние 10 дней</p>

<p>about variables from other research areas, such as subjective well-being, tend to simply reflect altered psychological states. In particular, it has been shown that people use their momentary affective state for judging how happy and satisfied they are with their lives in general. A depressed patient for example will usually see his/her well-being, social functioning, and living conditions worse than they would appear to an independent observer, or to himself/herself after recovery [2]. Thus, self-reports are affected by the state of mind at time of reporting, and the social desirability of the reported behavior. Together, these factors introduce significant noise to infrequently recorded data. Clinician-rated psychometric tests, entail the risk of being similarly biased, since assessments of experts are not entirely objective. In this context, the term “objective” assessment is misleading and should be replaced by “external”, as this evaluation might reflect the subjective view of the assessor himself [2].</p>	<p>(что было бы интересной переменной для изучения использования / зависимости в Интернете). Кроме того, сообщения о переменных из других областей исследований, таких как субъективное благополучие, имеют тенденцию просто отражать измененные психологические состояния. В частности, было показано, что люди используют свое мгновенное аффективное состояние для оценки того, насколько они счастливы и удовлетворены своей жизнью в целом. Например, пациент с депрессией обычно видит свое благополучие, социальное функционирование и условия жизни хуже, чем он мог бы показаться независимому наблюдателю или самому себе после выздоровления [2]. Таким образом, на самоотчеты влияют состояние души на момент сообщения и социальная желательность сообщаемого поведения. Вместе эти факторы вносят значительный шум в редко записываемые данные. Оцененные клиницистами психометрические тесты влекут за собой риск того же предвзятого отношения, поскольку оценки экспертов не являются полностью объективными. В этом контексте термин «объективная» оценка ошибочна и должна быть заменена на «внешнюю», поскольку эта оценка может отражать субъективную точку зрения самого оценщика [2].</p>
<p>In short, data gathered by traditional means thus capture the situation of a study’s participant or patient rather poorly. It is too coarse to show temporal patterns, and generally lacks dynamics. Additionally, it commonly employs shallow scales, thus quickly encounter floor effects. Most questionnaires regarding depression, for example, only permit answers on each item on a scale of 0–3 [3–5]. The effects of these coarse measurements are dramatic, because novel psychotropic substances frequently become stuck during the development phase, because (visible) positive effects cannot be quantified reliably [6,7]. Clearly, innovation in methodology has long been overdue.</p>	<p>Короче говоря, данные, собранные традиционными способами, таким образом, довольно плохо отражают ситуацию участника исследования или пациента. Это слишком грубый, чтобы показать временные модели, и, как правило, не хватает динамики. Кроме того, он обычно использует мелкие чешуйки, поэтому быстро сталкивается с полом. Например, большинство вопросников о депрессии разрешают только ответы по каждому пункту по шкале от 0 до 3 [3–5]. Эффекты этих грубых измерений являются существенными, потому что новые психотропные вещества часто застревают во время фазы развития, потому что (видимые) положительные эффекты не могут быть надежно определены количественно [6,7]. Очевидно, что инновации в методологии давно назрели.</p>
<p>As early as the seventies, researchers circulated the idea of <i>actigraphy</i> as</p>	<p>Еще в семидесятых годах исследователи распространили идею</p>

<p>a simple and non-invasive method for monitoring human rest and activity cycles. Inter alia, they measured sleep patterns [8,9] and circadian rhythms [10] via specific <i>actimetry sensors</i>, worn on the body of the patient. While this approach overcame some of the obstacles faced by questionnaires, it did not quite hit the mark. Early technology was rather simple, rendering sensors complex, expensive and socially awkward, thus requiring substantial compliance and discipline from the patient. In some areas of research such as neuropharmacology, actigraphy only was administered in very few cases [11]. In recent years, miniaturization of digital devices has given new rise to the methodology. Sensors have become smaller, less power-hungry, and can independently transmit their data. While finally practical, the central obstacle to actigraphy remains: the patient/participant must be coaxed into carrying a sensor for a substantial period of time. In sum, actigraphy has only been used sporadically in most areas of psychiatry and psychology. Due to miniaturization, it is about to enjoy a second lease of life, but will ultimately be made redundant by sensor-less methods of tracking.</p>	<p>актиграфии как простого и неинвазивного метода мониторинга циклов отдыха и активности человека. В частности, они измеряли характер сна [8,9] и циркадные ритмы [10] с помощью специальных датчиков актиметрии, которые носили на теле пациента. Хотя этот подход преодолел некоторые из препятствий, с которыми сталкиваются вопросники, он не совсем достиг цели. Ранние технологии были довольно простыми, что делало сенсоры сложными, дорогими и социально неудобными, что требовало от пациента значительного соответствия и дисциплины. В некоторых областях исследований, таких как нейрофармакология, актиграфия применялась только в очень немногих случаях [11]. В последние годы миниатюризация цифровых устройств породила новую методологию. Датчики стали меньше, потребляют меньше энергии и могут независимо передавать свои данные. Хотя, в конечном счете, это практично, главное препятствие для деятельности остается: пациент / участник должен быть вынужден носить датчик в течение значительного периода времени. В целом, актиография использовалась спорадически только в большинстве областей психиатрии и психологии. Из-за миниатюризации, эта технология может получить вторую жизнь, но в конечном итоге она станет ненужной из-за бессенсорных методов отслеживания.</p>
<p>In this paper, we propose observing behavior directly on digital devices and services, such as laptops, social networks, or even cars. Specifically, we focus on user interaction with smartphones. Carried on the person, around the clock, and used for a wide range of (informal) communication, these devices constitute a particularly rich and intimate source of information. The gathered data is of highest quality, gathered entirely in the background, and automatically forwarded to a central server. The method thus burdens neither patient/participant nor researcher. Most importantly, avoids the dominant sources of bias, commonly encountered by self-reports and questionnaires.</p>	<p>В этой статье мы предлагаем наблюдать за поведением непосредственно на цифровых устройствах и сервисах, таких как ноутбуки, социальные сети или даже автомобили. В частности, мы ориентируемся на взаимодействие пользователей со смартфонами. Эти устройства, которые круглосуточно используются человеком и используются для широкого (неформального) общения, представляют собой особенно богатый и интимный источник информации. Собранные данные имеют высокое качество, собраны полностью в фоновом режиме и автоматически передаются на центральный сервер. Таким образом, метод не обременяет ни пациента / участника, ни исследователя. Что наиболее важно, избегает доминирующих источников предвзятости, обычно встречающихся в самоотчетах и анкетах.</p>

<p>For several areas of research, the proposed methodologies constitute the only viable solution. Most notably, it constitutes the only valid measure for usage and abuse of digital media. Kimberly Young [12] first saw a problem for the human condition when excessively using the Internet, an issue also put forward for the usage of mobile phones [13]. Whether the observed phenomena constitute a ‘new disorder’ is a matter of heated debate [14]. Although excessive use of the Internet is not a distinct disorder in the DSM-V, evidence from both psychology, psychiatry and the neurosciences suggest that “Internet addiction” constitutes a substantial challenge [15,16]. While a high daily “dosage” does not qualify for an addiction, a rising number of hours spent with the phone over a certain time could indicate developing tolerance. In any case, such behavior must be recorded directly on the device. Ordinary patients/participants cannot be expected to accurately answer how often they unlock their phone each day (up to 200 times, according to our preliminary experimental findings). The particularly poor recollection in this context arises due to the “virtual” character of phone behavior. Alcohol consumption for example, is significantly easier to quantify, if only by the number of empty bottles.</p>	<p>Для нескольких областей исследований предлагаемые методологии представляют собой единственное жизнеспособное решение. В частности, он представляет собой единственную действительную меру для использования и злоупотребления цифровыми медиа. Кимберли Янг [12] впервые увидела проблему состояния человека при чрезмерном использовании Интернета, а также проблему использования мобильных телефонов [13]. Являются ли наблюдаемые явления «новым беспорядком» вопрос горячих споров [14]. Хотя чрезмерное использование Интернета не является явным расстройством в DSM-V, данные психологии, психиатрии и нейробиологии свидетельствуют о том, что «интернет-зависимость» представляет собой серьезную проблему [15,16]. В то время как высокая ежедневная «дозировка» не подпадает под зависимость, увеличение количества часов, проведенных с телефоном в течение определенного времени, может указывать на развитие толерантности. В любом случае, такое поведение должно быть записано непосредственно на устройстве. Нельзя ожидать, что обычные пациенты / участники точно ответят, как часто они разблокируют свой телефон каждый день (до 200 раз, согласно нашим предварительным экспериментальным данным). Особенно плохое воспоминание в этом контексте возникает из-за «виртуального» характера поведения телефона. Например, потребление алкоголя значительно легче измерить, хотя бы по количеству пустых бутылок.</p>
<p>The proposed methodology is about to equally revolutionize the work of researchers with more classic research agendas, such as personality or behavior. Recently, Kosinski et al. impressively inferred personality traits from the behavior on the Internet platform Facebook [17]. Yet, such research endeavors only mark the beginning of tight collaboration between psychology/psychiatry and informatics. After all, Facebook usage ‘only’ represents a rather narrow glimpse on people’s lives. By comparison, how much can we learn about the human condition when monitoring mobile phones 24 h/7 days a week? The socially outgoing (extraverted) person could easily be detected by the amount of inand</p>	<p>Предложенная методология собирается в равной степени революционизировать работу исследователей с более классическими программами исследований, такими как личность или поведение. Недавно Kosinski et al. Впечатляюще выявляются личностные черты поведения на интернет-платформе Facebook [17]. Тем не менее, такие исследования только отмечают начало тесного сотрудничества между психологией / психиатрией и информатикой. В конце концов, использование Facebook «только» представляет собой довольно узкий взгляд на жизнь людей. Для сравнения, сколько мы можем узнать о состоянии человека при мониторинге мобильных телефонов</p>

<p>outcoming calls, indicating a large active social network. The introverted person in contrast might display longer reading sessions, perhaps using an e-book application. The person being open for new experiences (another of the Big Five Factors of Personality describing human characteristics by McCrae and John [18]) might often install and test new apps. Numerous such dependent variables can be detected by observing humans through their mobile phone interactions. These measures will capture the human condition more precise than ever. For the first time, psychiatrists and psychologists can observe human behavior on a large scale, in the finest temporal granularity. They can thus assess the course of treatment and disease in a temporal continuum, instead of relying on selective snapshots.</p>	<p>24 часа в сутки 7 дней в неделю? Социально исходящего (экстравертированного) человека можно легко обнаружить по количеству входящих и исходящих вызовов, что указывает на большую активную социальную сеть. Напротив, интроверт может показывать более длительные сеансы чтения, возможно, с помощью приложения для электронных книг. Человек, открытый для новых впечатлений (еще один из «Большой пятерки факторов личности», описывающих человеческие качества МакКрей и Джона [18]), часто может устанавливать и тестировать новые приложения. Множество таких зависимых переменных можно обнаружить, наблюдая за людьми через их взаимодействие с мобильным телефоном. Эти меры позволят зафиксировать состояние человека более точно, чем когда-либо. Впервые психиатры и психологи могут наблюдать за поведением человека в больших масштабах, с лучшей временной детализацией. Таким образом, они могут оценить ход лечения и болезни во временном континууме, вместо того чтобы полагаться на выборочные снимки.</p>
<p>Equally, the proposed methodology is about to revolutionize clinical therapy, a role in which it will affect our everyday lives to an even higher degree. In this scenario, patients track a wide range of personal data, from phones, cars, and fridges. From this raw (and rather cryptic) data, large-scale analysis extracts meaningful indices, such as an “activity index”, or a “social interaction index”. The patient can then self-track his condition. He is reassured that it is not worsening. Or, if a worsening of his health condition occurs, he could confidently ask for an ad-hoc appointment with his doctor. In addition, he can explore interdependencies between his health condition and his lifestyle, such as staying up late, or working out. Most importantly, he can provide (selected) data access to his coach, therapist or doctor.</p>	<p>Точно так же предложенная методология собирается революционизировать клиническую терапию, роль которой будет влиять на нашу повседневную жизнь в еще большей степени. В этом сценарии пациенты отслеживают широкий спектр персональных данных, начиная с телефонов, автомобилей и холодильников. Из этих необработанных (и довольно загадочных) данных крупномасштабный анализ извлекает значимые индексы, такие как «индекс активности» или «индекс социального взаимодействия». Затем пациент может самостоятельно отслеживать свое состояние. Он уверен, что это не ухудшается. Или, если происходит ухудшение состояния его здоровья, он может спокойно попросить о специальной встрече со своим врачом. Кроме того, он может исследовать взаимозависимости между состоянием своего здоровья и образом жизни, например, лечь спать допоздна или заниматься спортом. Самое главное, он может предоставить (выбранный) доступ к данным своему тренеру, терапевту или врачу.</p>

<p>For the clinician, this methodology enables an entire range of new options. For the first time, he does not have to rely on the (poor) self-report of his patient. Instead, he receives clear indicators of the patient's mental state, and changes therein, in a finegranular temporal resolution. He can thus observe the continuous changes of health parameters over time (to follow the course of a disease, or the progress of therapy). The clinician will also be able to investigate changes of his patient throughout the day, and finetune timing and dosage of medication, providing a highly individualized therapy. For example, he thus could match medication doses in a patient suffering from schizophrenia. The clinician can even prescribe a range of dosage, from which the patient can independently choose, according to his or her latest data. The therapist can be automatically alarmed when symptom data indicates a critical situation. In this case, he can intervene via phone, video conference, or an ad-hoc appointment. At the same time, regular appointments can be spaced further apart.</p>	<p>Для клинициста эта методология предоставляет целый ряд новых возможностей. Впервые ему не нужно полагаться на (плохое) самоотчет своего пациента. Вместо этого он получает четкие показатели психического состояния пациента и его изменения в точном временном разрешении. Таким образом, он может наблюдать постоянные изменения параметров здоровья с течением времени (следить за течением заболевания или ходом терапии). Клиницист также сможет исследовать изменения своего пациента в течение дня, а также точно настроить сроки и дозировку лекарств, предоставляя высоко индивидуализированную терапию. Например, он, таким образом, мог подбирать дозы лекарств у пациента, страдающего шизофренией. Врач может даже назначить диапазон дозировки, который пациент может выбирать самостоятельно, в соответствии с его или ее последними данными. Терапевт может автоматически встревожиться, когда данные о симптомах указывают на критическую ситуацию. В этом случае он может вмешаться по телефону, с помощью видеоконференции или специальной встречи. В то же время регулярные встречи могут быть расположены дальше друг от друга.</p>
<p>Most importantly, the proposed methodology is significantly cheaper than personal interaction with a therapist. This profane observation has vast implications, opening the application area towards wellness and prevention for large amounts of people. Currently, society focuses its limited therapeutic resources on sick patients. In the future, data driven early warning systems will enable us to help people a long time before their conditions becomes serious or chronic. Raising red flags early, some people might just need to attend a seminar on sustainable usage of digital media, or an extended vacation, or the HR department talk to their chaotic manager. Eventually, most corporations will deploy data driven preventive mental health programs. The ethical perspective (as discussed below) only constitutes a fraction of the challenges these services face. The integration into the processes and structures of large corporations might turn out far more difficult. Yet, occupational doctors can serve as a blueprint for a data driven occupational mental health service, leading to</p>	<p>Самое главное, предложенная методология значительно дешевле, чем личное общение с терапевтом. Это профанное наблюдение имеет огромные последствия, открывая область применения для здоровья и профилактики для большого количества людей. В настоящее время общество концентрирует свои ограниченные терапевтические ресурсы на больных пациентах. В будущем системы раннего предупреждения, основанные на данных, позволят нам помогать людям еще долгое время, прежде чем их состояние станет серьезным или хроническим. Рано поднимая красные флаги, некоторым людям может просто понадобится посетить семинар по устойчивому использованию цифровых медиа, или длительный отпуск, или отдел кадров поговорить со своим хаотичным менеджером. В конце концов, большинство корпораций развернут программы профилактического психического здоровья, основанные на данных. Этическая перспектива (как обсуждается ниже)</p>

their widespread deployment much sooner than anticipated.	составляет лишь часть проблем, с которыми сталкиваются эти службы. Интеграция в процессы и структуры крупных корпораций может оказаться гораздо более сложной. Тем не менее, профессиональные врачи могут служить образцом для службы гигиены труда, основанной на данных, что приведет к их широкому распространению гораздо раньше, чем ожидалось.
The remainder of this paper is structured as follows. Next, we outline the underlying technological vision, comprised of various data sources, and means to store and analyze the data. Subsequently, we introduce two current studies and respective hypotheses. One study tracks depression, the other investigates the misuse of mobile phones. We then touch upon the ethical aspects of the proposed methodology, a topic we feel very strongly about. The article ends with an outlook on the anticipated changes in research and therapy. As we outline, the proposed methodology will shape, if not revolutionize, psychiatry and psychology. The envisioned shift will be massive, touch every aspect of both sciences, and eventually create its own field of research: PsychoInformatics.	Остальная часть этой статьи структурирована следующим образом. Далее мы опишем основную технологическую концепцию, состоящую из различных источников данных, и средств для хранения и анализа данных. Затем мы представляем два текущих исследования и соответствующие гипотезы. Одно исследование отслеживает депрессию, другое расследует злоупотребление мобильными телефонами. Затем мы коснемся этических аспектов предлагаемой методологии, тема, к которой мы очень сильно относимся. Статья заканчивается обзором ожидаемых изменений в исследованиях и терапии. Как мы отмечаем, предлагаемая методология будет формировать, если не революционизировать, психиатрию и психологию. Предполагаемый сдвиг будет массовым, затронет каждый аспект обеих наук и в конечном итоге создаст свое собственное поле исследований: психоинформатику.
Underlying technological vision	Основное технологическое видение
This paper's is based on a single central thesis. The user's mental state, we claim, affects the way he interacts with a machine. A stressed user may thus generate more typographic errors than ordinarily; a depressed user may communicate less over his phone than previously. Conversely, so the claim continues, changes in his interaction with a machine reflect changes in his mental state. Modern computer science enables us to automatically gather the appropriate data, transfer, and analyze it, all at very little cost. The proposed methodology, so we hypothesize, might outperform traditional methods in both data quality as well as quantity.	Эта статья основана на одном центральном тезисе. Мы утверждаем, что психическое состояние пользователя влияет на то, как он взаимодействует с машиной. Таким образом, подверженный стрессу пользователь может генерировать больше типографских ошибок, чем обычно; подавленный пользователь может общаться по телефону меньше, чем раньше. И наоборот, так продолжает утверждение, изменения в его взаимодействии с машиной отражают изменения в его психическом состоянии. Современная компьютерная наука позволяет нам автоматически собирать соответствующие данные, передавать и анализировать их, причем при минимальных затратах. Мы предполагаем, что предложенная методология может превзойти традиционные методы как по качеству данных, так и по количеству.

<p>The twenty-first century will likely be remembered as the age of <i>Big Data</i>. Recent advancements in hard and software enable us to store and analyze massive amounts of data, at surprisingly little cost. Currently, such technology is most prominently employed by search engines, social networks, credit card issuers and insurance companies. These very different businesses gather massive amounts of data, e.g., to seek underlying patterns or assign scores to individual users. Based on these quantitative measures, they automatize decision-making processes, such as which advertisement to show, whether to approve a transaction, or to grant a credit line. In the context of psychology and psychiatry, we propose Big Data to produce psychometric parameters and to trace the course of a disorder.</p>	<p>Двадцать первый век, скорее всего, запомнится как век больших данных. Последние достижения в области аппаратного и программного обеспечения позволяют нам хранить и анализировать огромные объемы данных при удивительно небольших затратах. В настоящее время такие технологии наиболее широко используются поисковыми системами, социальными сетями, эмитентами кредитных карт и страховыми компаниями. Эти очень разные предприятия собирают огромные объемы данных, например, для поиска базовых шаблонов или присвоения баллов отдельным пользователям. На основе этих количественных показателей они автоматизируют процессы принятия решений, например, какую рекламу показывать, одобрить ли сделку или предоставить кредитную линию. В контексте психологии и психиатрии мы предлагаем большие данные для получения психометрических параметров и отслеживания развития расстройства.</p>
<p>Unfortunately, terms like “Big Data” or “data mining” are surrounded by a significant amount of buzz. Frequently fed by marketing departments from IT companies, the hype often obfuscates the actual potential and limits. Yet, as this section should illustrate, the potential of these technologies is indeed immense. To complicate the terminology further, there is a significant overlap between the areas and concepts. Terms such as “data mining” and “Big Data” are often used interchangeably, even by experts. We thus next clarify the core concepts, before outlining their applicability in the context of psychiatry and psychology.</p>	<p>К сожалению, такие термины, как «большие данные» или «интеллектуальный анализ данных», окружены значительным количеством ажиотажа. Часто рекламируемые отделами маркетинга из ИТ-компаний, реклама часто скрывает реальный потенциал и ограничения. Тем не менее, как следует из этого раздела, потенциал этих технологий действительно огромен. Чтобы еще больше усложнить терминологию, между областями и концепциями существует значительное совпадение. Такие термины, как «интеллектуальный анализ данных» и «большие данные», часто используются взаимозаменяемо, даже экспертами. Таким образом, мы затем проясним основные понятия, прежде чем обрисовать их применимость в контексте психиатрии и психологии.</p>
<p>Big Data applications commonly comprise a range of various complex components. Data is extracted, collected, cleaned and transformed, stored and managed, analyzed, indexed and searched, as well as visualized. Accordingly, these applications touch many areas of computer science, such as database systems, text retrieval, data mining, machine learning and data visualization. For the scope of this paper, we thus define Big</p>	<p>Приложения для больших данных обычно содержат ряд различных сложных компонентов. Данные извлекаются, собираются, очищаются и преобразуются, хранятся и управляются, анализируются, индексируются и ищутся, а также визуализируются. Соответственно, эти приложения затрагивают многие области компьютерной науки, такие как системы баз данных, поиск текста,</p>

<p>Data as the union over the range of tools and disciplines involved in collecting, storing, and analyzing large amounts of data originating from observing the interaction between users and devices (phones).</p>	<p>интеллектуальный анализ данных, машинное обучение и визуализация данных. Таким образом, для целей данной статьи мы определяем большие данные как объединение всего спектра инструментов и дисциплин, связанных со сбором, хранением и анализом больших объемов данных, возникающих при наблюдении за взаимодействием между пользователями и устройствами (телефонами).</p>
<p>Indeed, independent of this paper's focus, the data in Big Data applications is frequently generated by logging user interaction. Examples range from recording queries from search engines to transactions from online shops. This data is commonly characterized by the three "v"s, its (i) velocity, (ii) volume, and (iii) variety. The former two indicate the speed at which data records arrive, and the large volume they amount to. The latter refers to the wide range of different data types and sources commonly involved in even a single application. Consider Google as a prime example for a Big Data use case, the collected data origins from a wide variety of services, such as its search engine, email service, and smartphone offerings (Android). Consequently, this data comes in a wide range of formats. Furthermore, already recording the queries issued to Google's search engine generates data tuples at a frightening rate, requiring extensive server farms.</p>	<p>В самом деле, независимо от того, на чем сфокусирован этот документ, данные в приложениях больших данных часто генерируются путем регистрации взаимодействия с пользователем. Примеры варьируются от записи запросов от поисковых систем до транзакций из интернет-магазинов. Эти данные обычно характеризуются тремя «v», его (i) скоростью, (ii) объемом и (iii) разновидностью. Первые два указывают скорость, с которой поступают записи данных, и большой объем, который они составляют. Последнее относится к широкому спектру различных типов данных и источников, обычно используемых даже в одном приложении. Рассмотрим Google в качестве основного примера варианта использования больших данных, когда собранные данные поступают от широкого спектра услуг, таких как поисковая система, служба электронной почты и предложения для смартфонов (Android). Следовательно, эти данные поступают в широком диапазоне форматов. Кроме того, уже запись запросов, отправленных поисковой системе Google, генерирует кортежи данных с пугающей скоростью, что требует обширных ферм серверов.</p>
<p>Big Data applications commonly focus on data analysis. Researchers either attempt to (i) detect hitherto unknown trends and patterns, or (ii) "learn" new properties about known entities. The former discipline is commonly referred to as <i>data mining</i>, the latter as <i>machine learning</i>. One common application of data mining is the so-called market basket analysis, generating observations such as "customers who buy bread and butter commonly also buy beer". Machine learning in contrast attempts</p>	<p>Приложения для больших данных обычно сосредоточены на анализе данных. Исследователи либо пытаются (i) обнаружить ранее неизвестные тенденции и закономерности, либо (ii) «изучить» новые свойства об известных объектах. Первая дисциплина обычно называется интеллектуальным анализом данных, вторая машинным обучением. Одним из распространенных применений интеллектуального анализа данных является так называемый анализ</p>

<p>to detect hitherto unknown properties of objects. In classification for example, a common machine learning task, the software is shown a reasonably large training set of objects, for which the property of interest is known. By observing this <i>labeled</i> data, the software “learns” a computational model. Later, this model allows to classify other data items, i.e. automatically assign their labels. A credit card company could for example train a classifier by showing a large number transactions which have been labeled as either correct or fraudulent. In this example, features of interest could contain the country in which the card was issued, the one in which the transaction takes place, the type of shop and the amount involved. The resulting model could later be used to (dis-)approve transactions on a large scale.</p>	<p>корзины, генерирующий наблюдения, такие как «клиенты, которые покупают хлеб с маслом, обычно покупают пиво». Машинное обучение, напротив, пытается обнаружить до сих пор неизвестные свойства объектов. Например, в классификации, обычной задаче машинного обучения, программному обеспечению показан достаточно большой обучающий набор объектов, для которых известно свойство интереса. Наблюдая за этими помеченными данными, программное обеспечение «изучает» вычислительную модель. Позже эта модель позволяет классифицировать другие элементы данных, то есть автоматически назначать их метки. Компания, выпускающая кредитные карты, может, например, обучить классификатора, показывая большое количество транзакций, которые были помечены как правильные или мошеннические. В этом примере интересующие функции могут содержать страну, в которой была выпущена карта, страну, в которой происходит транзакция, тип магазина и сумму, которую он совершил. Получившаяся модель впоследствии может быть использована для (отклонения) одобрения сделок в крупном масштабе.</p>
<p>Generating the computational models for data analysis is inherently labor intensive. In a trial-and-error fashion, data scientists repeatedly test the data for a certain hypothesis, fail, reconsider their assumptions and start over. In particular for machine learning, they have to (i) label a large training set, (ii) identify the most expressive data features to expose to the learning software, and extract these features from the raw data. Note, only the creation of the initial computational model is labor intensive. Later, this model can be used on a large scale, at little additional cost, thus benefiting from an economy of scale.</p>	<p>Генерация вычислительных моделей для анализа данных обычно трудоемка. Методом проб и ошибок ученые-данные неоднократно проверяют данные на предмет определенной гипотезы, терпят неудачу, пересматривают свои предположения и начинают заново. В частности, для машинного обучения, они должны (i) маркировать большой обучающий набор, (ii) идентифицировать наиболее выразительные функции данных, которые можно использовать в обучающем программном обеспечении, и извлекать эти функции из необработанных данных. Обратите внимание, что только создание исходной вычислительной модели является трудоемким. Позже, эта модель может использоваться в больших масштабах, с небольшими дополнительными затратами, таким образом, выигрывая от эффекта масштаба.</p>
<p>(ii)</p>	<p>(iii)</p>
<p>In contrast to what the title might suggest, Big Data is not so much</p>	<p>В отличие от того, что может предложить заголовок, большие</p>

<p>characterized by the size of the accumulated data, but by the underlying motivation. Massive amounts of data are commonly collected <i>without an immediate business case</i>, but simply because it is affordable. This data, so it is hoped, will <i>later</i> answer questions, most of which yet have to arise. In addition to not knowing the later application scenario at the time of data collection, scientist also commonly cannot anticipate whether or not the data will display patterns of interest or support a certain hypothesis. In a Big Data context, data analytics thus become inherently post-hoc. The proposed architecture resembles those employed in many other Big Data scenario. As depicted in Fig. 1, it gathers data from various <i>sources</i>, preferably without requiring user interaction. It forwards this data to a central data <i>storage</i>, where it is analyzed and mined for patterns, trends, and outliers. The resulting knowledge is then used in various <i>application</i> scenarios, such as academic research, therapy, or entirely new products.</p>	<p>данные характеризуются не столько объемом накопленных данных, сколько их основной мотивацией. Огромные объемы данных обычно собираются без непосредственного экономического обоснования, но просто потому, что это доступно. Эти данные, как мы надеемся, позже дадут ответы на вопросы, большинство из которых еще предстоит выяснить. В дополнение к незнанию более позднего сценария применения во время сбора данных, ученый также обычно не может предвидеть, будут ли данные отображать представляющие интерес модели или поддерживать определенную гипотезу. Таким образом, в контексте больших данных аналитика данных становится неотъемлемой частью ситуации. Предлагаемая архитектура напоминает те, которые используются во многих других сценариях больших данных. Как показано на рис. 1, он собирает данные из различных источников, предпочтительно без необходимости взаимодействия с пользователем. Он передает эти данные в центральное хранилище данных, где их анализируют и анализируют на предмет закономерностей, тенденций и выбросов. Полученные знания затем используются в различных сценариях применения, таких как академические исследования, терапия или совершенно новые продукты.</p>
<p>Any piece of machinery allowing user interaction constitutes a potential data source. The user's <i>desktop</i> computer for example allows documenting the number of typographic errors, a potential symptom of stress and tiredness. Yet, modern IT has penetrated our everyday lives to an even higher degree. From coffee machines to cars and fridges, all gadgets in this world of <i>ubiquitous computing</i> (will) have some computational power, as well as the ability to communicate over the Internet. They can opaquely track the number of cups we require to get started in the morning, the frequency of out-of-schedule snacks, as well as our propensity to tailgate on the highway. Changes in any such parameter, so we claim, might indicate a shift in e.g., stress or aggression.</p>	<p>Любой механизм, позволяющий взаимодействовать с пользователем, представляет собой потенциальный источник данных. Например, настольный компьютер пользователя позволяет документировать количество типографских ошибок, потенциальных симптомов стресса и усталости. Тем не менее, современные информационные технологии проникли в нашу повседневную жизнь в еще большей степени. От кофемашин до автомобилей и холодильников все гаджеты в этом мире повсеместных вычислений (воля) обладают некоторой вычислительной мощностью, а также способностью общаться через Интернет. Они могут непрозрачно отследить количество чашек, которые нам необходимы, чтобы начать утро, частоту внеплановых перекусов, а также нашу склонность к задним ходам на шоссе. Мы утверждаем, что изменения любого такого</p>

	<p>параметра могут указывать на изменение, например, стресса или агрессии.</p>
<p>The introduction of smart watches and eyeglasses has further enhanced the ability to track human behavior. Additionally, recent advances in miniaturization have made <i>wearable sensors</i> small and affordable enough for everyday use. Only a few years ago, wristmounted accelerometers were the size of a bar of soap, and cost more than a desktop PC. Nowadays, the same technology fits to the size of a USB-stick, and sells for as little as 50\$US. While it is thus possible to track the user's movements in the real world, we must also track online behavior. The average user spends a significant amount of our lives interacting with <i>Web applications</i>, such as social networks, shopping sites or online games. Most of these services offer a programming-interface which straightforwardly allows recording with who we communicate, for how long, what we buy, at which time, and how much we enjoy bashing orcs with magic swords in computer games. Any such behavior might correlate with traits of personality, while reported changes therein might aid a clinical practitioner in treating a psychiatric disorder. The currently foremost source of behavioral data represents <i>Smartphones</i>, featuring a constant broadband connection, and exceeding the computational power of an early Pentium PC. They are intimate devices, carried on the body throughout the day. We commonly interact with our phone within the last half an hour before falling asleep, and no later than 30 min after waking up. (We will report the first exact data on this topic in the near future). The phones provide our most central platform to communicate, professionally, as well as to friends and spouses. The operating system allows tracking any phone-call or SMS, as well as any interaction with the installed software, the so-called apps. In addition, smartphones feature an entire set of sensors, such as for GPS signals and acceleration. It is thus possible to track the user's macro movement throughout the day, as well as e.g., her phones.</p>	<p>Внедрение умных часов и очков еще больше расширило возможности отслеживания поведения человека. Кроме того, последние достижения в области миниатюризации сделали носимые датчики небольшими и достаточно доступными для повседневного использования. Лишь несколько лет назад акселерометры, установленные на запястье, были размером с кусок мыла и стоили дороже, чем настольный ПК. В настоящее время та же технология соответствует размеру USB-флешки и продается всего за 50 долларов США. Таким образом, можно отслеживать движения пользователя в реальном мире, но мы также должны отслеживать поведение в Интернете. Средний пользователь проводит значительное количество наших жизней, взаимодействуя с веб-приложениями, такими как социальные сети, торговые сайты или онлайн-игры. Большинство из этих услуг предлагают программный интерфейс, который позволяет напрямую записывать, с кем мы общаемся, как долго, что мы покупаем, в какое время и насколько нам нравится избивать орков магическими мечами в компьютерных играх. Любое такое поведение может коррелировать с чертами личности, в то время как сообщаемые в нем изменения могут помочь практикующему врачу в лечении психического расстройства. В настоящее время основным источником данных о поведении являются смартфоны, которые имеют постоянное широкополосное соединение и превосходят вычислительные возможности ранних компьютеров Pentium. Это интимные приспособления, которые носят на теле в течение дня. Обычно мы общаемся с нашим телефоном в течение последних полчаса до засыпания и не позднее 30 минут после пробуждения. (Мы сообщим первые точные данные по этой теме в ближайшем будущем). Телефоны обеспечивают нашу самую центральную платформу для общения, как на профессиональном уровне, так и с друзьями и супругами. Операционная система позволяет отслеживать любой телефонный</p>

	звонок или SMS, а также любое взаимодействие с установленным программным обеспечением, так называемыми приложениями. Кроме того, смартфоны оснащены полным набором датчиков, например, для сигналов GPS и ускорения. Таким образом, можно отслеживать макроперемещение пользователя в течение дня, например, его телефонов.
On the central server, the data is subject to various levels of analysis. These can range from simple counting (the number of times someone flicks on his phone), to a complex form of data mining and machine learning, especially for temporal patterns. Computer scientists have generated an entire range of methods for data analysis that yet have to be broadly applied to behavioral data. While far from trivial, one can essentially detect all patterns that are also apparent to a (trained) human eye. Riding on a train next to a passenger who keeps checking his phone, one might deduce that he is nervous, or bored. The human eye might come to the conclusion of the passenger to be nervous instead of being bored if the human brain then connects the observation of constant phone checking with the trembling of hands. The above software architecture might draw similar conclusions, albeit on the scale of thousands of phone users. Frequently, modern machine learning even outperforms the observational powers of humans. Duhigg reports the prominent case of a concerned father complaining to an American retailer for marketing pregnancy related products to his teen-age daughter [19]. The company's marketing department had indeed analyzed user's shopping behavior, trying to identify women in the second trimester of their pregnancy, who would soon make baby-related purchases. As it turned out, the girl was indeed pregnant. The retailer's data-analytics team had outperformed the observational skills of her own father, who she was living with.	На центральном сервере данные подвергаются различным уровням анализа. Они могут варьироваться от простого подсчета (количество раз, когда кто-то нажимает на свой телефон) до сложной формы сбора данных и машинного обучения, особенно для временных моделей. Ученые-компьютерщики создали целый ряд методов для анализа данных, которые еще должны широко применяться к поведенческим данным. Хотя это далеко не тривиально, можно по существу обнаружить все паттерны, которые также очевидны (обученному) человеческому глазу. Когда вы едете в поезде рядом с пассажиром, который продолжает проверять свой телефон, можно сделать вывод, что он нервничает или ему скучно. Человеческий глаз может прийти к выводу, что пассажир должен нервничать, а не скучать, если человеческий мозг затем связывает наблюдение постоянной проверки телефона с дрожанием рук. Приведенная выше архитектура программного обеспечения может сделать аналогичные выводы, хотя и в масштабе тысяч пользователей телефонов. Зачастую современное машинное обучение даже превосходит наблюдательные способности людей. Duhigg сообщает о известном случае, когда обеспокоенный отец жаловался американскому ритейлеру за маркетинг продуктов, связанных с беременностью, его дочери подросткового возраста [19]. Отдел маркетинга компании действительно проанализировал покупательское поведение пользователей, пытаясь определить женщин во втором триместре беременности, которые вскоре будут делать покупки, связанные с ребенком. Как оказалось, девушка действительно была беременна. Команда аналитиков данных ритейлера превзошла наблюдательные навыки своего собственного отца, с которым она жила.
For the academic community, the above methodology yields an entire range	Для академического сообщества приведенная выше методология дает

<p>of benefits. First, it generates behavioral data of unprecedented quantity. A single cellphone may well produce a thousand data points per day. Combined with the methods of data mining and machine learning, the approach thus enters an uncharted area of behavioral patterns and trends. Second, it is entirely opaque to the user, and does not require any explicit interaction. It can thus be employed to track user behavior for months and years, without overburdening the participant. Third, the approach avoids the dominant sources of bias in studying human behavior. In particular, it does not remind the user that he is participating in a clinical trial. Most notably, collected data measures the objective behavior (i.e., how he actually interacted with the phone), in contrast to subjective self-reports, which commonly suffer from a distorted or faulty recollection. Finally, data is transferred and analyzed automatically, rendering the proposed approach significantly cheaper than traditional paper-based methods. In summary, the approach thus increases data quantity and quality, while simultaneously unburdening participants as well as researchers.</p>	<p>целый ряд преимуществ. Во-первых, он генерирует поведенческие данные беспрецедентного количества. Один мобильный телефон может производить тысячи точек данных в день. В сочетании с методами интеллектуального анализа данных и машинного обучения этот подход, таким образом, входит в неизведанную область поведенческих моделей и тенденций. Во-вторых, он абсолютно непрозрачен для пользователя и не требует явного взаимодействия. Таким образом, его можно использовать для отслеживания поведения пользователя в течение нескольких месяцев и лет, не перегружая участника. В-третьих, этот подход избегает доминирующих источников предвзятости при изучении поведения человека. В частности, это не напоминает пользователю, что он участвует в клиническом испытании. В частности, собранные данные измеряют объективное поведение (то есть то, как он фактически взаимодействовал с телефоном), в отличие от субъективных самоотчетов, которые обычно страдают от искаженного или ошибочного воспоминания. Наконец, данные передаются и анализируются автоматически, что делает предлагаемый подход значительно дешевле, чем традиционные бумажные методы. Таким образом, подход, таким образом, увеличивает количество и качество данных, одновременно обременяя участников и исследователей.</p>
<p>Any conclusion automatically deducted from observations of human behavior can naturally be off. Typographic errors for example may as well be due to stress, as to a four-year-old yanking our sleeve. Thus, we do not propose to automatically generate diagnoses, but to quantitatively assist the medical practitioner, akin to a complete blood count. Following a multi-modal approach, simultaneously taking several different sensors into account, can further minimize error rates. For example, the combination of an increasingly monotonous voice, and progressing social withdrawal might indicate the worsening of a case of depression. Commonly, it will not be possible to determine absolute thresholds. Instead, one will rather investigate intra-personal change over time. For example, it will not be possible to deduce stress from the rate of typos exceeding x per 100 lines of written text. In contrast, a 50% increase in typos over 6 months could imply</p>	<p>Любой вывод, автоматически вычитаемый из наблюдений за поведением человека, может быть естественным. Типичные ошибки, например, могут быть вызваны стрессом, например, четырехлетнему дерганию за рукав. Таким образом, мы не предлагаем автоматически генерировать диагнозы, но количественно помогать практикующему врачу, похожему на полный анализ крови. Следуя мультимодальному подходу, одновременно принимая во внимание несколько различных датчиков, можно дополнительно минимизировать частоту появления ошибок. Например, сочетание все более монотонного голоса и прогрессирующей социальной самоотдачи может указывать на ухудшение ситуации депрессии. Обычно невозможно определить абсолютные пороговые значения. Вместо этого каждый будет скорее исследовать внутриличностные изменения с течением времени.</p>

<p>a substantial change in the patient's condition. Research will thus focus on detecting changes over time, the first statistical derivation of the observed signal. Likewise, it will frequently not be possible to detect trends applicable to an entire population of users. Yet, it will be feasible to detect typologies, i.e. classes of people exhibiting similar behavior. Data mining, in form of clustering and classification, can help find these typologies as well as the characteristic patterns in behavior.</p>	<p>Например, будет невозможно вывести стресс из числа опечаток, превышающих x на 100 строк письменного текста. Напротив, увеличение опечаток на 50% в течение 6 месяцев может означать существенное изменение состояния пациента. Таким образом, исследования будут сосредоточены на обнаружении изменений во времени, первом статистическом выводе наблюдаемого сигнала. Аналогичным образом, зачастую невозможно определить тенденции, применимые ко всей группе пользователей. Тем не менее, будет возможно выявить типологии, то есть классы людей, демонстрирующих сходное поведение. Интеллектуальный анализ данных в форме кластеризации и классификации может помочь найти эти типологии, а также характерные модели поведения.</p>
<p>Current research hypotheses in psychiatry and psychology</p> <p>In two current studies, we monitor smartphones to track (i) the severity and course of depression as well as (ii) conspicuous usage of the Internet and phone. While these studies are decidedly smallscale, at least compared to the above technological vision, they are primarily intended to evaluate the validity as well as practicability of the proposed methodology.</p>	<p>Актуальные исследования гипотез в психиатрии и психологии</p> <p>В двух текущих исследованиях мы проводим мониторинг смартфонов, чтобы отслеживать (i) степень тяжести и течение депрессии, а также (ii) заметное использование Интернета и телефона. Хотя эти исследования явно маломасштабны, по крайней мере, по сравнению с вышеупомянутым технологическим видением, они в первую очередь предназначены для оценки обоснованности, а также осуществимости предложенной методологии.</p>
<p>Monitoring depression</p> <p>In a first study, we currently employ smartphones to monitor depression. In particular we observe (i) app usage, (ii) social interaction, and (iii) macro movement of patients. In the context of depression, our central thesis is that these (and eventually most other) signals show less energy and dynamics, reflecting anhedonia and social withdrawal as central symptoms. To collect the necessary usage data, we have developed an app running on Android phones (version 4.0 and up), comprised of a software module for each of the three parameters, and one for encrypted data transfer to the server.</p>	<p>Мониторинг депрессии</p> <p>В первом исследовании мы в настоящее время используем смартфоны для мониторинга депрессии. В частности, мы наблюдаем (i) использование приложения, (ii) социальное взаимодействие и (iii) макро движение пациентов. В контексте депрессии наш центральный тезис заключается в том, что эти (и в конечном итоге большинство других) сигналов демонстрируют меньшую энергию и динамику, отражая ангедонию и социальную абстиненцию как центральные симптомы. Чтобы собрать необходимые данные об использовании, мы разработали приложение, работающее на телефонах Android (версия 4.0 и выше), состоящее из программного модуля для каждого из трех параметров и одного для передачи зашифрованных данных на сервер.</p>
<p>For app usage, we introduce the concept of an app-session, the time interval</p>	<p>Для использования приложения мы вводим концепцию сеанса</p>

<p>during which the user interacts with any one particular app. This implies that the app runs in the foreground, and the patient actively interacts with the phone. A session ends, when the patient (i) switches to another app, (ii) closes the app, or (iii) stops interacting with the phone. We thus record data tuples of the format: [app name, start-time, end-time]. In this context, we treat both, the lock-screen as well as the home menu, as just two additional apps. We thus also record, how often the user flicks the phone on (without actually unlocking it), and how much time he spends in the main menu.</p>	<p>приложения, временного интервала, в течение которого пользователь взаимодействует с любым конкретным приложением. Это означает, что приложение работает на переднем плане, и пациент активно взаимодействует с телефоном. Сеанс заканчивается, когда пациент (i) переключается на другое приложение, (ii) закрывает приложение или (iii) прекращает взаимодействие с телефоном. Таким образом, мы записываем кортежи данных в формате: [имя приложения, время начала, время окончания]. В этом контексте мы рассматриваем и экран блокировки, и домашнее меню, как два дополнительных приложения. Таким образом, мы также записываем, как часто пользователь включает телефон (фактически не разблокируя его) и сколько времени он проводит в главном меню.</p>
<p>Regarding communication patterns, we log inand outgoing SMS as well as phone calls. These measures allow to infer the size and usage of the users social network. In this context, we only document with how many contacts the user regularly interacts, and who initiates the communication. The actual content of the calls (or SMS) is of no concern. Also, we are not interested in the actual identity of the communication partner. Hence, we anonymize their phone numbers, using cryptographic hash functions (SHA-512). Akin to app-session, we record time stamped data tuples, namely for calls: [anonymized number, start-time, end-time, in/out], for outgoing SMS: [anon. number, length in characters, time-sent], and for incoming SMS: [anon. number, length in characters, time-received, and time-read]. Additionally, we monitor the phone's address book, to estimate the total number of contacts, their changes over time, and the fraction, which the user actively communicates with.</p>	<p>Что касается шаблонов общения, мы регистрируем входящие и исходящие SMS, а также телефонные звонки. Эти меры позволяют вывести размер и использование социальной сети пользователей. В этом контексте мы только документируем, сколько контактов пользователь регулярно взаимодействует, и кто инициирует общение. Фактическое содержание звонков (или SMS) не имеет значения. Также нас не интересует реальная личность партнера по общению. Следовательно, мы анонимизируем их телефонные номера, используя криптографические хеш-функции (SHA-512). Аналогично сеансы, мы записываем кортежи данных с меткого времени, а именно для вызовов: [анонимный номер, время начала, время окончания, вход / выход], для исходящих SMS: [анон. номер, длина в символах, время отправки] и для входящих SMS: [анон. число, длина в символах, время получения и время чтения]. Кроме того, мы отслеживаем адресную книгу телефона, чтобы оценить общее количество контактов, их изменения во времени и долю, с которой пользователь активно общается.</p>
<p>For macro movement, we regularly record the user's geographic position. In the context of severe depression, we are commonly concerned whether a user even just leaves the apartment. Every 20 min, we thus estimate his location using GPS, recording the data tuple [time, latitude, longitude, accuracy]. Location estimation is additionally supported through available</p>	<p>Для макроперемещения мы регулярно записываем географическое положение пользователя. В контексте тяжелой депрессии мы обычно обеспокоены тем, покидает ли пользователь квартиру. Таким образом, каждые 20 минут мы оцениваем его местоположение с помощью GPS, записывая кортеж данных [время, широта, долгота, точность]. Оценка</p>

<p>Wi-Fi signals and triangulation of cell-phone towers. Such functionality is readily provided by the phone's operating system. The resulting accuracy can be expected to range between 10 and 100 m, enough for macro movement, but not sufficient to track movement inside a building. We are currently embarking on a study, observing patients suffering from major depressive disorder (MDD) over a period of four months. These patients are either treated with electro convulsive therapy (ECT), magnetic seizure therapy (MST) or deep brain stimulation (DBS) in the Department of Psychiatry and Psychotherapy at the University Hospital of Bonn. In order to establish a baseline, we install the app 1 month prior to the treatment, and continue measuring over the following 3 months of treatment. For the assessment of the current depressive symptomatology, we apply the Montgomery-Åsberg Depression Rating Scale (MADRS) as well as the Beck Depression Inventory (BDI-II) on a biweekly basis, as clinician and self-rated questionnaires respectively. Due to drastic improvements commonly caused by the employed treatment methods, we expect to observe a significant signal-to-noise ratio. Furthermore we hypothesize that there is a significant correlation between the data concerning app usage, social interaction, macro movement and the applied questionnaires. We moreover hypothesize that this new method for monitoring the severity and course of depression is more sensitive to change, i.e. improvement of the depressive symptomatology (increase in activity patterns, resumption of social contacts) can be detected earlier.</p>	<p>местоположения дополнительно поддерживается посредством доступных сигналов Wi-Fi и триангуляции вышек сотовой связи. Такая функциональность легко обеспечивается операционной системой телефона. Можно ожидать, что полученная точность будет колебаться от 10 до 100 м, что достаточно для макроперемещения, но недостаточно для отслеживания движения внутри здания. В настоящее время мы начинаем исследование, наблюдая пациентов, страдающих от серьезного депрессивного расстройства (MDD) в течение четырех месяцев. Эти пациенты получают лечение с помощью электросудорожной терапии (ЭСТ), магнитотерапии (MST) или глубокой стимуляции головного мозга (DBS) в отделении психиатрии и психотерапии Университетской больницы Бонна. Чтобы установить базовый уровень, мы устанавливаем приложение за 1 месяц до начала лечения и продолжаем измерение в течение следующих 3 месяцев лечения. Для оценки текущей депрессивной симптоматики мы применяем Шкалу оценки депрессии Монтгомери-Осберга (MAD-RS), а также Инвентаризацию депрессии Бека (BDI-II) на двухнедельной основе, в качестве опросников для врачей и самооценок соответственно. Из-за радикальных улучшений, обычно вызываемых используемыми методами лечения, мы ожидаем увидеть значительное отношение сигнал / шум. Кроме того, мы предполагаем, что существует значительная корреляция между данными, касающимися использования приложений, социального взаимодействия, макродвижения и прикладных вопросников. Кроме того, мы выдвигаем гипотезу о том, что этот новый метод мониторинга степени тяжести и течения депрессии более чувствителен к изменениям, то есть улучшение депрессивной симптоматики (увеличение моделей активности, возобновление социальных контактов) может быть выявлено ранее.</p>
<p>Internet addiction & online social networks</p>	<p>Интернет-зависимость и онлайн социальные сети</p>
<p>Our second current study addresses a topic inherently linked to smartphone usage – Internet Addiction. As it does not constitute a distinct disorder in the DSM-V, but can only be found in the Appendix, further research</p>	<p>Наше второе текущее исследование затрагивает тему, связанную с использованием смартфонов, интернет-зависимость. Поскольку он не представляет собой явное расстройство в DSM-V, а может быть</p>

<p>efforts are required to better characterize this emerging threat to our mental health. Until now, the diagnosis of Internet Addiction heavily relies on questionnaires and to a lesser extent on structural interviews. Given the mentioned shortcomings of both methods, the detection of behavioral addictions will profit enormously from the collaboration with computer science. Indeed, illustrating the validity and practicability of Psycho-Informatics represents one of the most important aims of this our research endeavor.</p>	<p>обнаружен только в Приложении, необходимы дальнейшие исследовательские усилия, чтобы лучше охарактеризовать эту возникающую угрозу нашему психическому здоровью. До сих пор диагноз интернет-зависимости в значительной степени опирается на вопросники и, в меньшей степени, на структурные интервью. Принимая во внимание упомянутые недостатки обоих методов, выявление поведенческой зависимости извлечет огромную выгоду из сотрудничества с информатикой. Действительно, иллюстрирование обоснованности и практичности психоинформатики представляет собой одну из наиболее важных целей этого нашего исследования.</p>
<p>In a pilot study, we currently monitor the mobile phone behavior of $N = 100$ healthy participants for a duration of 6 weeks. This data thus provides the first reliable longitudinal statistics on diverse facets of smartphone usage and (mobile) Internet Addiction. We are particularly interested in fluctuations in the mobile phone behavior across the 6 weeks, but also search for stable – non-changing – facets. Questions to be answered are: How often is a phone exactly used each day? What are the most commonly used applications? When does a person go online for the first time in the morning? How often do they check their phones, mails or news? All these questions are impossible to grasp accurately by just asking participants. Instead, behavior needs to be recorded on the device. Indeed, a central point to the current study is the vast discrepancy between questionnaire based self-reports and actual phone usage. Additionally, we try to better characterize social activity. In this context, we explicitly do not limit ourselves to online social networks (such as Facebook), but study the size of the actual active social network reflected by phone-calls and SMS. Since we also record brain scans (structural MRI and resting state fMRI) and genetic material from participants, we are then able to correlate biological markers with social activity. Ideally, we would find a correlation between the active social network and certain areas of the brain. A recent study addressed a similar question, correlating amygdala volumes with the size of the social network [20]. This work however had to rely on self-reports, whereas we are able to measure the actual social activity, and in much finer detail. Ultimately, correlating Big</p>	<p>В пилотном исследовании мы в настоящее время отслеживаем поведение мобильного телефона $N = 100$ здоровых участников в течение 6 недель. Таким образом, эти данные обеспечивают первую надежную продольную статистику по различным аспектам использования смартфонов и (мобильной) интернет-зависимости. Мы особенно заинтересованы в колебаниях поведения мобильных телефонов в течение 6 недель, а также в поиске стабильных, не меняющихся аспектов. Вопросы для ответа: как часто телефон используется каждый день? Каковы наиболее часто используемые приложения? Когда человек выходит в Интернет впервые утром? Как часто они проверяют свои телефоны, почту или новости? Все эти вопросы невозможно точно понять, просто задавая вопросы участникам. Вместо этого поведение должно быть записано на устройстве. Действительно, центральным моментом текущего исследования является огромное расхождение между самоотчётами, основанными на вопросниках, и фактическим использованием телефона. Кроме того, мы стараемся лучше охарактеризовать социальную активность. В этом контексте мы явно не ограничиваемся сетевыми социальными сетями (такими как Facebook), а изучаем размер реальной активной социальной сети, отражаемой телефонными звонками и SMS. Так как мы также записываем сканирование мозга (структурная МРТ и МРТ в состоянии покоя) и генетический материал от участников, мы можем соотнести биологические маркеры с социальной активностью. В</p>

<p>Data with neuroscientific measures should carry far enough to eventually establish its own research direction of Psychoneuroinformatics.</p>	<p>идеале мы бы нашли корреляцию между активной социальной сетью и определенными областями мозга. Недавнее исследование касалось аналогичного вопроса, связывающего объемы миндалины с размером социальной сети [20]. Эта работа, однако, должна была опираться на самоотчеты, в то время как мы можем измерить реальную социальную активность и во многих деталях. В конечном счете, корреляция Больших Данных с нейробиологическими мерами должна продвинуться достаточно далеко, чтобы в конечном счете установить свое собственное направление исследования Психонейроинформатики.</p>
<p>Ethical aspects and data privacy issues of ‘Big Data’ research</p>	<p>Этические аспекты и проблемы конфиденциальности данных исследования «Большие данные»</p>
<p>The use of Big Data in research and therapy necessarily raises ethical concerns. Bordering mass surveillance, it realizes the vision of a “Gläserner Mensch”, a transparent human. Data privacy thus takes on a central role, and the potential of abuse cannot be overestimated. While monitoring depression in a medical scenario fulfills the highest ethical standards, it could equally well be misused by an employer to secretly monitor his staff, or by an insurance company to reject at-risk applicants. This research is however not aimed at Digital Taylorism, a strategy that would surely backfire, but preventing, detecting and curing psychological disorders. Both, medicine and psychology, have worked on the vision of a transparent human since their very beginning. And for the same time-frame, both have had to handle sensitive data. There thus exists a proud tradition of confidentiality, whose methods can serve as blueprints for the deployment of Big Data technologies. Scaling and extending these concepts to an entirely new dimension is no mean feat, and will generate a significant amount of work for researchers, practitioners and occupational bodies.</p>	<p>Использование больших данных в исследованиях и терапии обязательно вызывает этические проблемы. Граничащий с массовым наблюдением, он реализует видение «Gläserner Mensch», прозрачного человека. Таким образом, конфиденциальность данных играет центральную роль, и вероятность злоупотребления не может быть переоценена. Хотя мониторинг депрессии в медицинском сценарии соответствует самым высоким этическим стандартам, работодатель может также неправильно использовать его для тайного мониторинга своих сотрудников или в страховой компании для отклонения заявителей, подвергающихся риску. Это исследование, однако, не нацелено на Цифровой Тейлоризм, стратегию, которая, несомненно, будет иметь обратную силу, но предотвращает, выявляет и лечит психологические расстройства. И медицина, и психология с самого начала работали над видением прозрачного человека. И за один и тот же период времени оба должны были обрабатывать конфиденциальные данные. Таким образом, существует гордая традиция конфиденциальности, методы которой могут служить основой для развертывания технологий больших данных. Масштабирование и расширение этих концепций до совершенно нового измерения — это немалый подвиг, и оно создаст значительный объем работы для исследователей, практиков и профессиональных организаций.</p>

<p>From a different perspective, privacy concerns constitute somewhat of a side effect. Their denial would be entirely unethical. But, as this paper should outline, Big Data holds the potential to facilitate treatment of mental diseases. Hence, it would be equally unethical to bluntly deny its usage due to privacy concerns. Rather, the medical sciences have to follow another of their proud traditions: balancing risks and benefits on a case-by-case basis.</p>	<p>С другой стороны, проблемы конфиденциальности представляют собой некоторый побочный эффект. Их отрицание было бы совершенно неэтично. Но, как следует из этой статьи, большие данные обладают потенциалом для облегчения лечения психических заболеваний. Следовательно, было бы одинаково неэтично прямо отрицать его использование из-за проблем конфиденциальности. Скорее, медицинские науки должны следовать другой из своих гордых традиций: балансировать риски и выгоды в каждом конкретном случае.</p>
<p>On a practical level, we currently follow a simple <i>two level privacy model</i>. We only collect usage data (i.e., behavior) that commonly needs to be aggregated to make any significant statement. In contrast, we refrain from collecting one-to-one written conversation, audio recordings, or video captures. Where applicable, text, audio and video are analyzed locally on the phone, and only the resulting markers and numeric values are sent to the server. In a current prototype for example, we estimate the size of the vocabulary from messenger applications on the phone, and only report resulting figures to the server.</p>	<p>На практическом уровне мы в настоящее время следуем простой двухуровневой модели конфиденциальности. Мы собираем только данные об использовании (т. Е. Поведение), которые обычно необходимо агрегировать, чтобы сделать какое-либо важное заявление. В отличие от этого, мы воздерживаемся от сбора письменных обращений один на один, аудиозаписей или видеозаписей. Где это применимо, текст, аудио и видео анализируются локально на телефоне, и только полученные маркеры и числовые значения отправляются на сервер. Например, в текущем прототипе мы оцениваем размер словаря из приложений мессенджера на телефоне и сообщаем только полученные результаты на сервер.</p>
<p>On a different note, this paper does <i>not</i> propose an automatic diagnosis of psychiatric disorders. Instead, it suggests to assist the medical practitioner with additional information, in order to monitor the course of a disease and treatment. In certain cases, it may point a clinician in a certain direction while making an initial diagnosis. The ultimate responsibility, for false positives as well as negatives, remains with the clinician, as it has for two-thousand years. Whether one may be able to fully automatically diagnose mental disorders at some point in the future, we dare not hypothesize. Neither case would render the clinician obsolete, as communicating the findings to a patient, and even whether to do so at all, is a complex matter, not to be left to machinery. Proposed methodology could potentially serve as an early warning sign to a</p>	<p>С другой стороны, эта статья не предлагает автоматическую диагностику психических расстройств. Вместо этого он предлагает помочь врачу с дополнительной информацией, чтобы контролировать течение болезни и лечения. В определенных случаях он может указывать врачу в определенном направлении при постановке первоначального диагноза. Окончательная ответственность за ложноположительные и отрицательные результаты остается на клиницисте, как и на протяжении двух тысяч лет. Будем ли мы в состоянии полностью автоматически диагностировать психические расстройства в какой-то момент в будущем, мы не осмеливаемся предположить. Ни один из этих случаев не сделает клинициста устаревшим, поскольку передача</p>

<p>medical practitioner indicating that a patient might suffer from a condition. In this spirit, the paper compared the proposed methodology to a complete blood count. Neither technique makes a diagnosis, but assists a diagnosis, made by a clinical practitioner. The ultimate responsibility, for false positives as well as negatives, thus remain with the clinician, as it has for two thousand years. By providing additional quantitative information, we would hope to reduce the inherent actalepsy.</p>	<p>информации пациенту и даже вопрос о том, делать ли это вообще, является сложным вопросом, и его не следует оставлять на усмотрение аппарата. Предлагаемая методология потенциально может служить ранним предупреждением. подписать к врачу, указывающий, что пациент может страдать от состояния. В этом духе статья сравнила предложенную методологию с полным анализом крови. Ни один из методов не ставит диагноз, но помогает диагнозу, поставленному практикующим врачом. Таким образом, основная ответственность за ложноположительные и отрицательные результаты остается на клиницисте, как и на протяжении двух тысяч лет. Предоставляя дополнительную количественную информацию, мы надеемся уменьшить присущую акталепсию.</p>
<p>Conclusions and vision for the future</p>	<p>Выводы и видение будущего</p>
<p>This paper introduces Psycho-Informatics, the application of Big Data to psychology and psychiatry. Highly sensitive, the suggested method collects, stores, and analyzes massive amounts of indicative data at little cost and without risks or stress for patients or participants. The paper outlines the technical vision, sketches the signals that can be detected, and illustrates the tremendous benefits over traditional methods of psychometrics. In particular, it suggests tracking user behavior with smartphones, a particularly rich and intimate source of data. This approach underlies two current studies, in the context of (i) depression and (ii) (excessive) usage of smartphones. The proposed methodology outperforms traditional methods in both quality as well as quantity. Namely, it avoids biased self-reports, and avoids altering the user's behavior, as the data is completely collected in the background. Furthermore, it collects data on a much finer granularity than conventional questionnaires and enables the search for temporal activity patterns. Additionally, there is no need to collect the data manually, as it is directly available in electronic form. While there are strong ethical concerns, these must not be allowed to evolve into thoughtterminating clichés. Instead, they are to be addressed on a detailed level, case-by-case, following a rich tradition in medicine as well as psychology. These</p>	<p>Эта статья знакомит с психоинформатикой, применением больших данных в психологии и психиатрии. Высокочувствительный, предлагаемый метод собирает, хранит и анализирует огромное количество индикативных данных при небольших затратах и без риска или стресса для пациентов или участников. В документе излагается техническое видение, дается набросок сигналов, которые могут быть обнаружены, и иллюстрируются огромные преимущества по сравнению с традиционными методами психометрии. В частности, он предлагает отслеживать поведение пользователей с помощью смартфонов, особенно богатого и интимного источника данных. Этот подход лежит в основе двух текущих исследований, в контексте (i) депрессии и (ii) (чрезмерного) использования смартфонов. Предложенная методология превосходит традиционные методы как по качеству, так и по количеству. А именно, это позволяет избежать предвзятых самоотчетов и избежать изменения поведения пользователя, поскольку данные полностью собираются в фоновом режиме. Кроме того, он собирает данные с гораздо меньшей степенью детализации, чем обычные опросники, и позволяет искать временные модели активности. Кроме того, нет необходимости собирать данные</p>

<p>concerns being addressed appropriately, Big Data is about to revolutionize both psycho-sciences in research as well as therapy.</p>	<p>вручную, так как они доступны напрямую в электронном виде. Хотя существуют серьезные этические соображения, нельзя допустить, чтобы они превратились в клише, ограничивающее мысли. Вместо этого они должны рассматриваться на детальном уровне, в каждом конкретном случае, следуя богатой традиции в медицине и психологии. Эти проблемы решаются надлежащим образом, Big Data собирается революционизировать как психологические науки в области исследований, так и терапии.</p>
<p>In the near future, the researchers will embark on numerous projects incorporating simple aspects of human-machine-interaction and wearable sensors. Already in the medium future however, focus will shift towards data analytics. Once the low hanging fruit has been picked (e. g. ‘simple’ descriptive statistical data on what is done how often on a smartphone), scientists will need to dig deeper inside data. Simple aggregate functions (e.g., count, or sum) will no longer suffice, but be replaced by mining for complex temporal patterns. Eventually, the entire range of methodologies from data mining and machine learning will have to be adapted to behavioral data. The effects on psychology and psychiatry will equal those of the massive change that the life sciences have undergone, and even fundamental research methodologies will have to be revisited. We will frequently hear of Psycho-Informatics and its sub-areas, such as Psycho-Neuro-Informatics. Most importantly, it will be possible, and not uncommon, to make an academic career, by solely studying data. Frequently, we will no longer design studies, but subject existing data to deeper analysis. Some of these datasets may have been conceived as a by-product of entirely different (non-academic) applications. In this context, scientists will have to learn to yield control. Research will shift from carefully constructed experiments on small parts of the population in a controlled environment, to massive longitudinal recorded data on tremendously large populations, full of errors and noise. Yet, so we hypothesize, signals will contrast from noise clearer than ever, due to the sheer amount of data.</p>	<p>В ближайшем будущем исследователи приступят к реализации многочисленных проектов, включающих простые аспекты взаимодействия человека с машиной и носимых датчиков. Однако уже в среднесрочной перспективе акцент сместится в сторону анализа данных. После того, как плоды с низкой степенью подвески собраны (например, «простые» описательные статистические данные о том, что и как часто делается на смартфоне), ученым нужно будет копать углубленные внутренние данные. Простые агрегатные функции (например, счет или сумма) больше не будут достаточными, но будут заменены извлечением данных для сложных временных шаблонов. В конечном итоге весь спектр методологий от интеллектуального анализа данных и машинного обучения придется адаптировать к поведенческим данным. Влияние на психологию и психиатрию будет таким же, как и в случае массовых изменений, которые произошли в науках о жизни, и даже придется пересмотреть фундаментальные методологии исследования. Мы будем часто слышать о психоинформатике и ее подразделах, таких как психо-нейроинформатика. Наиболее важно, что будет возможно, и весьма часто, сделать академическую карьеру, только изучая данные. Часто мы больше не будем проектировать исследования, а подвергнем существующие данные более глубокому анализу. Некоторые из этих наборов данных могли быть задуманы как побочный продукт совершенно разных (не академических) приложений. В</p>

	<p>этом контексте ученым придется научиться уступать контроль. Исследования перейдут от тщательно построенных экспериментов на небольших частях населения в контролируемой среде к огромным продольным зарегистрированным данным об огромных популяциях, полных ошибок и шума. Тем не менее, как мы предполагаем, сигналы будут контрастировать с шумом более четким, чем когда-либо, благодаря большому количеству данных.</p>
<p>To the same extend, Big Data will affect the daily routines of patients and clinical practitioners alike. The former will collect seemingly unrelated data to share with coaches and therapists. Provided with the necessary toolkits and expertise, the latter will be able to observe the course (and origins) of a disorder as well as the progress of treatment. This painted picture will be more accurate compared to previous self-reports, and of such fine granularity as to allow a highly individualized medication. Compared to traditional methods, this data-driven therapy will be cheaper, and consume less time from both, therapists as well as patients. Such technology requires tremendous research efforts. Already the establishment of meaningful metrics (as opposed to raw data tuples) as well as visual data exploration tools will be a laborious and ongoing effort. However, in the medium to long term, this data-driven therapy will become cheaper than traditional methods, consuming less time from both, therapists as well as patients. Most importantly, the approach extends the benefits of psychiatry and psychology far beyond treatment, into systems for early warning and mental wellness.</p>	<p>В то же время большие данные будут влиять на повседневную жизнь пациентов и практикующих врачей. Первый будет собирать, казалось бы, не связанные данные, чтобы поделиться с тренерами и терапевтами. При наличии необходимого инструментария и опыта последние смогут наблюдать за течением (и происхождением) расстройства, а также за ходом лечения. Эта нарисованная картина будет более точной по сравнению с предыдущими самоотчетами и такой тонкой гранулярностью, что позволит получать высоко индивидуализированные лекарства. По сравнению с традиционными методами эта терапия, основанная на данных, будет дешевле и займет меньше времени как у терапевтов, так и у пациентов. Такая технология требует огромных исследовательских усилий. Уже создание значимых метрик (в отличие от кортежей необработанных данных), а также инструментов визуального исследования данных будет трудоемким и постоянным усилием. Однако в среднесрочной и долгосрочной перспективе эта управляемая данными терапия станет дешевле, чем традиционные методы, что потребует меньше времени как у терапевтов, так и у пациентов. Самое главное, что этот подход расширяет преимущества психиатрии и психологии далеко за пределы лечения, в системы раннего предупреждения и психического благополучия.</p>
Conflicts of interest statement	
<p>None of the authors' reports a conflict related to the work described. The software mentioned is currently developed for research purposes only, no commercial exploitation of it is planned at this stage.</p>	
Acknowledgement	

This work was partially funded in part by a grant awarded to	
C.M. by the DFG (MO-2363/2-1) and an independent investigator grant for the assessment of effects of deep brain stimulation for treatment resistant depression by Medtronic Inc. to TS.	