5

SEEING THE INVISIBLE ALGORITHM

The practical politics of tracking the credit trackers

Joe Deville and Lonneke van der Velden

> To do a sociology of the invisible means to take on the erasing process as the central human behaviour of concern, and then to track that comparatively across domains. This is, in the end, a profoundly political process, since so many forms of social control rely on the erasure or silencing of various workers, on deleting their work from representations of the work.
>
> *(Star, 1991: 281)*

Introduction

Susan Leigh Star (1991) captures a central concern of much politically sensitive academic practice: how to make that, which is rendered invisible, visible. There is a politics to that which is unseen, in which forms of what she calls 'social control' become tied to the erasure of the practical activities – the forms of work – that go into making and stabilising the domains of the visible.[1]

We follow the spirit of Star's call for a sociology of the invisible, but our focus is not on human labour, but the labour of machines: the automated, unseen, digital work undertaken by 'trackers' (other terms include 'bugs', 'pixels', 'tags'). These are online data gathering tools, many provided by third party providers, activated when a user visits a particular website. They form part of what has been called the 'invisible web'. Many of us are familiar with the way in which cookies are placed in our browsers. Cookies, however, are only one of a range of tracking devices that are deployed by tracking and advertising companies. When they are activated, content provided by a third party is loaded onto the website in question (even if it remains unseen; in some cases, this content can even import new trackers, so called '4th party content'). In the case we will consider in this chapter, the data gathered by these devices are being turned towards a quite specific form of algorithmic calculation (Gerlitz and Helmond, 2013; Helmond, 2013).

There is more we can draw from Star. In pursuing the sociology of the invisible, Star calls for a process of *comparative tracking*. We have followed this injunction by seeking to track and compare the tracking devices that are tracking us – a project of 'tracker tracking' (Castellucia, Grumbach, and Olejnik, 2013; Gerlitz and Helmond, 2013; Koot, 2012; Tran *et al*., 2012). This has involved not only comparing empirical objects, but also comparing changing 'tracking toolkits'.

Our set of empirical objects cluster around a phenomenon that we call 'digital subprime'.[2] This is a new variant of online consumer credit lending grounded in the exploitation of diverse data to predict the behaviour of individuals, and, in particular, the likelihood of subprime borrowers repaying subprime loans.[3] Its work can be seen to exhibit a degree of continuity with similar activities in a range of different online settings. Amazon, Facebook, Google, Netflix and many more besides are each routinely acquiring, analysing and making use of more or less willingly released personal online data (Bollier, 2010; van Dijck and Poell, 2013; Hoofnagle *et al*., 2012).

Though we will touch on some of the issues that arise from the proliferation of this new form of credit, the chapter's principal focus is on how we might hold this object steady in the first place. Drawing on the results from a pilot project, its focus is on how we might seek to understand with more precision, and *from the outside*, the work being done by forms of online algorithmic calculation. This, then, is an analysis of some of the practical politics entailed by the production of knowledge about algorithmic behaviour. What does it mean to try to understand an (proprietary) algorithm from the outside? What are the limits and opportunities? What calculative practices lend themselves to being rendered visible, which cannot? Through which different registers can questions of visibility and transparency be articulated? Our answers to these questions sit in dialogue with the emerging field interested in the politics of knowledge production associated with a variety of digital methods (see Marres, 2012; Borra and Rieder, 2014; Weltevrede n.d.).

Encountering digital subprime

Readers in the UK will likely be well aware of the UK's largest digital subprime company, even if they may not recognise it as such. It is Wonga, the most controversial of the UK's growing number of providers of so-called 'payday loans' or 'short-term credit' (the preferred industry term). These loans are relatively low in value – £400 is the current Wonga maximum for first time customers (Wonga, 2014a) – and are usually due to be repaid within a month. They are also expensive, with a Wonga loan currently repayable at 1,509 per cent APR. This, however, is a lot less than it used to be; a cap on interest rates introduced at the start of 2015 has pushed down the cost of its loans from an eye-watering 5,843 per cent. Before this, it was by far the most expensive UK payday lender, which led to much criticism. Most famously, the current Archbishop of Canterbury drew parallels to historical practices of usury (Mendick, 2012).

Its historically higher than average APR aside, there are two further elements that make Wonga near-unique in the UK and especially relevant to our concerns here.[4] The first is its proprietary in-house credit scoring system. Wonga claims that by sorting through up to 8,000 different data points (Wonga, 2014a), it is particularly good at sorting borrowers who will repay from those who will not. The second is Wonga's speed. Credit decisions are fully automated, with cash being delivered into customers' bank accounts five minutes after approval (Wonga, 2014b). This has an important consequence: Wonga is dependent on data that is available more or less instantly.

Given the overall themes of this edited volume, we will not detail here why Wonga might be interested in such a large number of data points. Suffice it to say that, when it comes to its particular form of risk-oriented algorithmic calculation, it is important that the data it is able to access is as plentiful and diverse as possible. It is also worth noting at this point the specific form of credit-consumption that is engendered by such sites. All payday lenders, including Wonga, depend, like 'mainstream', less 'fringe' forms of credit,[5] on a user's credit rating(s).[6] However, payday lenders' very positioning as less concerned with its users institutionalised credit standing marks this territory out as somewhat distinct from their less controversial creditor cousins (credit card issuers, banks issuing personal loans, etc.). To a degree, then, this renders the credit score and its ongoing management (e.g., Langley, 2014; Marron, 2009) a less relevant concern for borrowers using payday lending services. More significant for the present analysis is the fact that, if potential borrowers were minded to look into the basis of Wonga's particular method of credit assessment, they might quickly realise that their credit score matters even less than is the case with some of Wonga's payday lender rivals. For, while Wonga buys credit reference data, and supplements this with other commercially available data,[7] it claims its own scores, only *partially* composed from this data, are 'unbelievably' and 'dramatically' more predictive than those provided by third parties (Shaw, 2011). This seems to suggest that Wonga, a relatively small, relatively new company, is – in this particular section of the consumer credit market at least – doing better than an industry that has spent decades trying to master such methods.

As with many other payday lending websites, Wonga's homepage is dominated by a twin device designed to facilitate customer calculation: two 'sliders', whose movement affect both the size of the loan and its duration. These sliders and, specifically, the way they moved during the phase in which we were conducting our initial research, provided our first way into rendering visible some of invisible processes Wonga was and still is undertaking that make it, in fact, quite different from many of its rivals. Using some rough experimentation, it was, when we began our investigations in 2013, possible to demonstrate that the *starting* position of Wonga's sliders – that is the position that the sliders are set at when a user visits the site – was not constant (we would encourage readers to experiment!). Their starting point was affected by a number of conditions, relating to data that is released by the visitor, perhaps without them fully realising it. This is what one could call 'leaked' data (Chun, 10 October 2013; Rogers, 2013). One major factor in this variance was the particular browser being used – so, for instance, Firefox as compared to Internet Explorer. This might seem a highly mundane variable; however, analysts have shown how it is in fact possible to map a user's choice of browser with their particular spending habits (Qubit, 2013; we will return to this later). Another seemingly important driver of this variance was the user's IP address, which provides a rough indication of the user's location. And finally, repeat visits to the site, measured using cookies, had an effect on the repayment time bar, with more visits pushing it downwards.

This has now (early 2015) changed. The sliders no longer move pre-emptively. We can speculate that Wonga was experimenting with a novel technology and, for whatever reason, decided not to pursue this further. Irrespectively of this, by opening a brief window into Wonga's algorithmically-driven processes, we have succeeded in rendering an aspect of its 'behind the scenes' practices visible. We can draw tentative conclusions about what we were observing. We can be reasonably confident that this shifting behaviour revealed that Wonga was undertaking a process of customer 'segmentation', by which people are sorted out according to their expected consumption behaviour (Seaver, 2012; Zuiderveen Borgesius, 2014). However, because of the fact that the algorithm driving this movement was and still is hidden from view, we cannot draw firm conclusions about the exact reasons for this segmentation. We can provide some hypotheses: first, that the starting position of the loan amount could be understood as a kind of base level loan for a user – an algorithmically driven, rough and undeclared confidence vote in the potential borrower. It would follow that Wonga was trying to encourage broadly 'riskier' borrowers, a measure arrived at by combining various elements of their leaked data, not to ask for too much, therefore potentially improving the chance of their application being successful. The movement in the repayment time, in turn, would be aimed at making loan being offered ever cheaper. We might hypothesise in this case that the site was experimenting with enticing hesitant borrowers, those who were making multiple visits to the site, but not following through with an application, to go ahead. In effect, what we would be seeing here, then, was the offer of a different credit product to different types of people, potentially at different points in a purchasing/borrowing trajectory.

This analysis of the condition-dependent movement of the slider, then, while not providing access to the full epistemological underpinnings of the algorithm, does render visible some of its *ontological effects*. Experimentation is thus one way of chipping away at the opacities surrounding certain forms of algorithmic apparatus. While in some domains of academic social research the art of deploying this tool has been lost, when it comes to algorithmic opacity, it may be crucial. Furthermore, even if precisely how the various data being collected are analysed and deployed is currently opaque (although our experiments are ongoing), and if the registers of visibility of these processes are changing over time, we have been able to show that certain online, leaked data – including location, as measured by

IP address, frequency of visit, and browser type – are being used and combined in order to stand as very quick, very rough proxies for an individual.

We may also venture a stronger claim: that this kind of mundane, 'leaked' data is of interest to Wonga not only when deciding what credit products to offer to potential borrowers, but also in its process of conducting credit assessments once an application has been made. We have had this confirmed to us by an industry source, familiar with Wonga's systems.[8] Wonga are not alone in this respect. It also has an international rival called Kreditech, which is preparing to launch new payday loan sites in Australia and across Eastern Europe and Central and Southern America, to add to existing sites in the Czech Republic, Poland and Spain (Kreditech, 2013a). It specialises in what it calls 'big data scoring' and claims to assess potential borrowers by even more data points than Wonga (10,000) (Kreditech, 2013b), using data derived from "social networking sites and online tracking" (Kreditech, 2013c). Similar companies are ThinkFinance (USA; UK) and Zestfinance (US). The mantra of Zestfinance encapsulates the promise of data for all of these organisations: that 'all data is credit data'. Its approach is summarised in a talk given by one of its co-founders, Douglas Merril (it should be noted that, different to Wonga, Zestfinance claim that their use of data can help bring down the costs of subprime borrowing):

> It turns out that there are hundreds of sources of data, trivially available on the net. And thousands if you include things like web-crawls etc. And if your view is that all data is credit data, you build a piece of mathematics, or in our case a whole bunch of mathematics, that consumes thousands of data points. And of those thousands many are missing, many are wrong, etc, but regardless you build a score. And suddenly you build a score that allows you to figure out people who are maybe not quite good enough to get a subprime credit card, but are a way better credit risk than the payday loan guys. So instead of offering them a 700% APR borrowing [sic], you can offer them something in-between.
>
> *(Merrill, 2012)*

Different goals aside, what these sites share is the scraping of vast amounts of data that could be tied to the *identity* of a potential borrower, which is duly filtered and then acted upon through apparatuses of algorithmic calculation in order to make predictions about the *behaviour* of those borrowers.

In examining these largely opaque practices, however, what tools might be used in addition to experimentation? How might we situate their methodological promise?

Registers of (in)visibility and the 'Tracker Tracker'

Even before the revelations that followed Edward Snowden's release of National Security Agency's files, online tracking and behavioural profiling by online corporations started to emerge as a public concern.[9] Users have become accustomed to cookies being placed in their browsers (whether with their consent or not), to personalised search results and to advertisements for particular products following them as they browse online.

In a sense, then, the ubiquity of online tracking has already achieved a measure of visibility amongst even a non-expert online audience. Counterstrategies range from the simple (e.g., deleting browser caches, using ad blockers or browser's in-built private browsing facilities) to the more sophisticated (e.g., using VPNs and other online anonymity software (e.g., Tor)), to protect or obfuscate communication (Raley, 2013).These strategies are less concerned with rendering the invisible visible, than with blocking the effects of technologies assumed to be operating in the background, unseen.

There is a further strand within user-led counterstrategies whose explicit focus is on making visible, in real time, the much more specific processes of online tracking. A number of tools and browser plugins have been developed, with the focus not on the achievement of anonymity *per se*, but rather on changing the online browsing experience, so as to amplify the user's *awareness* of the tracking technologies that are in operation, while also potentially giving the user the option of impeding their operations.[10] These tools pull an invisible market of data sharing direct to the screen, while expecting the user to act on and play with this information.[11] They therefore engage in a particular repertoire of transparency that assumes that getting people to see these third party connections will stimulate a different info-aware behaviour. We will not dwell here on the question of whether these tools are successful in their aims. We are interested, instead, in what the tools can do for the inquisitive social scientist. For, given their attention to making visible the specificities of online tracking, they also can be turned into tools for keeping track of trackers; into 'tracker trackers', in other words.

In our research, we have drawn on one tool in particular, called Ghostery.[12] Ghostery is a tracker detector, owned by a company called Evidon, which, according to their own framing, "shows you the invisible web".[13] After a user installs the plugin in their browser, it provides a drop-down display listing the third party trackers that Ghostery detects in the web page being visited. It also provides the user with the option to stop these trackers from running – to block them, in other words. Ghostery works by consulting a 'library of trackers' that Evidon has built up, in part by some of its users having opted to share the trackers detected during their browsing sessions to it. At the time of writing, the library contains information about more than 26 million websites, 1,600 companies and 4,100 different types of trackers.[14] On the basis of its data, Evidon is able to rank the most frequent occurring trackers on the web, which it visualises as a periodic table of trackers that updated on a bi-weekly basis (Figure 5.1).[15]

For the social scientist, the appeal of such repositories is that they contain rich information about a practice, access to which would otherwise be highly technically challenging. While third party trackers can be detected manually (Koot, 2012), doing this on a large scale requires both considerable time and a

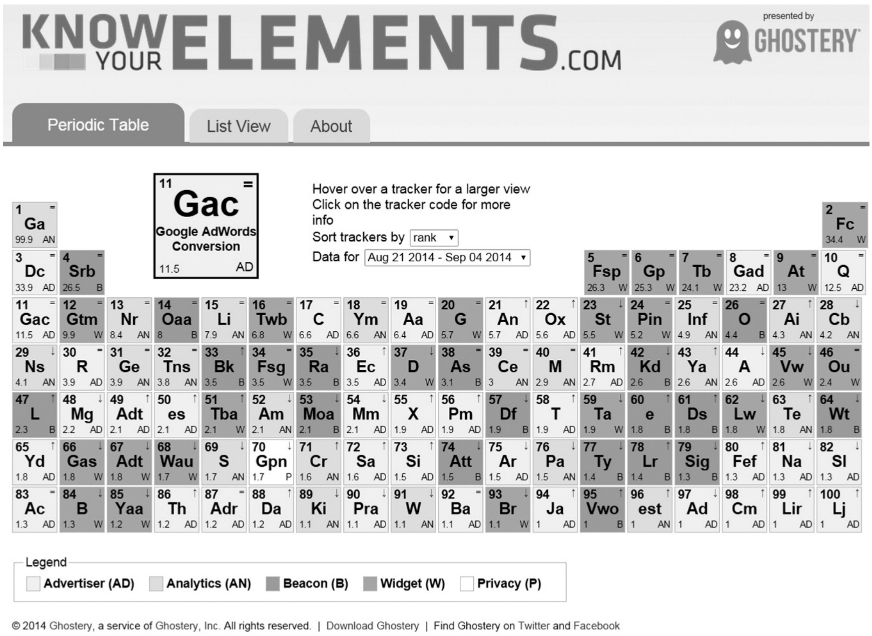

FIGURE 5.1 Know your elements: Ghostery’s tracker ranking visualisation

high degree of programming skill. Moreover, given the sheer volume of trackers and their rapidly changing configurations, an individual researcher would struggle to keep track of them. The use of such technologies is not without its issues. In using such tools, researchers are delegating part of the assembly of the empirical space s/he is intervening in to a third party (Marres, 2012). Moreover, they produce very particular epistemological affordances, being situated in particular device cultures (Rogers *et al*., 2013). In our case, this is partly a result of the fact that Evidon’s database is dependent on being populated by Ghostery’s user population. We are also dependent on how information is indexed in Ghostery’s library. For instance, we are interested in identifying commonalities amongst different ‘types’ of data being collected by these trackers. These ‘data types’ are categories including browser information, date and time, demographic data, hardware type, page views, and IP address. Their categorisation is the result of a manual process, in which Evidon employees group trackers’ according to their publicly accessible privacy policies. The transparency that Ghostery enacts is, therefore, inevitably partial and mediated.

The particular way we make use of the Ghostery plugin has emerged as part of a collaborative project undertaken with the Digital Methods Initiative (DMI) at the University of Amsterdam.[16] The DMI repurposes existing web devices for social and cultural research (Rogers, 2013). Examples of digital methods projects include scraping Google and Twitter for social data (Marres and Weltevrede, 2013), edit-scrapers in Wikipedia in order to follow controversies (Rogers and Sendijarevic, 2012), and software for co-link analysis to map the composition of issues on the web (Marres, 2005).

One of the tools developed by DMI researchers and developers that we have drawn on (and contributed to the ongoing development of) is the 'Tracker Tracker'. The Tracker Tracker repurposes Ghostery's method of detecting and ordering trackers: it inspects web pages for particular traces of trackers (scripts) and compares them with Ghostery's database. In addition, its interface allows the researcher to insert multiple websites (urls) for inspection. The results of this analysis are made available to users in a spreadsheet that facilitates the systematic comparison of the trackers found. This can reveal which websites share similar third party trackers and also which companies are main actors within this – usually – invisible 'fabric of the web' (Gerlitz and Helmond, 2013: 1349).

Early research into the accuracy of the Tracker Tracker has been promising: a recent study (Van der Velden, 2014) used the tool to simultaneously research a dataset of 1100 URLS being investigated by another researcher, using a method for automatic browsing (Koot, 2012). The results were found to be almost identical.[17] Other related digital methods projects have, by using Ghostery's library, been looking more deeply into the trackers themselves with respect to what kind of data these technologies collect. These research projects have tried to 'characterise' the trackers, for instance in a Glossary of Trackers that, inspired by the life sciences, maps the behaviour of trackers.[18] Our study builds on such work, but differs from it in a number of crucial respects, as we will now proceed to outline, while looking at how we coupled it to particular strategies of visualisation.

Rendering visible digital subprime's tracking toolkits

When trying to render visible commercial forms of algorithmic calculation, a major challenge is their proprietary status. The opacities of algorithmic calculation are therefore deeply entwined with the logics of intellectual property. We have sought to overcome this challenge by moving from a focus on individual companies to a generalised industry. We are less interested in the specifics of individual companies' algorithms than in more generalised *tendencies* in the basis of algorithmic calculation across the industry.

The basic logic and promise of algorithmic calculation and econometric analysis are well understood by researchers: that, through the analysis of large datasets, individual variables, or combinations of variables, may be found that hold predictive power (e.g., Deville, 2012; Seaver, 2012). When it comes to digital subprime, what is less well understood is what the basis for such calculations are. Put simply, what *kinds* of data are digital subprime sites interested in?

Our pilot dataset consisted of seven websites, three owned by Wonga – including in the UK, Canada and South Africa – three European sites run by Kreditech – and one in the US run by Zestfinance. This is a small dataset, compared to other tracker research (for instance, Gerlitz and Helmond (2013) and Van der Velden (2014) use about a thousand sources). However, in this project the Tracker Tracker performs a different function: instead of looking into the larger networks and the actors within them, it acts as a way of rendering visible tendencies and commonalities in the tracking work being done by digital subprime sites, which, in turn, can provide the stimulus for delving deeper into the role specific trackers are playing. In our case, we also checked the tool's output by running manual sweeps on each of the websites, including collecting trackers present on the application page of a particular site, which may be missed by the Tracker Tracker's analysis.

The first step in our project was to measure the trackers on a dataset of digital subprime websites at two separate points in time (July and November 2013) and then to simply count the 'kinds of data' that we encountered. To do so, we employed Ghostery's categorisation, as outlined above. The results (Figure 5.2) reveal that, for these seven websites, trackers involved in the collection of browser information featured most prominently (60 times) as a collected data type. We already knew from our initial experiments with Wonga that different browser types were affecting the slider position. Here we see an interest in

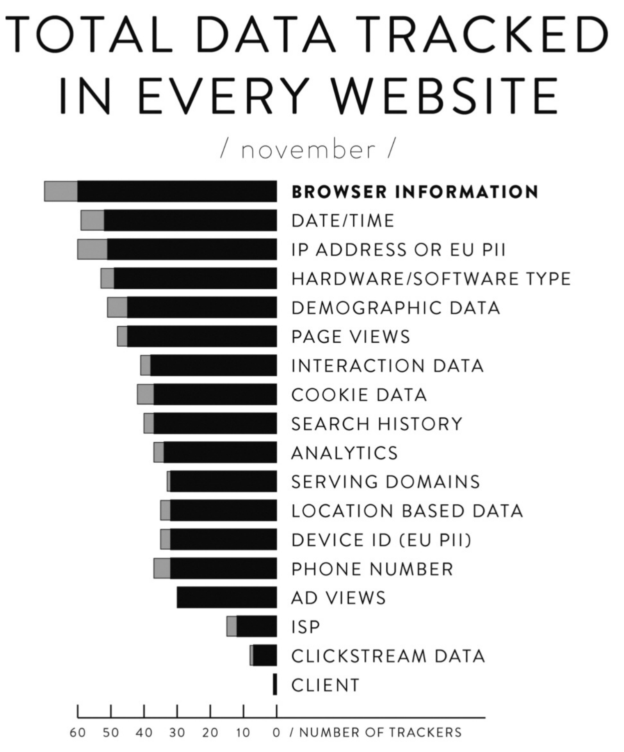

FIGURE 5.2 Tracker tracking preliminary results, July vs. November, 2013

browser information across the sector. Other prominent data types include the date and time of a visit, as well as the users IP address and hardware/software type. From conversations with those working in the industry we know that these are of potential interest to those working within digital subprime; here we can begin to render this interest visible.

In this short pilot experiment, however, we have been left with as many questions as answers: what accounts for the broad increase in trackers being used across the sector? Is this a general upward trend, or an anomaly? How do Ghostery's categories translate into what data is being *actually* connected. For instance, the results show trackers being involved in the collection of phone numbers (e.g., from mobile devices) – however, this may well relate to particular online telephony services, or to when phone numbers are volunteered by users.[19] Similarly, any demographic data being collected may not necessarily be able to be tied down to the level of the individual.[20] Moreover, there references to PII (Personally Identifiable Information) connect the categories to a particular term within US legal discourse – in Europe the preferred category is 'personal data' and does not necessarily refer to the same kinds of data (Borgesius, 2013). To a degree, then, these results are a prompt for further research.

That said, there are further ways in which this data can be set to work. By tracker tracking, we have been able to begin to develop profiles of both the work that individual trackers are undertaking, as well as the particular forms of tracking each website may be mobilising by *combining* individual trackers. For each is involved in a quite distinct form of 'socio-technical knitting', drawing on Jose Ossandon's term (Ossandon, 8 July 2013). That is, they are pulling together different online strands to each compose unique invisible 'tracking fabrics'. The question, however, is how to render this quite abstract, technical work visible and communicable.

For this, we can also turn to Ossandon. In some research on the ways that credit cards are passed between individuals and households in Chile, Ossandon asked his participants to map the whereabouts of their retail cards by pinning down woollen threads (Ossandon, 2012). Here some of the knitted socio-economic relations surrounding credit in Chile become visible and different profiles of movement comparable. In our case, the socio-economic knitting is undertaken not just by people but also by quite specific combinations of trackers. Trackers can each be seen as unseen, fully automated 'toolkits' that knit together unique assemblies of data points about internet users. We like using the concept of socio-technical knitting to describe the activity of credit trackers, because it renders the usually unseen and often apparently immaterial work of tracking more tactile, while allowing us to imagine how different patterns can emerge from their work. We have thus drawn on this metaphor in our attempts to visualise what is usually hidden, in order to construct distinct 'profiles' both of tracking toolkits and individual subprime websites. In so doing, we have worked with Frederica Bardelli and Carlo de Gaetano, designers at the Density Design Lab in Milan, who are experimenting with various ways to visualise digital relations.

Profiling trackers and digital subprime sites

Some of our initial collaborative outputs are presented below. Figure 5.3 shows a profile of particular tracking toolkits, listing the different kinds of data that they can collect. Each component in the toolkit stands for the collection of different data types. For example, in the visualisation of the Google+ widget, the buttons on the top row stand for ‘Ad Views’ (AV), ‘Analytics’ (A), and ‘Browser Information’ (BI). The second row contains buttons representing ‘Cookie Data’ (CO), ‘Date/Time’ (D/T) and ‘Demographic Data’ (DD). The measuring tape

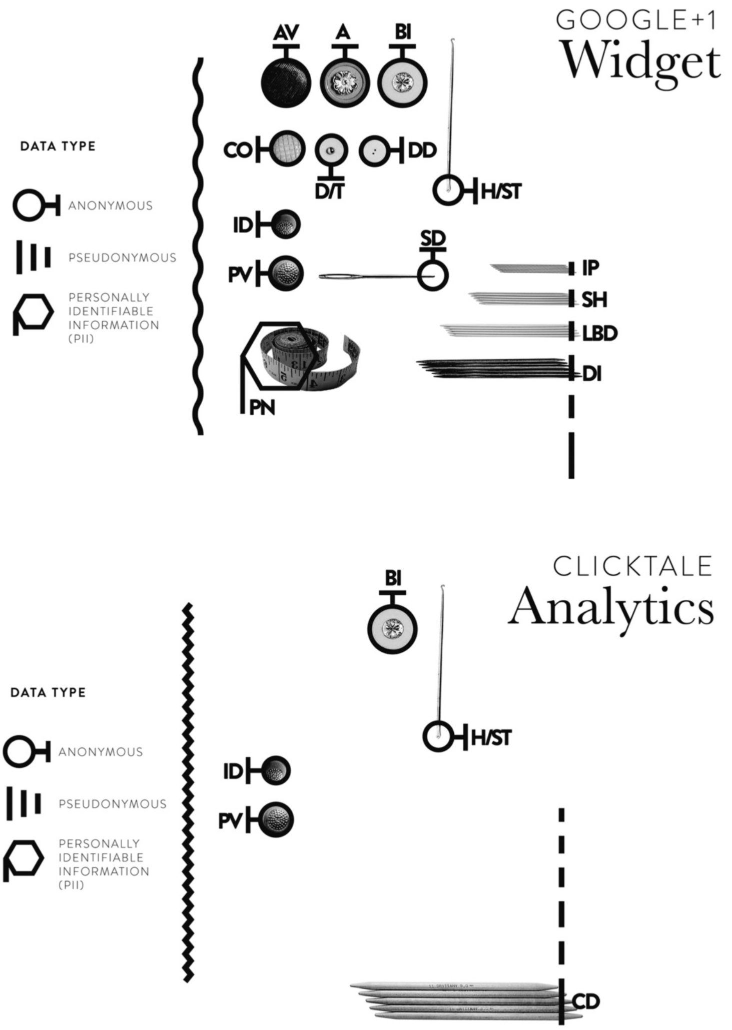

FIGURE 5.3 Comparing individual tracker profiles

stands for the collection of phone numbers and the black knitting needles towards the bottom-right stand for 'Device ID' (DI). These visualisations tell us that in order for companies to engage in online profiling, trackers need to stitch cookies into your browser, pin down your device ID, and obtain browser information and information about the date and time of a visit. Through these methods trackers can weave data together into more or less personal profiles. In order to give some indication of how personal this gets, the legend on the left side of the image shows three different symbols, each of which represents a sort of data: Anonymous, Pseudonymous and Personally Identifiable Data (PII) (these draw on Ghostery's own categorisations).

One of the results of these trackers-as-toolkits visualisations is that, when placed next to one another, they render more immediately visible the differing levels of sophistication possess by different trackers; here, for instance, how the Google+ tracker deploys a far greater variety of different ways to collect data than Clicktale.

This profiling of individual trackers has generic applicability. More specific to our research object is the profiling of how different digital subprime sites bring different trackers together. Three examples are shown (Figure 5.4). Each website is represented by a section of fabric, composed by all the trackers active in that website. The vertical length of this fabric indicates the number of trackers it contains. Each tracker, in turn, is identified with a number of threads proportional to the number of different types of data tracked by that tracker. This means that the 'density' of threads, the degree to which they are entangled, also allows for easy comparison between trackers. The design of this visualisation is quite deliberate: the curly/tortuous style of the threads is intended to give the images a sense of instability, signalling that the trackers and the collected data types may vary over time and over different browsing sessions. A dashed line indicates when no information is disclosed about what data a particular tracker collects. The icons to the left of the chart lines summarise the stated data retention policies of each tracker. For some trackers this is 18–24 months, for others it is a few years. More often than not, this remains undisclosed. Lastly, the icons on the right represent each tracker's 'data sharing' policy. It indicates what kind of data – for instance, aggregate data, anonymous data, and PII data – is shared with third parties.[21]

Comparing the 'tracker profile' of Kredito24.es (Spain, owned by Kreditech), Wonga (UK) and Spotloan's (USA, using technology licensed from ZestFinance (Hardy 2012)), we begin to be able to better detect important points of commonality and difference. Kredito24.es outweighs the other two in terms of the density of the data being tracked. All three are heavily reliant on trackers that do not disclose their data retention period.[22] They are also reliant on trackers that provide anonymous information about a particular user to third parties. For digital subprime lenders, what is important is the creation of specific profiles about their visitors in order to aid credit assessment. When combined with personalised data input by a potential borrower, this anonymous information can be tied to the

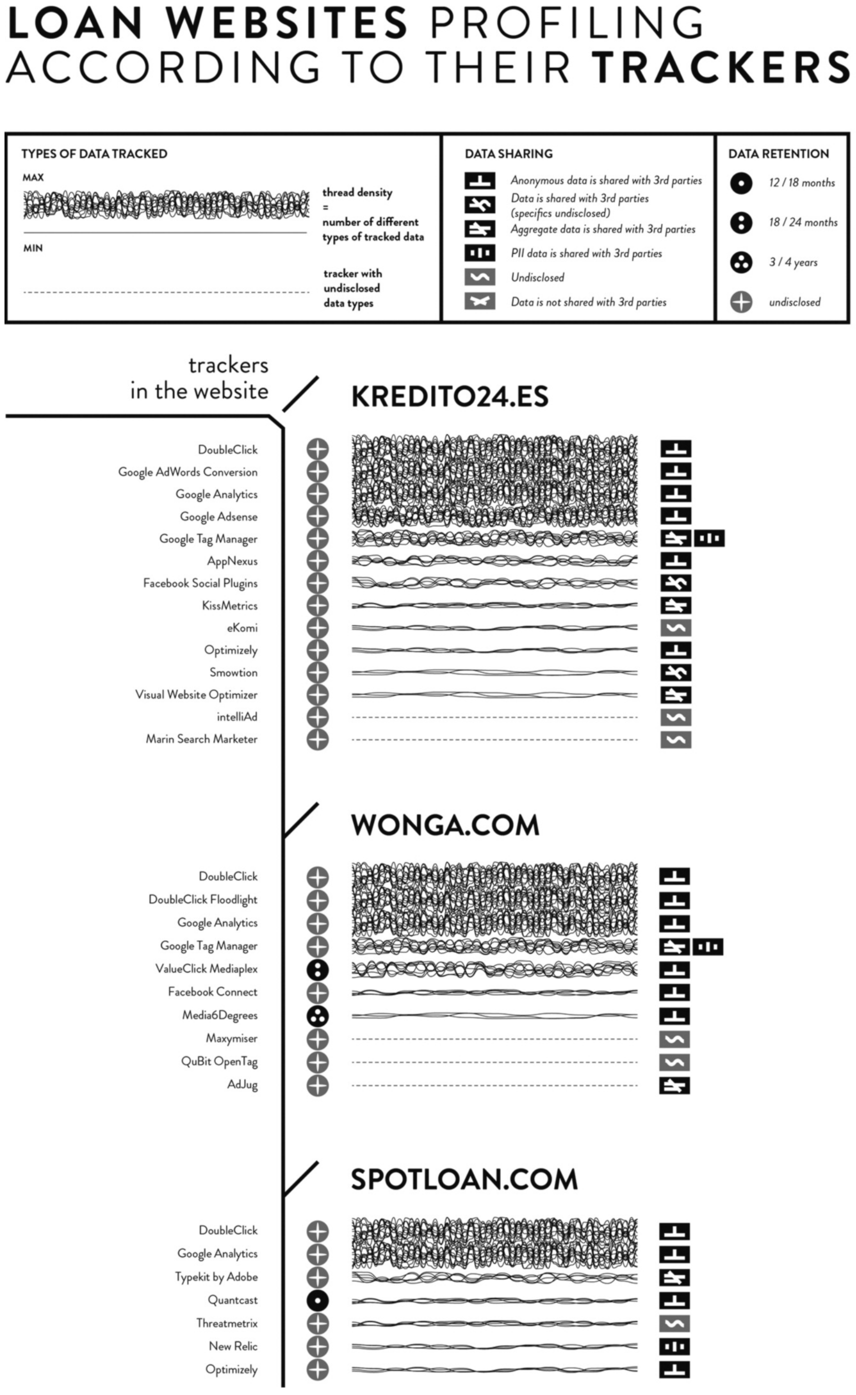

FIGURE 5.4 Comparing Kredito24.es, Wonga and Spotloan

individual. We can also see that Wonga is using a number of trackers whose data collection functions remain opaque – at least to Ghostery.

Further, the creation of individual profiles points us to what is unique in the tracking fabrics being composed by different digital subprime sites. In Wonga's case, a unique tracker is QuBit OpenTag. QuBit is a London based tech company, funded, perhaps coincidentally, by Balderton Capital, the same venture capitalist firm as Wonga. OpenTag itself is a tool partly designed to help companies improve their website's performance and monitoring. But QuBit also helps websites provide exactly the kind of real time personalised content, based on data such as browser type and IP address, that make Wonga's slider appear at different initial positions for different people. Thus, in a report designed to showcase the power of their analytics, QuBit describes how technology purchases by visitors using Safari are "around £30 more than any other browser", a conclusion designed to assist in practices of customer segmentation (Qubit, 2013: 14). Crucially this can be done virtually instantly, based on variables that many users might assume to be irrelevant.

Spotloan, meanwhile, is unique in using a tracker called 'ThreatMetrix'. In an industry sales briefing, ThreatMetrix is described as a provider of integrated cybercrime prevention solutions. The ThreatMetrix™ Cybercrime Defender Platform helps companies protect customer data and secure transactions against fraud, malware, data breaches, as well as man-in-the browser (MitB) and Trojan attacks. (Threatmetrix, 2013: 2).

The tracker is significant because it shows how the industry of online tracking, often associated with understanding and shaping consumer practices, is in this instance linking up with an industry concerned with combating cybercrime. Establishing the identity of a borrower has long been central to credit assessment practices. As these practices move online, new opportunities for potential fraudsters open up, thus generating new challenges for creditors. Such trackers are an indication of tailored attempts to manage the emergent risks involved.

Conclusion

Our attempts to understand the algorithmic basis of digital subprime calculation is ongoing. Since this initial pilot project we have expanded our dataset to incorporate twenty sites we suspect of using similar techniques and are now collecting data on a monthly basis. We also aim to further engage with industry figures. We are thus still in the process of chipping away at the opacities that characterise this industry. Digital methods are tools to do so, but they will need to be combined with others. In order to understand the rise and significance of so called 'big data' analytics, we as researchers will thus likely have to rely on a diverse palette of approaches, not just to keep our objects stable and 'detectable' (Law, 2009), but also to be able to understand and to become attuned to their transformations as they pass through diverse of socio-technical registers.

Our initial research has, however, provided both insights into the tracking work being done by digital subprime trackers, as well as into the challenges facing researchers seeking to understand online algorithmic calculation from the outside of an industry. In respect of the former, we can return to Ossandon (2013), who suggests that, while we *know* that credit practices produce networks, "what kind of collective or social formation are we talking about? At what level do these networks operate?" In the case of digital subprime, our initial findings suggest the creation of networks not just between potential borrowers and organisations involved in the credit industry (including both lenders and third party credit reference agencies), but also now involving the ever growing industry of online tracking. These sites do, then, seem to have an interest in using trackers to collect user data for the purposes of credit assessment and online behavioural profiling and segmentation. Consumer credit lending has long been accompanied by a range of controversies (see: Deville forthcoming). In the case of digital subprime, there is the potential for it to become wrapped up in the controversies surrounding the ethics of online tracking and the collection and retention of the data of users. Further, the deployment of 'custom' trackers and the common interest in particular data types also suggests an industry-specific 'professionalisation' of tracking practices. In other words, this is the highly emergent, likely experimental deployment of trackers that meet the specific needs of digital subprime websites.

Finally, we can reflect on what kinds of transparency such methods produce. Our findings are, to a degree, achievements of transparency, even if they remain incomplete. Striving to open up the opacities of digital subprime has also pointed us to the way in which digital methods itself is involved in the production of opacity. In our case, this has centred most clearly on our dependence on Ghostery's database and its process of categorisation. We have, however, departed from Ghostery's elementary understanding of trackers, and moved to a vocabulary of threads and density, which we consider more appealing to describe unseen 'work' of trackers (Star, 1991). Further, the role of rendering visual what is usually unseen is also centrally important to our work. As Tyler Reigeluth notes, digital traces tend to be naturalised and claims can too readily be made about their objectivity. He proposes to see such traces as 'in-formation' (Reigeluth, 2014: 253). For our emergent sociology of the invisible, the challenge has been, and continues to be, to grasp how trackers partake in forming digital traces and how they are also traces in formation themselves. One way we have begun to grapple with these issues is through visualisations that have emerged as the product of collaboration with designers. These reflections have helped us in turn to profile the different digital subprime websites, as different kinds and unstable textures. The challenge as we take this project forward is how to track and render visible these textures, as they continue to be re-shaped and knitted anew.

Acknowledgements

The authors would like to thank the editors and Paul Langley for their helpful comments on a previous draft of this chapter. Many thanks also to Frederica Bardelli and Carlo de Gaetano for their hard work on the visualisations and to Erik Borra and Emile den Tex for their assistance with the Tracker Tracker tool.

Notes

1 The way Star analyses 'social control' does not map cleanly into our case. Star is, amongst other things, concerned with the unseen, unrecognised and unpaid workers, such as (the many women) involved in unpaid (home) care. Their becoming invisible goes hand in hand with quite explicit forms of social control. Although the machines we study are indeed undertaking invisible work, their invisibility would connect up to very different forms of control – and would have less to do with their suppression as labouring subjects.

2 Although less relevant for present purposes, we are also interested and include under the 'digital subprime' umbrella another set of ventures that use alternative online methods to try to assess user behaviour (e.g., Lenddo and LendUp, both based in the US).

3 To be clear, these are individuals to whom the status 'subprime' is assigned by the credit industry, rather than a particular quality of personhood. Although the term is formally used within the industry to refer to borrowers who have fallen below a precise threshold in a risk-based analysis of creditworthiness, it is also used to refer more generally to those categories of borrowers that are perceived, irrespective of any formal evaluation, as undesirable from the point of view of mainstream lenders (see Langley, 2008: 473).

4 A site possibly using similar techniques is Sunny (www.sunny.co.uk). It is run by Think Finance, known for its use of big data analytics.

5 The 'mainstream'/'non-mainstream' distinction is a placeholder used for convenience. Payday lending is very much on a spectrum of credit products available to potential borrowers, and it would be incorrect to label it as in any way a separate domain. This is in particular given both the quantitative increase in such businesses in many countries, including the UK, and the fact that this industry relies extensively on ostensible 'mainstream' credit scoring technologies. It might better, therefore, be considered an example of what Rob Aitken (2006, 2010) calls 'fringe finance' (see also: Langley (2008: 170)), given that the metaphor evokes a continuity.

6 In the UK, an individual may have a number of such ratings, each generated by competing credit reference agencies. Wonga, for instance, draws data from both Callcredit and Experian (Wonga, 2014c).

7 At present it is unclear exactly what this additional third party data comprises (which is unsurprising, given that the information is proprietary). One source in a recent article in *The Guardian* newspaper speculates that Wonga draws on the wealth of free information that is available instantly online: electoral roll details, estimates of house values, for instance (Lewis, 2011). Wonga asks for users' vehicle registration details in the application process (if they own a vehicle). This might suggest they are tapping into the database of registered owners, perhaps to verify identity, perhaps to feed this into their risk calculations. Another source we spoke to speculated that they might also look into databases containing stolen mobile numbers (users are also asked to provide their mobile number), which could, again, be used to feed into their risk calculations.

8 Discussion with an anonymous industry source, 28 October 2013.

9 See, for instance, research into public views on targeted advertising (Pew Research Center, 2012), discussions about 'Do Not Track' (www.eff.org/issues/do-not-track),

which dates from 2007, but also the writings by public intellectuals, such as Evgeny Morozov, about consumer surveillance (and about 'big data' in relation to credit assessments) (Morozov 2013).

10 See for instance, Lightbeam (www.mozilla.org/en-US/lightbeam/) and Disconnect (https://disconnect.me/).

11 For example, through moving the nodes of your online 'data body' (Lightbeam), testing the effect of blocking trackers to your browsing experience such as connection speed (Disconnect), or by engaging with an analysis of rankings (Ghostery).

12 See: www.ghostery.com.

13 Ghostery. 'How It Works.' www.ghostery.com/how-it-works (accessed on 20 January 2014).

14 Evidon. 'Ghostery Sees What Scanners Alone Can't.' www.evidon.com/analytics (accessed on 8 March 2014).

15 See: http://knowyourelements.com (accessed 21 September 2014).

16 https://tools.digitalmethods.net/beta/trackerTracker/. The tool was created in a collaborative project by Yngvil Beyer, Erik Borra, Carolin Gerlitz, Anne Helmond, Koen Martens, Simeona Petkova, JC Plantin, Bernhard Rieder, Lonneke van der Velden, Esther Weltevrede at the Digital Methods Winter School 2012.

17 The Tracker Tracker found third party content on 72 per cent of the sites in the sample; using manual methods the figure was 73 per cent.

18 Project page: https://wiki.digitalmethods.net/Dmi/TrackersGuide.

19 See discussion here: https://twitter.com/Ghostery/status/433349897471799296.

20 For example, IP addresses could until recently only give an indication of geographical location and could not match the geodemographic precision of, say, a UK postcode (on which see Burrows and Gane (2006)), although this kind of research is progressing quickly (Lowenthal, 20 April 2011).

21 'Sharing PII data with third parties' does not necessarily mean that data is shared with any third party, such as the digital subprime website itself. It could also include another company that a tracker collaborates with, or an advertising network or broker.

22 Data retention period likely depends on the particular legislation in the country where the tracker company is based.

References

Aitken, R. (2006) "Capital at its fringes". *New Political Economy*, 11(4): 479–498.

Aitken, R. (2010) "Regul(ariz)ation of fringe credit: payday lending and the borders of global financial practice". *Competition & Change,* 14(2): 80–99.

Bollier, D. (2010) *The Promise and Peril of Big Data*. Washington, DC: The Aspen Institute.

Borgesius, F. Z. (2013) "Behavioral targeting: a European legal perspective". *IEEE Security & Privacy*, 11(1): 82–85.

Borgesius, F.Z. (2014) *Behavioral Targeting: How to Regulate?* (Diss.). Institute for Information Law (IViR), University of Amsterdam.

Borra, E.K. and B. Rieder. (2014) "Programmed method. Developing a toolset for capturing and analyzing Tweets". In K. Weller and A. Bruns (eds). *Aslib Journal of Information Management,* 66(3): 262–278.

Burrows, R., and N. Gane. (2006) "Geodemographics, software and class". *Sociology,* 40(5): 793–812.

Castellucia, C., S. Grumbach, and L. Olejnik. (2013) "Data harvesting 2.0: From the visible to the invisible web". *The Twelfth Workshop on the Economics of Information Security.* http://hal.inria.fr/hal-00832784 (last accessed 31 October 2014).

Chun, W. (2013) "Habitual New Media: Exposing Empowerment", Barnard College, New York City, 10 October 2013. http://vimeo.com/78287998 (last accessed 31 October 2014).

Deville, J. (2012) "Regenerating market attachments: consumer credit debt collection and the capture of affect". *Journal of Cultural Economy,* 5(4): 423–439.

Deville, J. (2015). *Lived Economies of Default: Consumer Credit, Debt Collection and the Capture of Affect*. London: Routledge.

Gerlitz, C. and A. Helmond. (2013) "The like economy: social buttons and the data-intensive Web". *New Media & Society*, 15(8): 1348–1365.

Hardy, Q. (2012) "Big Data for the Poor". *The New York Times: Bits Blog*. http://bits.blogs.nytimes.com/2012/07/05/big-data-for-the-poor/ (last accessed 31 October 2014).

Helmond, A. (2013) "The algorithmization of the hyperlink." *Journal of Computational Culture*, 3. http://computationalculture.net/article/the-algorithmization-of-the-hyperlink (last accessed 31 October 2014).

Hoofnagle, C.J., Soltani, A., Good, N., Wambach, D.J., and M. Ayenson. (2012) "Behavioral advertising: the offer you cannot refuse". SSRN Scholarly Paper ID 2137601. Rochester, NY: Social Science Research Network. http://papers.ssrn.com/abstract=2137601 (last accessed 31 October 2014).

Koot, M. (2012) "A survey of privacy and security decreasing third-party content on Dutch websites", October. www.madison-gurkha.com/press/2012-10-SurveyOfThirdPartyContent.pdf (last accessed 31 October 2014).

Kreditech. (2013a) "Where We Are | Kreditech.com." www.kreditech.com/ (last accessed 16 August 2015).

Kreditech. (2013b) "Kreditechnology". www.kreditech.com/ (last accessed 31 October 2014).

Kreditech. (2013c) "Online Loan and Online Lenders in Austria". www.kreditech.com/ (last accessed 31 October 2014).

Langley, P. (2008) "Sub-prime mortgage lending: a cultural economy." *Economy and Society,* 37(4): 469–494.

Langley, P. (2014) "Equipping entrepreneurs: consuming credit and credit scores". *Consumption Markets & Culture,* 17(5): 448–467.

Law, J. (2009) "The materials of STS", version 9 April 2009. www.heterogeneities.net/publications/Law2008MaterialsofSTS.pdf (last accessed 31 October 2014).

Lewis, T. (2011) "With Wonga, your prosperity could count on an algorithm", 25 October. *The Guardian.* www.theguardian.co.uk/money/2011/oct/16/wonga-algorithm-lending-debt-data (last accessed 31 October 2014).

Lowenthal, T. (2011) "IP address can now pin down your location to within a half mile", April. *Arstechnica.* http://arstechnica.com/tech-policy/2011/04/getting-warmer-an-ip-address-can-map-you-within-half-a-mile/ (last accessed 31 October 2014).

Marres, N. (2005) *No Issue, No Public: Democratic Deficits After the Displacement of Politics*. Amsterdam: Universiteit van Amsterdam.

Marres, N. (2012) "The redistribution of methods: on Intervention in digital social research, broadly conceived". *The Sociological Review,* 60(S1): 139–165.

Marres, N. and E. Weltevrede. (2013) "Scraping the social? Issues in live social research". *Journal of Cultural Economy*, 6(3): 313–335.

Marron, D. (2009) *Consumer Credit in the United States: A Sociological Perspective from the 19th Century to the Present.* New York: Palgrave.

Mendick, R. (2012) "Justin Welby, the Bishop of Durham, Calls for Cap on Total Cost of Payday Loans", 25 November, *Telegraph.* www.telegraph.co.uk/news/religion/

9700796/Justin-Welby-the-Bishop-of-Durham-calls-for-cap-on-total-cost-of-payday-loans.html (last accessed 16 August 2015).

Morozov, E. (2013) "Your social networking credit score", 30 January, *Slate*. www.slate.com/articles/technology/future_tense/2013/01/wonga_lenddo_lendup_big_data_and_social_networking_banking.html (last accessed 31 October 204).

Ossandon, J. (2012) "Sowing financial consumers: Retail credit in Chile". In Isaconf. https://isaconf.confex.com/isaconf/forum2012/webprogram/Paper21552.html (last accessed 31 October 2014).

Ossandon, J. (2013) "Low and high finance studies after performativity: A speculative workshop report, Charisma – Consumer Market Studies", 8 July. www.charisma-network.net/markets/low-and-high-finance-studies-after-performativity-a-speculative-workshop-report (last accessed 31 October 2014).

Qubit. (2013) "The Power of Personalisation: How Understanding Visitor Behaviour Helps You Create the Perfect Customer Experience". www.qubitproducts.com/sites/default/files/pdf/Power_of_Personalisation_Qubit_whitepaper.pdf (last accessed 31 October 2014).

Pew Research Center. (2012) "Search Engine Use 2012". Washington: Pew Research Center's Internet and American Life Project, 9 March. www.pewinternet.org/2012/03/09/search-engine-use-2012-2/ (last accessed 16 August 2015).

Raley, R. (2013) "Dataveillance and countervailence". In L. Gitelman (ed.). *'Raw Data' is an Oxymoron*. Cambridge, MA: MIT Press.

Reigeluth, T.B. (2014) "Why data is not enough: digital traces as control of self and self-control". *Surveillance & Society,* 12(2): 243–254.

Rogers, R. (2013) *Digital Methods.* Cambridge, MA: The MIT Press.

Rogers, R. (2009) "Post-demographic machines". In A. Dekker and A. Wolfsberger (eds). *Walled Garden*. Amsterdam: Virtueel Platform, pp. 29–39.

Rogers, R. and E. Sendijarevic. (2012) "Neutral or national point of view? A comparison of Srebrenica articles across Wikipedia's language versions." Conference paper, Wikipedia Academy 2012, Berlin, Germany, 29 June–1 July 2012.

Rogers, R., Weltevrede, E., Niederer, S. and E. Borra. (2013) "National Web Studies: the case of Iran online". In J. Hartley, A. Bruns and J. Burgess (eds). *A Companion to New Media Dynamics*. Oxford: Blackwell, pp. 142–166.

Seaver, N. (2012) "Algorithmic recommendations and synaptic functions". *Limn*, 2. Crowds and Clouds.

Shaw, W. (2011) "Cash machine: Could Wonga transform personal finance?" *Wired*, May 11. www.wired.co.uk/magazine/archive/2011/06/features/wonga/page/4 (last accessed 31 October 2014).

Star, S.L. (1991) "The Sociology of the Invisible: The Primacy of Work in the Writings of Anselm Strauss". In D.R. Maines (ed.). *Social Organization and Social Process: Essays in Honor of Anselm Strauss*. Transaction Publishers, pp. 265–283.

Threatmetrix. (2013) "ThreatMetrix for insurance: Remaining agile in a changing risk environment". http://threatmetrix.com/wp-content/uploads/2013/03/Industry-Brief-Insurance.pdf (last accessed 31 October 2014).

Tran, M., Dong, X., Liang, Z., and X. Jiang. (2012) "Tracking the trackers: Fast and scalable dynamic analysis of web content for privacy violations". *Applied Cryptography and Network Security*, *Lecture Notes in Computer Science,* 7341: 418–435.

van Dijck, J., and T. Poell. (2013) "Understanding social media logic". *Media and Communication*, 1(1): 2–14. www.librelloph.com/mediaandcommunication/article/view/MaC-1.1.2 (last accessed 31 October 2014).

van der Velden, L. (2014) "The third party diary: tracking the trackers on Dutch governmental websites". *NECSUS, European Journal of Media Studies*, 3(1): 195–217.

Weltevrede, E. (n.d.) 'Introduction: device-driven research' (working paper). In *Repurposing Digital Methods: Exploring the Research Affordances of Platforms and Engines*. Dissertation. Digital Methods Initiative, Department of Media Studies, University of Amsterdam.

Wonga. (2014a) "How Your Wonga Loan Works." www.wonga.com/how-it-works/ (last accessed 17 August 2015).

Wonga. (2014b) "About Us". www.wonga.com/money/about/ (last accessed 17 August 2015).

Wonga. (2014c) "Our Privacy Policy". www.wonga.com/money/privacy/ (last accessed 17 August 2015).