

Est.
1841

YORK
ST JOHN
UNIVERSITY

Denham, Jack and Spokes, Matthew (2022) Little Data: Negotiating the 'New Normal' with Idiosyncratic and Incomplete Datasets. *International journal of Social Research Methodology*, 26 (6). pp. 679-691.

Downloaded from: <https://ray.yorks.ac.uk/id/eprint/6329/>

The version presented here may differ from the published version or version of record. If you intend to cite from the work you are advised to consult the publisher's version:
<https://www.tandfonline.com/doi/full/10.1080/13645579.2022.2087850>

Research at York St John (RaY) is an institutional repository. It supports the principles of open access by making the research outputs of the University available in digital form. Copyright of the items stored in RaY reside with the authors and/or other copyright owners. Users may access full text items free of charge, and may download a copy for private study or non-commercial research. For further reuse terms, see licence terms governing individual outputs. [Institutional Repositories Policy Statement](#)

RaY

Research at the University of York St John

For more information please contact RaY at
ray@yorks.ac.uk

Little Data: Negotiating the ‘New Normal’ with Idiosyncratic and Incomplete Datasets

Dr Jack Denham & Dr Matthew Spokes

Abstract

In this paper we make a case for ‘Little Data’, which is real-time, self-collected, idiosyncratic datasets maintained by individuals about themselves on myriad topics. We develop and offer a methodology for combining these messy, highly personal insights, to make deductive observations about collective practices. In testing this approach, we use the case study of the 2020-21 stay-at-home orders imposed in the U.S.A., U.K. and Western Europe during the Coronavirus pandemic to operationalise and demonstrate the applicability of this method. Our main finding is to show that whilst stay-at-home orders did have a significant impact on habits during the COVID-19 pandemic, these changes were often counterintuitive, of an insightful nature on topics that would otherwise not be investigated, and always short-lived. Our main contribution is to present Little Data, despite and because of its fragmented and disparate nature, as a viable and useful tool to understand personal habits at finite junctures.

Keywords: Big Data, Little Data, tracking, lifelogging, COVID-19

Introduction

Located at the intersection between Dodge and Kitchin’s (2007) discussions on the totality of ‘life-logging’ and Lupton’s (2016) work on ‘self-tracking’, this paper uses user-collected, personal, introspective data composed and publicly posted by anonymous internet users over the period of the 2019-2021 Coronavirus pandemic to consider broader methodological issues

around engagement with large digital datasets, alongside the confluence of individual, subjective experiences of data collection, and shared reflection and analysis after-the-fact as the paper was co-constructed by the authors. Drawing on theoretical literature including that on flawed recording practices (Garfinkel 1967), longitudinal classificatory change (Uprichard 2011), and Benjamin's (2009) reflections on authorship and translation, we explore the ways in which these approaches might be situated within discussions of 'Big Data', identity and the quantification of self, arriving through Becker's (2007) concept of 'makers' and 'users' to present a methodological case for 'Little Data'.

In testing our approach with Little Data, we use the spring and winter lockdowns that took place in most of Western Europe and North America during 2020 to see changes in habits and lifestyles, as individually recorded by personal collectors on myriad topics. We sampled 43 datasets of various time lengths, all of which covered this period at least. Our main finding is to show that whilst stay-at-home orders did have a significant impact on habits during the COVID-19 pandemic, these habitual changes were often counterintuitive, of an insightful nature on topics that would otherwise not be investigated, and always short-lived. We demonstrate the frail nature of changes forced or motivated through external circumstances – quick as they were to revert back to their pre-pandemic norms in all 43 cases – running counter to media and political narratives of the 'new normal' which sought to cement lifestyle changes in semi-permanence (see, for example, Brickman 2021; Woodcock 2020). Alongside this we consider recent discussions around how personhood is increasingly constructed in response to data (Lupton 2019) or that 'we informational persons have become our data' (Koopman 2019:9).

Our main contribution is to present and situate Little Data, despite and because of its fragmented and disparate nature, as a viable and useful tool to understand personal habits at finite junctures. As Uprichard (2011:110) states: 'the social world has changed and so too must

our ways of knowing change with it'; to this end, we use Becker's (2007) concept of 'makers' and 'users' in offering an approach to unpacking, interpreting, and 'telling' with this relatively new, messy form of data.

In the first section (literature review), we situate this form of data alongside existing work on 'life-logging', 'self-tracking' and 'Big Data'. In section two (method), we define the parameters of our approach and propose a methodological framework that can be used to unpack these data. Then (on coronavirus), we put this method into practice, offering findings on the COVID-19 pandemic. Finally (discussion), we consider Little Data's strengths, weaknesses, and opportunities for future work, concluding our findings on the pandemic.

Literature review

As outlined in the introduction, our opening contention could be understood as a binary, so to begin to challenge that assumption it is vital to situate Little Data robustly in extant literature on data use, framed by research around how data and our everyday lives intersect and, crucially, are recorded and rendered in particular ways for particular reasons.

The first aspect to highlight is that these intersections are facilitated by technological developments in terms of how data are recorded. We are not claiming that prior to this moment in time people have not recorded a variety of data about their everyday experiences – Ibn Banna's diary observing life in Baghdad can be dated back to 1068 (Makdisi 1986) for instance – but rather that developing technologies are now 'directed at monitoring aspects of human lives and rendering them into digital datasets' (Lupton 2018:1) through the collection of what Beer (2013) terms 'digital byproduct data'. These data are routinely framed under the umbrella of 'Big Data', a multifarious term that is often understood as a catch-all encompassing the aggregation of huge amounts of information and those techniques used to understand social

action within these data, though the exact definition is of course open for indefinite debate (see Favaretto et al 2020).

This is where the difficulty with Big Data, in a terminological sense, emerges: Big Data implies *scale*. A case-in-point, in the first issue of the journal *Big Data and Society* Burrows and Savage (2014:3) consider Big Data as revealing ‘patterns of social order, movement and engagement with the world – and on such a scale – that [...] nothing less than a fundamental *re-description* of what it is that needs to be explained and understood by the social sciences’ is required. Or, elementally, we begin to see ‘the metricization of social life’ (ibid).

However, as we will go on to demonstrate in relation to personalized datasets and the interpretive mechanisms used to understand them, the scalar reading of Big Data is misleading in that it deprioritizes the agency of the individual and the context in which they produce, interpret and share their data. Burrows and Savage (2014) are similarly concerned about the polarizing effect Big Data has on exploring the social, so recapitulating the value of different types of data and their uses is important in avoiding binary distinctions, beyond the deliberate provocation we have used in coining a seemingly oppositional term like ‘Little Data’.

The second aspect to consider is what we mean by Little Data in the context of broader research connecting everyday activities and forms of tracking and data collection. There is substantial literature that speaks to facets of Little Data as outlined in the introduction, and there are analogies that can be drawn to lifelogging in particular, though our later methodological intervention shows the lines of flight from what is being highlighted in this section.

Relatedly, Lupton claims that ‘digital data are often de-humanised and de-materialised in discourses’ (2018:2). In response to this, Little Data can be understood as ensuring the human and the material are at the forefront of analyses, and the relationship between the two

are fully fleshed out. One site of enquiry where this is happening is the area(s) of ‘lifelogging’ and ‘self-tracking’, which routinely include site-specific data recording.

Dodge and Kitchin (2007:432-33) see lifelogging as a positional form of capture, involving what they term ‘capta’, namely those chosen data that have been selected from the total data available. The key shift in lifelogging is that capta are generated from a first-person perspective, with the individual vicariously consuming their chosen data through ‘intimate technology’ – wearable tech like smart watches, though these were largely unavailable when Dodge and Kitchin proposed the term – enabling a controllable archive (the by-product of which might be that these data are automatically, uncontrollably shared with companies who produce the tech or the databases/cloud storage where the data are stored).

Building on this, Sellen and Whitaker (2010) suggest lifelogging is bifurcated: total capture and situation-specific capture. The former involves a digital record of all experience, determined in part by technological limitations or reach; the latter focuses on rich data within areas intentionally restricted by the agent. Sellen and Whitaker may have been unable to predict the speed and direction that lifelogging would travel, but their conclusion – that access to the archives created by lifelogging will potentially be of limited value in the sense that the real focus is on the collection of data rather than the analysis – has important implications for our argument here, especially when considered in tandem with Selke’s (2016:3) claim that ‘lifelogging promises the possibility of breaking bad habits and turning one’s life for the better’.

Lupton (2014a) sees lifelogging as contributing toward ‘the reflexive monitoring self’, an agglomeration of data about the self, gleaned from intentional recording practices – so situation-specific capture – and by-product collection which allow for reflexive assessment and interpretation ‘working towards the goal of becoming’. Later (2014b:12), Lupton expands on the way that these technologically-mediated practices allow individuals to contextualize their ‘selfhood, embodiment and social relations’ but also, as Selke highlighted, to use ‘the

information they collect on themselves to achieve self-awareness and optimise or improve their lives' (2016:105). This is part of a historical trajectory that Serfaty (2004:5) argues connects contemporaneous forms of personal record-keeping with the tradition of self-improvement connected to Puritanism, where the diary was used as a means of examining the self and finding a path towards betterment.

This narrative of self-improvement accentuates agency and offers a counterpoint to earlier discussions that veer towards technological determinism. It also underscores a practical consideration of the use of data at a personal level. Our interest in this paper is to explore how these personal datasets – which are routinely shared across digital platforms – are used, and a corollary issue pertains to problems associated with highly individualized data collection methods. These data, as we suggest later, are often messy, incomplete or super-specialized, so it is useful to tackle the issue of idiosyncratic and incomplete datasets head on.

The third and final aspect to consider is that mess is not something specific to Little Data, but rather something that researchers routinely experience as part of social research (see Law 2004). There is a long history of addressing and acknowledging what Garfinkel (1967:190) describes as making 'a silk purse out of a sow's ear'. In Garfinkel's case, this is data pertaining to patient records, identifying that there are a variety of 'good organizational reasons for bad records' (187), some of which speak to everything from methodological approaches through to the use of data in particular settings. More recently, Uprichard's (2011) exploration of the gradual shift of recording practices over time firstly demonstrates that category change – and the reasons behind category change – is nothing new (see also Menard 2002) and secondly underscores how interpreting change at key junctures requires the combination of quantitative and qualitative methods where the data is understood in situ without being tidied up: 'leave the data and categories alone, even if this creates more mess in practice!' she argues in conclusion (110). Our contention in this paper is that exceptional

circumstances, like the pandemic, are opportunities to utilize Little Data to add to the picture of both social and classificatory change, albeit this is sometimes short-lived as we will highlight later on.

In summary, there is a notable history of personal record-keeping that has proliferated in recent years through technological change, affordable wearable devices and the default collection of by-product data from these devices. We can consider these developments in the context of the macro-level discourse(s) around Big Data, but equally we can use personalized, idiosyncratic datasets to explore agglomerated responses to specific events, in this case COVID-19, with the necessary observation that data are inherently messy. Next, we consider how we unpack these data, and the practical considerations associated with this.

Methodology

To collect these data, we used the anonymous social media forum *Reddit*, with the search terms ‘I tracked’ and ‘I collected’, aiming to sample the first 50 datasets which included a timeframe of at least six months either side of February 2020, towards assessing habitual movement during changes to daily routines resulting from the coronavirus pandemic lockdowns. Data collection took place during August 2021. We captured 43 datasets (coded as P1-43) before reaching a point of saturation: in 26, impacts associated with coronavirus were clearly visible, in 7 changes occurred that could not clearly be associated with coronavirus, and in 10 no visible change occurred in the sample timeframe at all.

There are several limitations for these data. The first issue, as we have argued, is that it is messy, sometimes incomplete, and always highly subjective. To double down, ‘clean’ sources are rarely ever clean (Garfinkel 1967), and even ‘Big Data’ is exposed to subjective decisions on the ways in which data are collected and aggregated, even as by-products.

Second, we cannot assume that digital data is an accurate record. Certain technologies are routinely employed by the user to collect, store and produce graphics to illustrate this personal data, which can influence practices in and of themselves (Beer 2016). Individuals make subjective decisions about what they do and do not record, and the act of recording and maintaining an up-to-date record itself is one that can encourage or discourage certain actions: if one is recording, there is an incentive to maintain a streak of desirable behaviour.

A third is conversion and equalisation. Data are recorded by different individuals on myriad topics, using different recording methods, against divergent sets of parameters. This is a native unreliability, but can be seen as an opportunity for the purposes of this methodological experiment since our control factor is a certain timeframe, and what we are interested in understanding are these variances in behaviour around that period.

However, there are also opportunities here. For starters, this method counters aspects of recall bias. According to Becker, the majority of social research methods are ‘structured forms of asking people to tell you what happened’ (2017:188), like surveys or interviews, which we should not confuse with the more advantageous ‘being there, seeing for yourself what happens and recording it soon afterward’ (ibid). Here, Becker is referring to *observations* – this method does not allow the researcher to ‘be there’, but it does allow data collectors to record in real time or soon after, potentially reducing inaccuracies that can come from recall methods.

An additional opportunity for these data is that they track movement over time, and they are more longitudinal than some traditional survey methods. Findings can be more accurately tied to certain external pressures during specific timeframes, and the ‘normal’ value of a certain set of behaviours is clearer to see and more easily compared against. A survey approaching habits during a pandemic is limited in that it can measure something new more accurately than it can the zero-level that the data should be compared against.

A challenge to be acknowledged is the distinction between what Dodge and Kitchin (2007) call ‘thin’ and ‘thick’ memory – ‘for people who know the individuals in the archive, context provides emotional connections and further meaning, but for others the archive simply provides a factual account’ (428). Thin memory is data devoid of emotional context; it is easily formalised and can be straightforwardly represented as fact. Thick memory is ‘embedded, emotional, context rich [and] immaterial’ (428). With Big Data, what we have is exclusively ‘thin’, devoid of emotional context and expression – and with records like diaries, we have something exclusively ‘thick’, extremely subjective and sensemaking mostly to the original author or their close relations. Little Data sits somewhere in the gulf between, with markedly personal insights and subjectivities that are most meaningful to the agentic individual – but that has been meticulously and numerically collected, can be aggregated against certain external measures like time or event, and analysed without emotional investment, to tell a story about society.

This translates into both a drawback and an opportunity – these data are not useful without ethnographic input, but an ethnographic approach to Little Data can provide useful insights. Beuving (2020) tells us that there is a conflict between Big Data and ethnography, in that Big Data treats ‘motivation’ as observable preferences revealed in fact, whereas for an ethnographer, ‘motivations’ are social behaviours requiring of interpretation. Furthermore, ‘[Big Data] may be very capable of mapping digital patterns [...] but the claims that it makes with regard to *Verstehen* are problematic from the reflexive viewpoint that characterizes ethnography’ (1633). With Little Data we have mapped patterns, which contain some individual interpretation in the way that they are collected and presented but that require further empathic understanding. For this reason, Beuving argues, following Halford and Savage (2017), that *a hybridized approach is necessary*. This is a suggestion visible also in Pink et al’s *Data Ethnographies* (2016), which should include ‘investigation into the different way that

data is produced by people, the reasons for this, and what habits and routines are reinforced or altered through their engagements with data' (3). In these terms, ethnography is the interpretive work that uses Big Data to tell a story – and ethnography is 'uniquely positioned' (6) to answer the questions that Big Data poses. Halford and Savage label this a 'symphonic' approach in their call for 'wide data', 'thick data' as well as 'Big Data' (2017:1140). Our approach to this important interpretive work draws from Howard Becker.

Firstly, Becker (2007:135) tells us that 'for every form of telling about society, we should look for (as a possibility, not an inevitability) a moral community of makers and users'. Becker starts by considering telling in relation to who is involved in the process. He describes telling as action in a community of people actively engaged in interpretive work, an 'organization [...] who routinely make standardized representations of a particular kind ("makers") for others ("users") who routinely use them for standardized purposes' (2007:7).

In describing the process, Becker delineates between two related groups who are both involved in collaborative works of representation. These representations make sense in relation to those groups, to those who collect data on themselves and those who consume that data posted publicly. In suggesting two groups, there is an implied tendency towards binary distinctions again, but Becker wards us off that, explaining that the categories are more complex, that people are potentially both makers and users depending on the circumstances in that group. Becker characterises the maker/user relationship as a service relationship, where the interests of each group routinely differ. He writes that this tends to be the case 'when makers are professionals who make those representations full time for pay and the users are amateurs who use them occasionally, in a habitual and uninspected way' (2007:26). In our study, interests differ where researchers (users) are interpreting and contextualising the representations of makers (data collectors). Becker suggests that following a process of interpretive work can help in untangling what stories are being told, by whom, and for what

purpose. This process – or ‘work’ – has four stages: selection; translation; arrangement and interpretation (2007:20).

Selection involves a response to the initial conundrum of the necessarily partial. Makers decide upon those aspects of the story that are included and those that are not, resulting in a representation that tells some of a narrative whilst leaving out something else. In terms of understanding the social aspects of this type of work, Becker suggests we look for who finds the choices acceptable and unacceptable (2007:21). For our data, and for us as users, this is visible change in relation to the global pandemic, over a certain time period.

Translation can be thought of as the mapping of one collection of features – those selected in the first type – on to another set of features that are conventionally understood. For our data this is the overlaying a set of previously understood characteristics about the pandemic which we seek to test for their presence against the criteria selected.

Arrangement is required to ensure that the selected and translated elements of a representation cohere. Becker likens this to making a story from seemingly disassociated fragments, a part of the process that is unavoidable as ‘users of representations see order and logic even in random arrangements’ (2007:24). For our data, this would be the process of forming a narrative which presents the selected and translated criteria as relational.

Interpretation is intrinsically connected to the representation itself, where users, through their consumption of and engagement with the representation, construct ‘around themselves a reality out of what the maker has shown them’ (2007:25). Becker discusses this in terms of the constraints it puts on makers in terms of how much their representation can achieve – that users will approach representations with a knowledge of the conventions of a genre, for example, but makers at the same time need to be wary of the variability of user knowledge: in essence, a representation cannot be one thing in this scenario, but rather something that is co-produced through contested divisions of labour (2007:30)

For our data, this is the point at which makers and users overlap. As users, we form a belief around what Little Data can show the reader – this is our interpretation, which forms our argument. But as makers, we present a representation of reality around the COVID-19 lockdowns that the readers of this work, as users, with knowledge of the conventions of social research methodology, will receive varyingly. By positioning Little Data collectors as *makers*, and the researchers as *users*, we are able to decode and communicate these fragments in order to tell a story about habitual changes during COVID-19 lockdowns.

On Coronavirus

Emergent research which has focused on habitual change in relation to COVID-19 has favoured ‘Big Data’ such as Zattoni et al’s (2020) study on increased use of online pornography via Google metrics, or survey methods, such as Izzo et al’s (2021) work on Italian dietary habits. These works have been able to show a spike in the use of internet pornography during stay at home orders (Zattoni et al 2020), worsening diets – such as an increase in alcohol and frozen food consumption and a decrease in exercise - (Izzo et al 2021), weight gain in the Middle East and North Africa (Ismail et al 2020) and in Saudi Arabia (Aljohani 2020), declining sleep quality and higher reported stress in Cyprus (Philippou’s 2020), an increase in house-plant buying in the USA (Campbell, Rihn & Campbell 2020), an impact on salty food consumption in China (Xu et al’s 2020) and changing media habits (Newman & Gallo 2020). Opinion work has also been conducted, using existing consumer trends to predict changes in the way that we purchase food (Martin-Neuninger & Ruby 2020). More holistic research aiming to consider a wider breadth of habits has been conducted using India as a case study, but sampling a short time period of one-month (Chauhan & Shah 2020) neglects to reveal whether habitual changes are long-lived, and recall approaches (Kumar & Dwivedi 2020) may reveal a disparity between what participants say that they did, and what they actually did pre-lockdown.

Each of these works considers a trend in a specific habit, usually health-based, ostensibly through recall methods – whereas here we make a case for Little Data’s ability to see across the spectrum of habits formed and altered, individually collected in real time, to provide a more holistic overview of habitual change during lockdown. By using Little Data, it has been possible to show that 1) common-sense assumptions about the impacts of lockdowns and stay-at-home orders, unemployment and personal circumstantial changes on habits during the February-May 2020 ‘first wave’ of the pandemic in the US, UK and Western Europe, are not reflected in the data and appear in ways that conventional approaches to research design would not have highlighted. Additionally, 2) it has been possible to show that where habits have changed, often in counter-intuitive ways, they have largely been circumstantial, short-lived, and quick to regress back to pre-lockdown norms during the June-September 2020 easing of restrictions.

To offer an initial example, using Little Data it was possible to see more niche habitual changes that researchers may not routinely look for, or ones where agglomeration of data may have obscured the findings, and changes that appear to run counter to media narratives about the impact of COVID (Amrani et al. 2020). P5’s data tracked a steady incline in their chess rating as an avid amateur player, up until late February 2020 when stay-at-home orders interrupted normal routines. The participant did not alter their time spent playing, but it appears that other players finding more time for hobbies briefly skewed the difficulty of chess metrics, so our participant appeared to get worse at their hobby despite a likelihood of no actual change.

As previously highlighted, one of the most popular lifelogging topics is health and wellbeing. When it comes to health, we found a variety of unexpected trends in individual Little Data sets. P9’s data showed a significant increase in workouts in March 2020, despite them having resolved to work out more three months prior at the start of the year, contradicting early narratives of poorer overall wellbeing and health habits during the global crisis (Morgan

and Rose 2020). Despite evidence previously presented to show stress levels and diet changes moving wildly, P12 showed significantly more stability in their defecation, indicating more constancy in their intake during a period of reduced travel, socialising and eating out. Despite supermarkets posting record alcohol sales, P21's data on alcohol consumption for a friendship group of three tapers to almost nothing, repeated across both major UK lockdowns, showing drinking is still a social activity for many, tied to the use of pubs and bars – or further evidence that stay-at-home orders triggered prompt introspection and affirmative change about health and wellness.

P13, who has been recording their personal finances for ten years, was able to save significantly less of their income during the lockdown months, despite record increases in savings, and stock market gains doubling their net worth, contravening a popular media narrative that lockdowns had increased the disposable income of those whose incomes were not impacted (see, for example, Romei 2021).

The effect of time spent apart can be seen in the context of relationships, with P38s text messaging with his partner increasing tenfold during the lockdown months. With these data it is also possible to see, by distribution of messages, the couple staying awake significantly later into the night – presumably with no work, or no commute, to contend with in the morning. This change, driven by lockdown circumstance, reverts immediately once the first pandemic wave has subsided.

Fears of a new baby-boom were widely reported early in COVID coverage due to couples being locked down together (see Puffet and Hall 2021; NZ Herald 2021), but P16, who recorded every orgasm they achieved with his partner, had a 76% decrease in average orgasms per month during the stay-at-home order, citing stress and unrest due to the pandemic. P2 had a significant spike in masturbation during the month of March, as well as the most time recorded all year dedicated to hobbies, but by June this trend had levelled out towards the norm.

P3 also recorded a spike in masturbation during March 2020, at least 25% higher than the next closest month displayed on a graph with a difficult to decipher Y axis. The aforementioned P21, again recording masturbation, showed spiking and then returns to normal during April and November American lockdown periods.

There are always outliers when it comes to changes of statistical significance, but what we see across these data is important nuance that trends established through Big Data can neglect. Of note is the speed at which habitual changes return to normal – something particularly visible in Little Data, since it pinpoints a specific activity or set of activities over time rather than sampling a time period and looking for activities.

Campbell, Rihn & Campbell (2020) found that when new habits were formed online, like plant purchasing, they were more likely to persist post-lockdown, whereas traditional shopping methods did not have the same longevity. What we see with Little Data is that habits and practices altered and formed during the lockdowns of the early-2020 first wave of Coronavirus were exclusively short-lived. In no instance out of 43 cases, 26 of which clearly showed a change during stay-at-home periods, were changes sustained.

Once again, the popular tracking of health and wellbeing showed rapid rebounds. P17's exercise peaked in March/April, and then again in November, mirroring periods where the two most significant sets of US restrictions were introduced but returned to the normal level within four weeks of each instance. P36 also started working out more in March and April, and their workouts had shifted indoors: by July, the quantity and distribution had normalised.

Both P42 and P43 provide data which shows notable declines in their mental health. For P42, 'mood' ranked lowest in April and May, with 'anxiety' and 'emptiness' heightened, but June and July were great months, and a normal distribution of good and bad days was resumed by August, despite pandemic case rates beginning to rise again across the world. For P43, their spikes in anxiety map neatly onto stock market dips – suggesting that they may be

an investor – although these rises last only a couple of days before returning to their normal levels.

With regard to hobbies, P34's time spent playing video games increased by 50% during the March/April 2020 lockdown, but dipped back normal by the end of the summer. P39's comprehensive tracking of every act in every hour of their life – an archetypal 'lifelogger' - shows an emergent hobby of music making during our sample time period, which is indulged regularly for two months before becoming a once-a-month treat through the remainder of the year.

With these data we are able to show marked social and cultural resilience in the face of change-of-circumstances, despite its idiosyncratic and incomplete nature when considered in isolation. The onset of the global pandemic is a visible juncture in personal, introverted data collection, demonstrated in often in eccentric ways. Collecting these data, in its essence, is a form of self-reflection carried out by individuals who are introspective and perhaps seeking to better themselves in some way (Lupton 2018a). These betterments, as well as detriments, were given space to flourish during isolation orders where other passions, responsibilities and pleasures fell dormant. What Little Data has been able to show is that, whether a positive personal goal (such as more exercise or a better diet), a negative consequence of social and economic circumstances (such as declining mental health), or something muddily in-between (like increased masturbation or video gaming), changes to habits which are triggered or enforced by circumstance lack longevity once conditions have normalised.

Discussion and Conclusion

With this method we have been able to show the interim, non-permanent nature of habitual changes made, and dedicated to, during altered life conditions. These 'caretaker' habits, both positively-minded lifestyle goals and commitments and circumstantial detriments were visible

in the vast majority of our data: in all cases these lacked longevity and the robustness to weather a transition back to established quotidian life. This finding is stark when considered against the sort of person who is engaging with their body, fitness, lifestyle and habits as mindfully as life-loggers.

As Beuving (2020:1634) points out – where we get our data from online is highly biased: ‘more highly educated, politically vocal groups are overrepresented on Twitter, for instance, and analysis of Tweets cannot be assumed to represent the entire population’. The anonymity afforded by spaces like Reddit has accommodated the sharing of datasets that could be seen as socially embarrassing, or endeavours that have failed, many of which we have documented in our analysis, but it is safe to assume that particularly motivated, or at least introspective and relatively self-aware, individuals are logging their practices (Lupton 2014a), and individuals who feel successful in this pursuit are more likely to post them publicly. In this case study where changes have been revealed as fugacious – it is credible that this trend would be exacerbated were the sample widened.

Our contribution is to propose this approach to complement to more established methods like self-report recall surveys, as a way to assess more subtle, unpredictable alterations in behaviour during specific events that disrupt aspects of everyday experience. Pink (et al 2016:6) points to the nature of these data as rooted and concrete in time: ‘self-tracking data is visualized as cut-through representations of human activity, and there is made static [...] these visualisations are ongoingly produced, they stand for moments in a process, not end results’. Even though these datasets span months or years, in their individuality they are murky when considered alone. As a result of this, our approach has been one of stacking datasets on myriad topics, revealing longitudinal findings where the timeframe (COVID-19) is controlled for, rather than the participant or subject matter.

To achieve this, our quantitative datasets also required ethnomethodological translation and qualitative interpretation, for which we can draw on Walter Benjamin (2009:31) for additional clarity:

Translatability is an essential property of certain work. That is not to say that translation of them is essential as such; what it means is that a certain significance possessed by the originals finds expression in their translatability. Clearly a translation, no matter how good it is, can never mean anything so far as the original is concerned. Yet thanks to the original's translatability, the translation is very closely connected with it. Indeed, the connection is all the closer for the fact that it [the connection] no longer means anything [...]. It may be called a natural connection – more precisely, in fact, a living connection.

Here Benjamin, on the task of the translator, gives us a useful description for the analysts' dual role as both maker and user. The original datasets possess significance for their maker, but as users we are able to provide a translation which – although closely connected to the original – has taken on a new meaning. Benjamin's remarks on the 'connection' between the translation and the original no longer meaning anything reflect the translation's ability to stand alone, to provide value in its ability to speak to a wider audience and, for certain qualities of the data, to 'find expression' in a new form. This translatory process can also be understood as the difference between self-reflection and social scientific sense. Methodologically, both the original datasets and their interpretation here are valid to makers and users are essentially different. Whilst the insights present in these Little Data sets are not completely lost-in-translation when viewed individually, their significance can be understood, and finds expression, through translation.

This is where we can understand Little Data in comparison to big, as well. Little Data acknowledges the value of the original maker's idiosyncrasies in the process of translation, rather than seeking to blend them together by subsuming them in volume and scale. In a Big Data scenario, with only a small handful of cases referencing something as distinctive as masturbation, for example, no conclusions could be drawn. Alternatively, with a million cases referencing masturbation, only masturbatory habits can be analysed. Here, when coupled with a multitude of other habitual practices, both sexual and non-sexual, we can translate them into a useful narrative about something entirely different.

Little Data also builds on the sort of tension previously highlighted in the literature review and by Rawassizadeh (2009) in relation to life-logging: it operates at the intersection between individual control and identity management, and the trends towards agglomeration routinely associated with Big Data, the primary difference being the agency afforded the maker in terms of what is shared and when (here the context being temporary lifestyle changes mitigated by a pandemic). Where Big Data can be accused of indiscriminatory collation, Little Data allows for a nuanced reading of datasets voluntarily shared as part of sense-making actions.

Given that Big Data routinely involves collecting generally unfiltered information (O'Hara, Tuffield and Shadbolt, 2008), Little Data can be seen as contributing to the translating of these data. In line with Becker's process of 'telling', we can see clearly how those who share aspects of their lives publicly select their data, translate their data for a particular audience and arrange that data in a way that is legible; the act of 'interpretation' is then tacitly a collaborative enterprise on the part of those sharing the data (makers) and those responding to the data online (users) - or indeed those combining smaller datasets for a journal article. The outcome, when combined with other idiosyncratic datasets is a viable snapshot of societal responses to particular events that disrupt the everyday.

There is, of course, a wider applicability for Little Data in the context of social, political and cultural upheaval. Pandemics have not been regular occurrences, although there are predictions for their increase in frequency (Penn 2021). In terms of the applicability of this approach for the future, there are opportunities for understanding habitual changes at significant waypoints in global circumstances more broadly, such as during periods of economic growth or recession. For example, during periods of recession, practices like smoking (Novo et al 2000) have been successfully measured using surveys – and other consumption dynamics have been shown between boom and recession using qualitative interviews (Sarmiento et al 2019) – both of which can be said to suffer from recall bias, and both of which could be successfully approached through Big Data techniques, which might focus on purchasing habits for example. However, habits that are not consumptive in nature are more challenging to identify using Big Data, and the ways in which more personal behavioural modification intersect with each other at points of change can be illuminated with Little Data. There is also the potential to study regular seasonal change in habits, allowing the researcher to access those more sensitive practices which are not regularly shared and/or not directly consumed.

Other significant waypoints could be natural disasters, as a way to study usually noiseless variations to behaviour brought on by climate change. This would require the researcher to pinpoint certain locale in order to understand the social impacts of things like floods, wildfires or the migration associated with regional heating. The majority of habitual and behavioural work that exists currently on climate change focuses on warming mitigation habits (like recycling, or using cleaner fuels) (see, for example Leger & Pruneau 2012), or on easily observable or survey-able purchasing decisions, like travel and holidays (see, for example, Schwirplies & Ziegler 2016). More personal methodologies, like diary methods, lifelogging, and our approach to these two as ‘Little Data’, provide opportunities to understand

where macro issues intersect the micro minutiae of people's more mundane, yet still meaningful, behaviours.

References

- Aljohani, N. (2020). The Effect of the Lockdown for the Coronavirus (COVID-19) Pandemic on Body Weight Changes and Eating Habits in Saudi Arabia. *Journal of the Saudi Society for Food and Nutrition*. 13(1) pp103-113.
- Amrani, I; Payne-Frank, N; Lamborn, K and C Phillips (2020) A new normal: work, life and balancing it all during coronavirus. In *The Guardian*. Available at <https://www.theguardian.com/commentisfree/video/2020/jul/22/a-new-normal-work-life-and-balancing-it-all-during-coronavirus-video>. Accessed 15th Oct 2021
- Becker, H. (2007). *Telling about Society*. London: The University of Chicago Press.
- Becker, H. (2017). *Evidence*. London: The University of Chicago Press.
- Beer, D. (2013). Playing with digital byproduct data: an indicative example. Available at <https://necus-ejms.org/playing-with-digital-by-product-data-an-indicative-example>. Accessed 8th September 2021
- Beer, D. (2016). *Metric Power*. London: Palgrave Macmillan.
- Benjamin, W. (2009) *One-way Street and Other Writings* (trans JA Underwood). London: Penguin.
- Beuving, J. (2020). Ethnography's Future in the Big Data Era. *Information, Communication & Society*. 23(11) pp1625-1639.
- Brickman, S. (2021) If Covid's 'new normal' makes you even more anxious than before, you're not alone. In *The Guardian*. Available at <https://www.theguardian.com/commentisfree/2021/oct/01/covid-new-normal-anxious>. Accessed 17th Nov 2021

- Burrows, R. and Savage, M. (2014). After the crisis? Big Data and the methodological challenges of empirical sociology. *Big data & society*, 1(1), pp1-6.
- Campbell, B.; Rihn, A. & Campbell, J. (2020). Impact of the Coronavirus Pandemic on Plant Purchasing in Southeastern United States. *Agribusiness*. 37 pp160-170.
- Chauhan, V. & Shah, H. (2020). An Empirical Analysis into Sentiments, Media Consumption Habits, and Consumer Behaviour during the Coronavirus (COVID-19) Outbreak. *Parakala (UGC Care Journal)*. 31(20) pp353-378.
- Dodge, M. & Kitchin, R. (2007). ‘Outlines of a World Coming into Existence’: Pervasive Computing and the Ethics of Forgetting. *Environment and Planning B: Urban Analytics and City Science*. 34(3) pp431-445
- Favaretto, M., De Clercq, E., Schneble, CO. & Elger, BS. (2020). What is your Definition of Big Data? Researchers’ Understanding of the Phenomenon of the Decade. *Plos One* 15(2) available at:
<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0228987>
- Garfinkel, H. (1967). Good Organizational Reasons for Bad Clinic Records. In H. Garfinkel (ed): *Studies in Ethnomethodology*. Englewood Cliffs, NJ: Prentice-Hall pp186–204
- Halford, S. & Savage, M. (2017). Speaking sociologically with big data: Symphonic social science and the future for big data research. *Sociology* 51(6) pp1132–48.
- Ismail, L.; Osaili, T.; Mohamad, M.; Marzouqi, A.; Jarrar, A.; Zampelas, A.; Habib-Mourad, C.; Jamous, D.; Ali, H.; Sabbah, H.; Hasan, H.; AlMarzooqi, L.; Stojanovska, L.; Hashim, M.; Obaid, R.; ElFeky, S.; Saleh, S.; Shawar, Z. & Dhaheri, A. (2020). Assessment of Eating Habits and Lifestyle during the Coronavirus 2019 Pandemic in the Middle East and North Africa Region: A Cross-Sectional Study. *British Journal of Nutrition*. 126(5) pp757-766.

- Izzo, L.; Santonastaso, A.; Cotticelli, G.; Federico, A.; Pacifico, S.; Castaldo, L.; Colao, A. & Ritieni, A. (2021). An Italian Survey on Dietary Habits and Changes during the COVID-19 Lockdown. *Nutrients*. 13(4) p1197
- Koopman, C. (2019) *How We Became Our Data: A Genealogy of the Information Person*. Chicago, Ill: University of Chicago Press
- Kumar, M. & Dwivedi, S. (2020). Impact of Coronavirus Imposed Lockdown on Indian Population and their Habits. *International Journal of Sciences and Healthcare Research*. 5(2) pp88-97.
- Law, J. (2004). *After method: Mess in social science research*. London: Routledge.
- Leger, M. & Pruneau, D. (2012). Changing Family Habits: A Case Study into Climate Change Mitigation Behaviour in Families. *International Journal of Environmental Education* 2(2) pp77-87.
- Lupton, D. (2014a) Beyond the Quantified Self: The reflexive self-monitoring self. Available at <https://simplysociology.wordpress.com/2014/07/28/beyond-the-quantified-self-the-reflexive-monitoring-self>. Accessed 9th September 2021
- Lupton, D. (2014b). Self-tracking Modes: Reflexive Self-Monitoring and Data Practices. Paper for the 'Imminent Citizenships: Personhood and Identity Politics in the Informatic Age' workshop, 27 August 2014, ANU, Canberra
- Lupton, D. (2016). The Diverse Domains of Qualified Selves: Self-Tracing Modes and Dataveillance. *Economy and Society*. 45(1) pp101-122.
- Lupton, D. (2018). How do data come to matter? Living and becoming with personal data. *Big Data & Society*, 5(2), pp1-11
- Lupton, D. (2019) *Data Selves*. Cambridge: Polity Press

- Makdisi, G. (1986). The diary in Islamic historiography: Some notes. *History and Theory*, 25(2), pp73-185.
- Martin-Neuninger, R. & Ruby, M. (2020). What does Food Retail Research tell us About the Implications of Coronavirus (COVID-19) for Grocery Purchasing Habits? *Frontiers in Psychology* 11. Available at:
<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7292029/>
- Menard, S. (2002) *Longitudinal Research*. London: SAGE
- Morgan, C and Rose, N (2020) A mental health epidemic caused by coronavirus can be avoided if we act now. In *The Independent*. Available at
<https://www.independent.co.uk/voices/mental-health-coronavirus-covid-nhs-b885840.html>. Accessed 12th Oct 2021.
- Newman, N. & Gallo, N. (2020). *Daily News Podcasts: Building New Habits in the Shadow of Coronavirus*. The Reuters Institute for the Study of Journalism.
- Novo, M., Hammarström, A. & Janlert, U. (2000). Smoking Habits – a Question of Trend or Unemployment? A Comparison of Young Men and Women between both Boom and Recession. *Public Health* 114(6) pp420-463.
- NZ Herald (2021) Lockdown baby boom. Available at
<https://www.nzherald.co.nz/lifestyle/lockdown-baby-boom-number-of-births-in-2021-highest-since-2015/CTGSOOZV62E462S7LNN4K3JYHQ/>. Accessed 11th Nov 2021.
- O’Hara, K.; Tuffield, M.; Shadbolt, N. (2008). Lifelogging: Privacy and Empowerment with Memories for Life. *Identity in the Information Society*. 1(1) pp155-172.
- Penn, M. (2021) Statistics Say Large Pandemics Are More Likely Than We Thought. Available at <https://globalhealth.duke.edu/news/statistics-say-large-pandemics-are-more-likely-we-thought>. Accessed 4th Nov 2021.

- Philippou, E.; Demetriou, C; Heraclides, A.; Mosquera, M.; Quattrocchi, A; Derlagen, A.; Pinnawala, N. & Kolokotroni, O. (2020). The Effects of Lockdown on Lifestyle Habits of Adults during the Coronavirus Pandemic in Cyprus. *European Journal of Public Health*. 30(5) pp210-210
- Pink, S., D. Lupton, M., Berg, P., Dourish, A., Dyer, V. Fors, E., Gómez Cruz, H., Horst, P., Lacasa, J., Postill, S., Sumartojo, & Witkowski, E. (2016). *Data Ethnographies 1: personal data in an uncertain world*.
- Puffett, N and Hall, R (2021) Lockdown baby boom may be on the way as NHS antenatal bookings rebound. Available at <https://www.theguardian.com/lifeandstyle/2021/apr/20/lockdown-baby-boom-on-the-way-as-nhs-antenatal-bookings-rebound>. Accessed 5th Nov 2021.
- Rawassizadeh, R. (2009). Towards Sharing Life-Log Information with Society. *Behaviour and Information Technology*. 31(11) pp1057-1067.
- Romei, V (2021) Global savers \$5.4tn stockpile offers hope for post-Covid spending. In *The Financial Times*. Available at <https://www.ft.com/content/8cbfe40d-1ce1-4dc6-bcb2-1314b77b9443>. Accessed 9th Oct 2021.
- Sarmiento, M., Marques, M. & Galan-Ladero, M. (2019). Consumption Dynamics During Recession and Recovery: A Learning Journey. *Journal of Retailing and Consumer Services* 50 pp26-234.
- Schwirplies, C. & Ziegler, A. (2016). Adaptation of Future Travel Habits to Climate Change: a Microeconomic Analysis of Tourists from Germany. *Tourism Economics*. 23(6) pp1275-1295
- Sefarty, V. (2004) *The Mirror and the Veil*. New York: Radopi.
- Selke, S. (ed.) (2016). *Lifeloggging: Digital self-tracking and Lifeloggging-between disruptive technology and cultural transformation*. Berlin: Springer.

- Sellen, A. & Whittaker, S. (2010). Beyond Total Capture: A Constructive Critique of Lifelogging. *Communications of the ACM*. 53(5) pp70-77.
- Uprichard, E. (2011). Dirty data: longitudinal classification systems. *The Sociological Review*, 59 pp93-112.
- Woodcock, A. (2020) *Boris Johnson says Britain cannot go back to 'old normal' after coronavirus*. In *The Independent*. Available at <https://www.independent.co.uk/news/uk/politics/boris-johnson-conservative-conference-coronavirus-building-back-better-b836324.html>. Accessed 17th Nov 2021.
- Xu, Ying.; Li, Z.; He, X.; Ma, Y.; Cai, F.; Liu, Z.; Zhao, R.; Wang, D.; Guo, Y. & Chen, J. (2020). The Association Between Subjective Impact and the Willingness to Adopt Healthy Dietary Habits after Experiencing the Outbreak of the 2019 Novel Coronavirus Disease (COVID-19): A Cross-Sectional Study in China. *Aging*. 12(21) pp20968-20981
- Zattoni, F.; Gül, M.; Soligo, M.; Morlacco, A.; Motterle, G.; Collavino, J.; Celeste Barneschi, A.; Moschini, M. & Dal Moro, F. (2020). The Impact of COVID-19 Pandemic on Pornography Habits: A Global Analysis of Google Trends. *Your Sexual Medicine Journal* [online].