Original Research Article

BIG DATA & SOCIETY

Big Data & Society
July–December: 1–16

DOI: 10.1177/2053951720949571
journals.sagepub.com/home/bds
SAGE

For a situational analytics: An interpretative methodology for the study of situations in computational settings

Noortje Marres

Abstract

This article introduces an interpretative approach to the analysis of situations in computational settings called situational analytics. I outline the theoretical and methodological underpinnings of this approach, which is still under development, and show how it can be used to surface situations from large data sets derived from online platforms such as YouTube. Situational analytics extends to computationally-mediated settings a qualitative methodology developed by Adele Clarke, Situational Analysis (2005), which uses data mapping to detect heterogeneous entities in fieldwork data to determine 'what makes a difference' in a situation. Situational analytics scales up this methodology to analyse situations latent in computational data sets with semi-automated methods of textual and visual analysis. I discuss how this approach deviates from recent analyses of situations in computational social science, and argue that Clarke's framework renders tractable a fundamental methodological problem that arises in this area of research: while social researchers turn to computational settings in order to analyse social life, the social processes unfolding in these envirnoments are fundamentally affected by the computational architectures in which they occur. Situational analytics offers a way to address this problematic by making a heterogeneously composed situation – involving social, technical and media elements – the unit of computational analysis. To conclude, I show how situational analytics can be applied in a case study of YouTube videos featuring intelligent vehicles and discuss how situational analysis itself needs to be elaborated if we are to come to terms with computational transformations of the situational fabric of social life.

Keywords

Computational methodology, science and technology studies, situational analysis, platform studies, autonomous vehicles, computational social science

Introduction[1]

It has been argued that digital data and computational tools make it possible to extend formal, automated approaches to data analysis to contextual phenomena. This newfound capacity of formal analysis to take ephemeral, tacit and latent aspects of social life into account is said to derive from sophisticated computational methods such as machine learning, and their application to new sources of social data made available by digital architectures in society (Cointet and Parasie, 2018). To mention a popular example, digital listening services are said to be capable of situational analysis, as they are able to pick a suitable song for 'your ride home after work' based on locative analysis of aggregate collective user choices, thus taking listening contexts into account (Seaver, 2015). Computational scientists' claim to knowing contexts puts advocates of interpretative social enquiry in an uneasy position, as the latter tend to regard their ability to grasp contextual phenomena – latency, situatedness, atmosphere and so on – as an unique feature of their own approaches, and what validates their contribution to knowledge.

In this article, I offer a critical review of recent claims by computational social scientists to have rendered contextual phenomena amendable to formal analysis, but I go on to argue that we will need to learn how to combine elements from both scientific and interpretative approaches if we are to develop

Centre for Interdisciplinary Methodologies, University of Warwick, UK

Corresponding author:
Noortje Marres, Centre for Interdisciplinary Methodologies, University of Warwick, Library Road, Coventry, CV4 7AL, UK.
Email: N.Marres@warwick.ac.uk

adequate contextual understandings of social life in computational environments. That is, I seek to undertake a *critical reconstruction* of the methodological claims of both computational social science and qualitative social research to have a unique capacity for knowing context in a digital society.

To really understand the limitations – and possibilities – of the computational analysis of social life, we must be willing to consider a counter-intuitive diagnosis: computational social science's problem with context does *not* derive from a *shortage* of commitment to elucidating the situated, local and embodied character of social life in this field, but arises because its commitment to this purpose is *too rigid, and too narrowly defined*. The formal approach to data analysis required by automation leaves computational researchers ill-equipped to perceive, let alone come to terms with, a crucial transformation of our time: what counts as 'context' appears to be undergoing transformation in a digital society. New computational architectures, such as social media platforms, have rendered social life reportable, interpretable, shareable and influenceable in potentially new ways. As a result of the expansion of these architectures across society, social activities are becoming more strictly formatted, thinly structured, and artificial (Alaimo and Kallinikos, 2019). The very same digital transformations that have made available new types of social data and enabled the application of new computational methods in social research, are equally affecting the role of locatedness, embodiement, latency, atmosphere and so on in social life, in short, its contextual – or as I will go on to call it 'situational' – character.

The aim of this article is then twofold, to show that (a) frameworks currently under development in computational social science for the contextual understanding of social life do not sufficiently take into account the methodological implications of the computational transformation of social life, and (b) that interpretative approaches are equipped to address this lacuna, and this by creatively combining elements from qualitative, formal and automated approaches. To do this, I draw on recent work in the sociology of science and technology. Following Knorr Cetina (2014), I propose that it is the *situational framing* of social life that is fundamentally affected by its enactment in computational settings. Drawing on Clarke (2005), I argue that interpretative methods of data mapping make it possible to address the methodological problems noted above, as they offer an empiricist approach to computational enquiry, one that is focused on surfacing *from the data* which heterogenous entities – social, technical, media-based – make a difference in a situation. This makes it possible to treat as a researchable question the issue of whether and how situations in contemporary society are inflected by the computational settings in which they unfold. To conclude, I will illustrate the potential of this approach through a discussion of a digital social research project still under development, a semi-automated, lexicon-based analysis of so-called 'test drive videos' on the online video platform YouTube, which report on the introduction of intelligent vehicles into the social environment of the street.

Computational social science: Extending formal analysis to contextual phenomena?

It has become de rigeur in social science to posit that the development of new forms of computational data analysis enables new ways of knowing society (Lazer et al., 2009; Marres, 2017; Ruppert et al., 2013; Salganik, 2019). While in the 2000s, debates about computational social research focused on the affordances of the Internet as a research environment (Hine, 2000), and later in that decade, on industry-led developments in digital data analytics such as the rise of geodemographics (Savage and Burrows, 2007), in recent years scholarly attention has shifted to the capacities of advanced, 'intelligent' computational methods for social analysis (Castelle, 2020; Selbst et al., 2019). While still tethered to industry hype cycles, todays debate about digital ways of knowing society has produced a distinctive methodological claim, namely the idea that sophisticated new forms of computational analysis, such as machine learning, natural language processing and computer vision, have endowed computational science with *the capacity to render contextual phenomena amendable to formal analysis* (Ampofo et al., 2015; Bechmann and Bowker, 2019). Social phenomena that were previously considered to require interpretative research of some kind – such as ethnographic fieldwork or discourse analysis – can today, they suggest, be brought within the remit of formal, fully automatable data analysis.

Many of today's proponents of the new computational social science have backgrounds in the sciences, and the current generation seems less inclined than their predecessors to produce summary methodological statements, but it is not difficult to detect in publications in this area the above methodological proposition. In the area of textual analysis,[2] the French scientists Cointet and Parasie claim that new, informatics and AI-enabled approaches can today be used to elucidate sociological phenomena, as methods like Natural Language Processing can be integrated into approaches that take into account 'the context of the production of textual inscriptions,' thereby recovering their 'social thickness' (Cointet and Parasie, 2018: 3; see also Zubiaga et al., 2017). Törnberg and Törnberg

(2016) propose that the statistical textual analysis method of topic modelling can be used to study discourse, which they defined as 'communication in context,' because 'it explicitly models polysemy (cf. DiMaggio, Nag, and Blei, 2013), i.e. the notion that words can obtain multiple meanings depending on the context they are used in. In fact, what topic modelling does can be summarized as tracing the multiplicity of contexts of every word in the corpus' (Törnberg and Törnberg, 2016: 5). Nguyen (2017), who conducts linguistic analysis of social media data, makes a similar claim when she states that social media platforms offer '(a) rich contextual data, such as social network information; (b) the opportunity to study language use and human behavior in a multitude of social situations' (Nguyen, 2017: 3).

While these commitments to contextual understanding seem largely in line with those of interpretative social enquiry, further probing reveals that the understandings of context invoked in the contributions above are unconventional in a number of respects. The contexts that computational social scientists claim to be able to render amendable to formal, automated analysis differ from the contexts that interpretative social researchers have long claimed are only accessible through qualitative methodologies such as ethnography. Take the project of Nguygen (2016), which is to 'automatically infer social variables' – such as age and gender – 'from text' (p. 25) in social media analysis, which in her view requires taking the situational character of social media interaction into account. Citing Judith Butler, Nguygen explains that it is not just that young women have a different social media style than, say, older men. Rather, discernment of these different styles requires taking the communicative context into account (e.g. 'speaking with friends,' 'flirting') [.], and she concludes that 'using contextual information is the only way to improve predictive accuracy of gender classification based on "text only"' (Nguyen, 2017: 35). However, in contrast to Judith Butler's performative theory of gender, Nguyen's justification for taking context into account is to identify co-relations (between gender and speech), something which leaves out of account *the effects* of the use of gender-specific speech on the level of the situation. Nguyen states that her aim is *not* to achieve interpretative accuracy, but to maximise the predictive capacity of social data analysis by generalising from situational language use to demographic attributes.[3] In this version of contextual analysis, then, the research objective is to abstract actionable generalisable features of human behaviour from situations (Kelleher and Tierney, 2018).

This commitment to determine generalizable features of human behavior can be found in many computational analyses of contextual phenomena, and it leaves its mark in the very conceptualisation of 'context' in these studies. Take the study of stop-search encounters with the police by Voigt et al. (2017), which uses statistical methods of textual analysis to study interactions during street encounters with police. Analysing verbal interactions captured by body cameras worn by police officers, they sought to determine whether there is racial bias in the ways police officers address citizens. What stands out in this study from my perspective is the focus on a highly ritualised situation, like 'stop and search.' Ritualised situations, such as encounters with the police, and, in a different way, the flirting situations analysed by Nguygen, are likely to have stable features, which repeat themselves across different instances, and can therefore be more easily inferred using quantitative methods. This focus on situations with repeat-able protocols can be contrasted to the *interpretative* framing of situations in sociology and anthropology, *where situations are considered valuable analytic foci precisely insofar as they present moments of disruption*: occasions in which interactional scripts break down, are pushed to their limit, or require repair or adjustment (Woolgar and Neyland, 2008). It is this latter understanding of why context matters that seems at risk of being bracketed in the new computational social science in pursuit of generalisable accounts of social behaviour.

We have a situation: The methodological importance of 'break down' in interpretative social enquiry

In interpretative traditions developed in sociology from the early 20th century onwards, the analytic importance of situations was established on very different grounds than in recent computational social science. For the former, the analytic value of situations derives at least in part from the resistance of social activity to stabilisation and generalisation that becomes apparent here. This point was forcefully made by Erving Goffman, who is well-known for undertaking fieldwork studies of everyday situations, and who chastised what he called 'correlational' analysis of situations for merely documenting 'the geometric intersection of actors making talk and actors bearing particular social attributes,' noting that 'I do not think this approach is always valid. Your social situation is not your country cousin' (Goffman, 1964: 134). Goffman rejected the *generic* understandings of situations produced by correlational analyses, insofar as they did not acknowledge the *underdeterminacy* of situations, which, in his useful characterisation, are marked by the difficulty of formulating a simple, single answer to the question 'what is going on here?' (Goffman, 1964). For Goffman, an adequate interpretation of situations

cannot be produced from a distance, by relying on abstract or typical understandings. This is because the answer to the question of 'what is going on here?' is at least partly an accomplishment of interaction within the situation itself, and can therefore only be achieved, and documented, by observing the situation 'from the inside.'

The idea that social life cannot be adequately understood through formal analysis was also taken up by ethnomethodologists, like Harold Garfinkel. Crucially, the latter extended this idea to *mediated* situations unfolding beyond the face-to-face. Anne Rawls, in her 2008 introduction of Harold Garfinkel's theory of information, criticises efforts to model situations for obscuring the *constitutive contingency* of social life, an open-endedness or uncertainty, if you will, that can only be made manageable as part of the unfolding of social life across settings. As she put it, formal analysis is ill-adapted to the analytic objective of surfacing *the constitutive process of the production of shareable interpretative frameworks* in social life. As she writes:

> Information is situated. [.] [We must] study those constitutive orders that naturally develop to manage and order contingencies. Abstract models do not help. What they do is obscure the contingencies that should be the focus. (Rawls, 2008: 35)

For Rawls, these contingent processes of mutual coordination that happen in situations also extend to communication across settings. For this reason, a science that disregards processes of contingent attunement between actors do not just limit our ability to understand what goes on in face-to-face interactions, but in social life as such.

In the interpretative sociologies inspired by Goffman and Garfinkel, situations are then valued as sites where mutual coordination happens among actors as an inevitably situated process: understanding of the situation can only emerge from the situation. Other interpretative sociological traditions, such as actor-network theory and pragmatist sociology, went further, in that they value situations as occasions where shared understandings and assumptions are called into question. For actor-network theory (Latour, 1993) and the sociology of critical capacity (Boltanksi and Thévenot, 1999), a situation is first and foremost marked by the possibility of *dispute* about 'what is going on here': situations arise *when it is no longer possible to carry on in the habitual way, by relying on conventional, engrained and repeated ways of doing* (Boltanksi and Thévenot, 1999). A similar approach is taken in American pragmatist sociology, where situations came to be defined as 'problematic, high-stake episodes that cast our prescribed roles and trajectories into question' (Missche and White, 1998: 697). This is not the place to provide a more detailed intellectual history of pragmatist sociology, but the main point is that situations here are relevant foci of analysis precisely *because* they are *not* routine, and do not repeat themselves exactly. It is not only because we cannot assume an agreement about 'what is going on' – among the actors involved, nor among analysts – that sociologists should pay careful attention to situations. Situations present moments in which *rule-following breaks down*. It is this insight – that social life cannot be conclusively defined in terms of 'rule-following' – that led pragmatist sociologists to posit that situations resist purely formal, rule-based forms of analysis.

To sum up, the claim that context has been rendered amendable to formal scientific analysis in computational data science relies on a particular understanding of the situatedness of social life, in terms of formalisable, generalisable features of situations. Such a formal understanding of contextual phenomena differs from those advanced in interpretative sociology, where situations are defined in terms of underdeterminacy, contingency and problematicness. From the latter perspective, formal analysis inevitably leaves key aspects of situations out of account: it fails to engage not only with the uncertainty of situations but with their *unresolved* character. In a given situation, which interpretation of the situation will prevail and prove adequate is *not just unknown but fundamentally in question*, the peculiar challenge of situations being that *the definition of the situation is at stake in the situatio*n; it is partly decided by how the situation unfolds. For authors like Goffman and Rawls, the analyst can only appreciate this formative feature of situations by adopting a position *inside the situation*. In other words, the type of contexts that computational social science claims to have rendered amendable to formal analysis with the aid of new data and new methods *is a different type of context* from the one interpretative social researchers have long claimed is only accessible through situated methodologies. I now turn to the question: if we take the interpretative definition of situations seriously, how then could computational methodologies inform our understanding of them?

The situation is dead, long live the situation: The transformation of aboutness in a digital society

Faced with confident claims by computational social scientists that they are able to analyse social phenomena previously deemed inaccessible to formal analysis, like context, it is tempting for interpretative social researchers to fall back on classic critiques of scientific

methodology, such as those developed by Goffman. But it is crucial that we critically review taken-for-granted assumptions *on both sides* of the debate between formalist and interpretative approaches. There are two reasons for this. First, *the rise of computational forms of social analysis are considered exciting because they open up alternative directions for methodology development: they may enable changes in the relation between interpretation and formalisation in social research*. Secondly, today's problems with the formal analysis of situations are potentially different from the problems identified by sociologists in the second half of the 20th century. For instance, Goffman and Rawls' criticisms discussed above focus on *interpretative accuracy*: formal analysis is not capable of producing adequate accounts of social life. However, today we are facing a somewhat different issue, that of the *analytic capacity* and actionability of computationally enabled social science. It is not just claimed that formal analysis has descriptive or explanatory power as regards context but a number of different advantages are attributed to it, such as being scalable, predictive and embeddable in computational infrastructures in society.

But formalist methods also face methodological problems in today's digital societies which challenge its analytic capacities. Knorr Cetina (2009, 2014) has made an important contribution to surfacing one of these problematics. She argues that in a digital society the very composition and nature of 'situations' is undergoing transformation. Starting from the often-made observation that the importance of face-to-face situations has diminished in technological, media-intensive societies (see also Smith, 1990; Marvick and Boyd, 2010), Knorr Cetina argues that this shift affects not just community bonds or sociality, but the *situational fabric* of social life: face-to-face situations – which 'are foundational for how we conceive of the emergence of sociality and effects like trust' (Knorr Cetina, 2014: 47) – are being gradually replaced by 'synthetic situations.' The latter do not require 'being there in person but allow for participants and objects to be dispersed and still process things interactionally and collectively' (Knorr Cetina, 2014: 47). And 'a synthetic situation is a composite, an assembly of information bits that may arise from many areas around the world and feature the most diverse and fragmented content' (Knorr Cetina, 2014: 49). To develop this theory, Knorr Cetina draws on field research on electronic trading, and also discusses marital conflict via Skype, but her aim is to offer a general diagnosis of digital societies, as marked by interactional conditions *that put the status of situations itself at risk*. In digital societies, 'conditions that were once central and held to be universal may change' (Knorr Cetina, 2014: 46).

Knorr Cetina's argument brings into view a fundamental, empirical transformation of situations in computationally-intensive societies, which has significant methodological implications for how we define the object of social analysis in these societies, but is likely to remain out of view as long as social analysis is focused on *routine* situations. The methodologically ordained pre-occupation in computational social science with repeated, regular, conventional, generalisable situations risks to leave out of consideration key defining features of situations in a digital society, such as the precarity and increased difficulty of accomplishing a shared interpretation of what is going on here. As Knorr Cetina (2014) puts it, 'situational integrity' is much harder to maintain in the mediatised setting of the synthetic situation than in the face-to-face. In mediated situations, '*the result is much more likely a muddle* [italics mine]: a disorderly interactional arrangement struggling with problems of differential access, orientation and perspective, and coordination' (Knorr Cetina, 2014: 47[4]). One of the possible results of digital transformations of society is that a key feature of situations, namely account-ability[5] between actors, comes under pressure. Where the situation presents a muddle, actors may be more inclined to opt for more generic and conventional forms of communication. Situations in mediated settings would then be *both* more disordered, and provoke more generic forms of communication – *actively contributing* to the demise of situations as defined by Missche and White (1998), in terms of problematicness, as 'cast[ing] prescribed roles and trajectories into question,' and offering occasions for actors to *account* for roles, trajectories and relations.

Knorr Cetina also points to an alternative methodological understanding of *what composes the situation*. The notion of the synthetic situation highlights the constitutive role of computational settings, like electronic trading platforms, and digital media architectures, such as Skype, in the organisation of situations. From this perspective, computational media architectures and devices do not just present a condition of possibility for sociality, and its analysis, they participate in the very articulation of the situation qua situation. It is this that is only rarely acknowledged in formal situational analyses, even if computational social scientists are not unaware of the methodological challenges posed by digital data architectures in this respect. In a recent introduction to the field, Salganik notes that 'the digital systems that record behavior are highly engineered to induce specific behaviours' (Salganik, 2019: 35). However, he still defines the computational approach to knowing society as an *observational science* (see also Lazer et al., 2009). This obliges him to assume, *on methodological grounds*, that societal architectures for data capture and analysis do not fundamentally inflect or

inform the phenomena under study, or at the very least, that such effects are containable: digital architectures may 'distort' social phenomena that unfold within them but cannot be assumed to positively inform their organisation. Similarly, when Voigt et al. (2017: 21) discuss the possible bias introduced by the presence of observers in stop-and-search situations – which may or may not include the body camera's worn by police offers (!) – they go on to show how this potential source of bias does *not* significantly affect their study.

The idea that the apparatus of knowledge should not contaminate the phenomenon under study, and to negatively define phenomena thus affected as 'experimental artefacts' (Rheinberger, 1997), is one generally held and respected across the sciences. However, in pursuing this methodological tenet uncritically in the computational analysis of social life, the effect is to significantly limit the analytic capacity of computational social science to engage empirically and conceptually with digital transformations of social life. The framing of situations in terms of generalisable scripted behaviours in computational social science may then, in the long term, limit our capacity to understand society through computational methods. An analytic focus on stable, circumscribed situations ('flirting'; 'stop and search') implicitly or explicitly defines social life in terms of stable rituals and interactional forms, and this may put computational social science at risk of excluding from empirical analysis phenomena that look like mere contextual noise or artefacts of machinic bias: the muddles we face when finding Twitter messages littered with too many hashtags, a comment space full of advertising and spam. However, if we follow Knorr Cetina's analysis of synthetic situations, such muddles *may precisely be constitutive of the situations in which actors find themselves in computational societies*. The 'aboutness' (Gross, 2016) of interaction, information and communication – their capacity to be 'about' something, to find a referent in social and cultural life, the determination that some*thing* is definitely going on here – is *not* as a matter of course accomplished in mediated settings (Lindgren, 2020). A conventionalist definition of 'situations,' in terms of successfully ritualised interaction, is likely to leave us – analysts, as well as actors – underequipped to understand what is going on in digital societies.

Consideration of what counts as 'a situation' in a digital society then brings into view the following methodological challenge: *If in computationally-inflected settings in society, technological infrastructures, media architectures and devices may play an active role in organising – or dis-organising – situations, how then should we analyse situations with the aid of computational methods?* In considering this methodological question, it becomes clear that Knorr Cetina's account has some limitations. For one, her definition of the 'mediatised setting' is mostly limited to the digital front-end, which she says is composed of screen-based technologies. As such, her concept of the synthetic situation more or less disregards the infrastructural layer of computational architectures including that of communication and interaction formats (publishing, sharing, friending) and algorithmic selection (rankings and ratings). However, the latter seem precisely key to possible transformations of situations in today's computational societies: it is precisely because of the relative invisibility of computational data architectures across society that the socio-technical framing – or possibly, de-framing – of social life is at risk of being ignored (Marres, 2017; Maguire and Winthereik, 2019). If we are to develop an understanding of how the situational fabric of social life is undergoing transformation today, we should therefore extend our analysis of situations to include this infrastructural layer.

Situational analysis: An empiricist approach in interpretative computational enquiry

In the remainder of this paper, I would like to outline one possible way in which interpretative traditions in social research can contribute to addressing the above methodological challenge, namely, by making the situation a unit of *empiricist analysis* in computational enquiry. In proposing this, I follow a particular qualitative approach to data analysis, namely Adele Clarke (2005) situational analysis (SA), a methodology for the study of situations with methods of data mapping. The aim of data analysis for Clarke is 'to specify which entities – of varying scale and composition – make a difference in a situation' (Clarke, 2005: 78). Such an empiricist approach does not presume to know beforehand which entities are relevant to the situation, how they relate, what their status is (human or non-human, social, technical or natural or media-based or conceptual), or even 'what is going on here.' Instead, the objective of SA is to determine *which entities are activated and deployed in the specification of the situation at hand*.[6] It seems to me that Clarke's approach has much to offer for a computationally-enabled, interpretative analysis of situations as they unfold in computationally-inflected settings, for the following reasons.

First, Clarke's SA explicitly recognises the participation of technical, mediated and environmental entities in situations: to specify the heterogeneous composition of a situation, SA proceeds by constructing so-called compositional maps, discursive data visualisations populated by diverse elements including non-humans, technical entities, ideas, issues, organisations and so on. This

interpretative cartographic method also allows SA to recognise the dynamic nature of situations: 'situational analysis favours analytics over theory, because the composition of the situation is always changing' (Clarke, 2005: 28) This makes it possible for situational analysts to recognise the constructive and/or destructive contributions of *different types* of entities – technical, social and environmental – to the unfolding of situations in computational settings, from a camera on the chest of a police man in a stop and search situation, to a like button on a Facebook page. Second, SA's aim is to surface latent, problematic realities: it does not 'wait for emergence from data (...) as we must "actively detect silences in data"' (Clarke, 2005: 75). SA, that is, specifies entities that compose the situation not in a purely descriptive mode, but defines this task as *articulation work*, actively attending to what may be difficult to express. Third, Clarke's approach is able to recognise the capacity of situations to surface account-ability requirements on the actors implicated in them. With its commitment to specify 'what makes a difference in situations,' SA makes it possible to operationalise situations as empirical occasions for accountability. Fourth, SA offers an *iterative approach* to data analysis: the construction of compositional maps and the specification of the situation is a qualifying operation, with involves the iterative curation of data and map, of figuring out the situation *and* determining what are its consequential elements, through a back-and-forth between empirical materials, data, concepts and visualisation.

As I will illustrate by discussing a pilot study below, Clarke's approach offers a possible way of analysing situations as they unfold in computational settings with digital methods of textual and visual analysis. However, doing so also brings to the fore an important limitation of the approach. Even if SA does not define what composes the situation at the outset, it still seems to presume *a bounded and recurrent situation*. Clarke's SA presumes a world in which situations are detectable as part of the process of data collection and analysis, without broaching the question of how socio-technical infrastructures problematise this very possibility. To be sure, Clarke recognises the importance of infrastructure, noting that 'taken-for-granted, invisible non-human actors like "electricity" are generally assumed [to be in place]' and that 'specifying such non-human actors is generally important' (Clarke, 2005: 87). But the approach rests firmly on the assumption that there are fields of social activity, which can be transformed into ethnographic material. The delineation of situations is itself not problematised on infrastructural grounds. However, situations as they unfold in computational settings often do not unproblematically belong or contribute to clearly defined fields of activity, and may present not-quite situations or semi-situations (what Knorr Cetina calls 'muddles').[7] In settings like these, the relative (un-)boundedness of situations or not-quite situations, is likely to be at least partly an effect of the digital media infrastructures in which they unfold.

Insofar as the object of SA is not only informed, but also problematised, by the computational settings in which they arise, it seems helpful to recognise the situations here unfold not in a field, but in a semi-field, to use the term proposed by anthropologist by Kelly (2012). Coined to characterise experimental huts, a kind of model home designed for the study of malaria in model villages in East Africa, Kelly defines the semi-field as 'a stage upon which to observe [.] phenomena, bridg[ing] the distinct empirical terrains and methodological registers of the laboratory and the field.' As Kelly points out, the semi-field 'is located in the field, but it is not quite of the field': these artificial environments are explicitly designed with the purpose to render monitor-able and analys-able what happens in them. Just as experimental huts, computational environments like social media platforms are sufficiently 'like' other environments in society, insofar as they enable social interaction, expression and organisation, yet 'they are controlled enough to facilitate intervention and manipulation of these activities, provid[ing] the artificial conditions required for the recording and analysis of these actions (Derksen and Beaulieu, 2011)' (Marres, 2017: 53). It is in comparison to this relative artificiality of digital social life as studied in platform-based social research, that it becomes clear how, by comparison, Clarke's approach is marked by what could be called *a residual naturalism*. The idea that the infrastructural environments in which social life happens can often be bracketed in the study of social life does not just mark the quantitative methodologies implemented in computational social science: traces of this assumptions can equally be detected in interpretative traditions in social research like Clarke's. This is understandable insofar as in the situations for which SA was developed infrastructural conditions often do not constitute the problem at hand (say, the availability of electricity not being the issue on a hospital ward under study). But if we are to develop an interpretative analysis of situations in computational settings, we will need to be able to recognise the participation of infrastructures, media and devices in the situation.

Situating intelligent vehicle testing in society: A semi-automated analysis of test drives on YouTube

To illustrate how SA may be operationalised – and developed – in computational social research, I will

briefly discuss a pilot study undertaken with colleagues in the Media of Cooperation Research Programme at the University of Siegen in 2017 and 2018. In this project, we turned to the online video platform YouTube to examine whether and how videos featuring self-driving cars undertake situated evaluations of this new technology in environments in society, notably the street. In examining this, we built on recent work in Science & Technology Studies and Human Computer Interaction, which has proposed that the appearance of 'intelligent' or smart vehicles in street environments presents an opportunity for social learning about technology 'in the wild' (Brown and Laurier, 2017; Stilgoe, 2018). Thus, Brown and Laurier (2017) have analysed YouTube videos featuring Tesla cars in Autopilot mode, showing how these videos situate, contextualise and problematise intelligent technology by reporting on 'real-life' experiences of driving and encountering these new technologies in the street. Building on this work, our project asked, can test drive video's featuring self-driving cars on YouTube be said to instantiate a *situational mode of evaluation* of the introduction of intelligent vehicle technology in environments in society? Do they render this event – the introduction of intelligent technology into the social environment – available for interpretation from the standpoint of the on-going happening of life in society, on the road?

Our question was informed by the following concern: while user-led evaluations of technology in the form of online video reviews have quickly gained currency in today's cultural economy, it remains in question whether and how this form of technology testing in everyday settings is capable of producing *evaluations* of new technology, and of ensuring the accountability of innovation. We wanted to establish whether and how user-generated videos featuring self-driving cars rendered available for public evaluation the type of situations that according to the sociology of technology enable social learning about innovation to take place: the testing situation, the moment in which the introduction of a new entity into social life disrupts habitual ways of doing, and in that moment compels social actors to engage in articulation work, specifying and evaluating features of technology, social life and their interrelations (Boltanksi and Thévenot, 1999; Star, 1999). We asked: do YouTube videos of intelligent vehicles in street environments surface such testing situations, helping to render explicit the implications of self-driving vehicles for society? That is also to say, in taking up methods of YouTube video analysis, we sought to determine whether video reports of intelligent vehicles in the streets surfaced situations, and whether they enabled the type of accountability relations that according to sociologists like Goffman and Knorr Cetina are facilitated by situations.

In focusing on the popular YouTube genre of 'tech review,' our study explicitly took digital media architectures into account. We approached situations, or rather, the situational, not as given in the data, but as entailing *a distinctive mode of publicity*, a mode of reporting that deploys contingent and contextual occurrences and encounters in everyday environments like the street in order to narrate and/or investigate the introduction of new technology into society.[8] However, this is also to say that we framed the relevance of media architectures in explicit reference to the situation that formed the object of our analysis: the introduction of self-driving vehicles into social environments. Because of this, we defined the significance of YouTube as a popular platform for technology review not primarily in terms of user-generated content (Arthurs et al., 2018), but in terms *of facilitating technology review from situated standpoints in mundane social environments*. This approach differs from other sociological research that relies on platform-derived video data to analyse situations, such as the work by Nassauer and Legewie (2018) who define online video data analysis as 'focuse[d] on situational dynamics and behaviors using video or other visual data to understand how people act and interact, and which consequences situational dynamics have for social outcomes' (Nassauer and Legewie, 2018: 2). That is, Nassauer and Legewie (2018) define situations in terms of what is depicted in video data, whereas our pilot study sought to establish whether and how YouTube, as a digital media architecture, enables situational modes of reporting on the introduction of new technology in society. We are interested in the extent to which online publicity platforms like YouTube are *configured* to enable the development of new, situational forms of evaluating technology, in this case, the reporting test drives and sightings of self-driving cars in the social environment.

To investigate this, we combined two different approaches to online video analysis, each of them adapting Clarke's SA in a different way: (1) an interpretative mapping of situational elements in a small corpus of online videos and (2) a semi-automated textual analysis of a larger corpus of YouTube video descriptions collected via this platform's API. Thus, in the first step, we conducted data sessions loosely structured on the in-depth interpretation of video recordings that are the specialty of ethnomethodology, which on this occasion we referred to colloquially as 'deep watching,' to mark its contrast from the larger-scale textual analysis reserved for the second phase of our study. Working with an interdisciplinary group of scholars with backgrounds in digital media studies, Science and Technology Studies and sociology, we selected 15 online videos featuring driverless cars which potentially matched our description above:

reporting on the appearance of driverless cars in the social environment in the situational mode.[9] We then watched and interpreted test drive videos featuring self-driving cars[10] en groupe over the course of a few days, with two aims: (a) to group the videos in our corpus into categories, and (b) to produce for each video type an initial mapping of constituent elements. After watching each video, we categorised the videos as follows: (a) company demos (featuring on-the-road vehicle demonstrations by automotive and tech companies); (b) DIY testing (amateur videos of test drives, featuring mostly Tesla vehicles in autopilot mode, recorded with dashboard cameras or smart phones and narrated from a driver's perspective); (c) the 'view from the street,' which consisted of recordings of third-party sightings of self-driving test vehicles (Google, Uber) in the street, by journalists and other external observers. Each of the video categories was marked by different cinematographic styles, with company videos tending to be professionally produced, while content in the categories 'DIY testing' and the 'view of the street' tended to follow home video and real-time reporting conventions.

Next, we identified notable elements in the videos in each of the three categories, loosely following Clarke's first step of SA, situational mapping (Clarke and Charmaz, 2019: 15)[11] using sticky notes. Watching each video in silence, participants were asked to identify *heterogeneous entities featuring in the videos* which, in their view, could help to answer the question: which objects, actors, concepts and values are invoked in the videos to specify what is at stake in the introduction of autonomous vehicles into the social environment? Our provisional findings suggest that the demo videos produced by automotive companies made the most significant effort to narrate the social environments in which self-driving cars operated, featuring women (non-)drivers, diverse road users including cyclists, urban and road environments that were clearly named, and vehicle engineers discussing the unpredictability of the street environment. However, they did so not necessarily in a situational mode, since, as one participant put it, everything in the videos is so clearly scripted. DIY testing videos did remarkably little to report the environments through which the test drives were passing, focusing instead on vehicle performance and the driving experience (as one participant provocatively summed it up, 'talking to self, in the fog'). By contrast, our situational map for 'views from the street' recorded a number of situational elements ('clumps of people on sidewalk,' 'a police car observing transgression,' 'test ground fenced off' and the enigmatic 'ruins of the automotive society,' referring to a graffiti covered underpass where one of the vehicle sightings occurred), but also includes viewer interpretations indicating that the curation of a testing situation was not quite accomplished in these videos ('not much happens,' 'car not strange enough,' 'people don't notice the vehicle'). For purposes of illustration, Figure 1 presents a transcription of this last mapping.

Seeking to extend our analysis beyond the speculative interpretation of our small, theory-driven selection, we then took the next step of conducting a semi-automated textual analysis of a larger corpus of YouTube video descriptions collected via this platform's API. To this end, we constructed a larger corpus of YouTube videos of street tests of intelligent vehicles in the following manner: from a custom-made Twitter data set, consisting of tweets containing the terms driverless car, self-driving car, autonomous vehicle, and related terms between 15 October 2017 and 15 April 2018, we extracted all YouTube URLs (total 4052 videos).[13] We then queried the YouTube API to extract the video descriptions produced by the creators of these YouTube videos. Based on a selection of the Top 500 most frequently tweeted videos, and informed by the deep watching exercise, our study group proceeded to construct a lexicon of relevant terms for the analysis of these self-descriptions, identifying terms that could help to specify 'what is going on here' (the situation) and could serve as indicators of the extent to which the video situated intelligent vehicles in environments in society. In doing so, we followed Gerlitz and Van der Vlist lexicon-based analysis of app videos on YouTube (see for a discussion Dieter et al., 2019): in populating our lexicon with terms (see Figure 2), we then constructed a model of the situation composed of heterogeneous entities extracted from our data, through a back and forth between our interpretative maps, our top 500 URLs, and constrained by the lexicon tool's technical limitations (for example, at this stage, we could not identify phrases). The resulting lexicon, consisted of two types of categories: (1) genres (news, demo, recording, humour, test) and (2) features (environment, business, technology, accidents). For each category we identified index terms, the aggregated occurrence of which in the video descriptions of our corpus would indicate the category obtained for the video at hand.

Using an R script, we then applied the lexicon to the full corpus of video descriptions (4052 videos),[14] allowing us to establish (a) the proportion of videos that featured in each of the categories of video genres in the lexicon – Demo, News, Promo, Test, Recording, – and of different categories of entities (features) and b) the co-occurrence of the categories of video genre and features in the videos analysed. We then hypothesised that these different entity types could be taken as indicators of a situational mode: a strong presence of entities in the category road

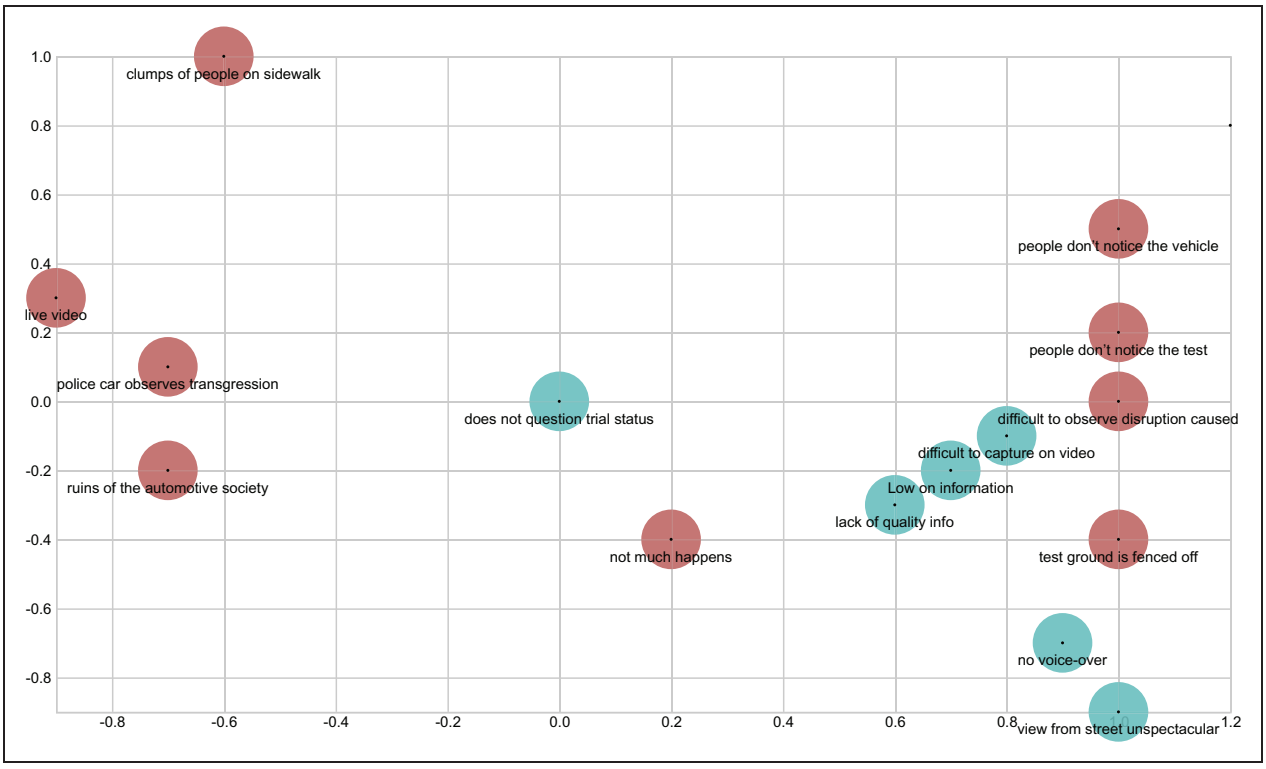

Figure 1. The 'view from the street' situational map of online videos recording third-party sightings of intelligent vehicles, December 2017.[12]

environment (zebra crossing, traffic light, side walk) would indicate a comparatively speaking more situationally grounded mode of reporting, then a large presence of entities in, for example, the category business (brand, invest, market). Through a correlational analysis (see Figure 3), we were then able to produce an overview of different types of entities featuring in the various genres of videos: media-specific terms (brand, promote) as well as actor types (pedestrian, cyclists, police), and environmental entities (traffic light; zebra crossing).[15] Noting that videos in the category 'test' correlate with a wider range of environmentally specific features, as compared to videos in the category 'demo,' we tentatively attribute to the former genre of videos a greater capacity to locate intelligent technology in environments in society, and the potential to evaluate technology in the situational mode.

To be sure, this analysis leaves many questions unanswered, including that of whether and how test drive videos on YouTube enable the enactment of accountability relations, both within the dramaturgy of each video, and as media circulating in YouTube infrastructures and beyond. Neither does our lexicon analysis enable us to specify in sufficient detail how the media architecture of the YouTube platform leaves its mark on the 'testing situations' in the videos under scrutiny, although it was clear to us that they do. In parsing YouTube video descriptions for our lexicon building exercise we encountered lots of material that pointed in this direction, from dedicated Tesla test drive channels set up to enable monetisation of YouTube content, to attention seeking content like a demo of how to put make-up on while driving in Autopilot mode. What we called above the 'infrastructural layer' of online platforms equally left its mark on our analysis. For one, in turning to a Twitter data set to extract the larger set of YouTube URLs featuring driverless cars, we gave the latter social media platform a role in the delineation of the 'testing situation' under scrutiny, the appearance of self-driving cars in street environments. This begs the questions: appearance in which street, where? At which level is 'the situation' constituted, in our semi-automated online data analysis, that of the individual video or of the entire data set? While our lexicon analysis suggests that situational mapping as an interpretative method can be scaled up with the aid of automat-able, lexicon-based methods of data analysis, these methods at the same time introduce platform effects into our very delineation of the 'situation' to be interpreted. However, to understand the participation of infrastructures in the situation under scrutiny, we would need to extend

Type	Category Grou	Category Grou	Category	Category Grou	Count	Queries						
Formats	1		**Test**	1_Test	17	prototype	sighting	realtime	log	spin	happens	try
Formats	2		**Talk**	2_Talk	12	interview	lecture	keynote	interview	presentation	comment	talk
Formats	3		**Demo**	3_Demo	6	hack	demo	how	works	experience	showcase	
Formats	4		**News**	4_News	11	report	news	media	story	stories	CBC	CNN
Formats	5		**Promo**	5_Promo	8	market	opportunity	investment	concept	preview	trailer	advertis
Formats	7		**Review**	7_Review	5	reaction	success	assess	review	rate		
Formats	6		**Record**	6_Record	10	dashcam	footage	material	recording	livestream	live	realtime
Formats	8		**Instruction**	8_Instruction	3	tutorial	explain	how				
Formats	9		**Humour**	9_Humour	8	prank	sketch	parody	comedy	comic	funny	joke
Formats	10		**Other**	10_Other	12	game	NBA	Championship	quarter	player	lyrics	basketball
Features	1	Accidents	**Accidents**	1_Accidents	5	crash	casualty	collision	collide	struck		
Features	1	Accidents	**Safety**	1_Safety	10	save	lives	warning	hazard	lives	risk	safe
Features	1	Accidents	**Death**	1_Death	5	kill	dead	wreck	manslaught	death		
Features	1	Accidents	**Errors**	1_Errors	9	error	mistake	deviant	failure	drunk	transgress	wrong
Features	2	Tech	**Automation**	2_Automation	9	machine	automat	robot	algorithm	autonomous	self-	agent
Features	2	Tech	**Intelligence**	2_Intelligence	10	predict	learn	decide	navigat	find	cognitive	percept
Features	2	Tech	**Computationa**	2_Computatio	13	adaptive	assist	detect	identif	lidar	radar	autopilot
Features	3	Social	**Human_Actor**	3_Human_Act	17	pedestrian	cyclist	driver	human	CEO	woman	friends
Features	3	Social	**Everyday**	3_Everyday	6	work	sleep	commute	workplace	live	home	
Features	3	Social	**Interaction**	3_Interaction	7	interact	avoid	behaviour	hands	steer	stop	society
Features	4	Business	**Innovation**	4_Innovation	14	leap	innovation	develop	startup	remake	change	perfect
Features	4	Business	**Business**	4_Business	5	company	invest	market	business	enterpr		
Features	4	Business	**Brands**	4_Brands	14	tesla	google	waymo	uber	bosch	bmw	mercedes
Features	5	Disruption	**Problematisat**	5_Problematis	5	controversy	delay	issue	disput	dilemma		
Features	5	Disruption	**Regulation**	5_Regulation	18	permit	legal	bans	law	sue	regulation	court
Features	5	Disruption	**Hacking**	5_Hacking	4	hack	holtz	DIY	comma			
Features	6	Environment	**Road**	6_Road	11	freeway	highway	street	curved	light	road	lane
Features	6	Environment	**Weather**	6_Weather	6	fog	storm	snow	rain	sun	clouds	
Features	6	Environment	**Environments**	6_Environmen	19	urban	city	world	neighbourhoo	copenhagen	greenwich	local
Features	6	Environment	**Mobility**	6_Mobility	11	miles	distance	race	speed	travel	kilometer	journey
Features	7	Car	**Car_Parts**	7_Car_Parts	6	door	wheel	pedal	brake	seat	modular	
Features	7	Car	**Cars**	7_Cars	6	vehicle	car	SUV	truck	wagon	automobile	
Features	8	Future	**Future**	8_Future	7	generation	future	forward	dream	inevitable	vision	years
Features	9	Other Transpo	**Other_Transp**	9_Other_Trans	8	metro	bus	taxi	bike	transport	minivan	bicycle

Figure 2. The Lexicon: Intelligent vehicle test drives on YouTube, Siegen, 21–22 April 2018.

Figure 2. Continued

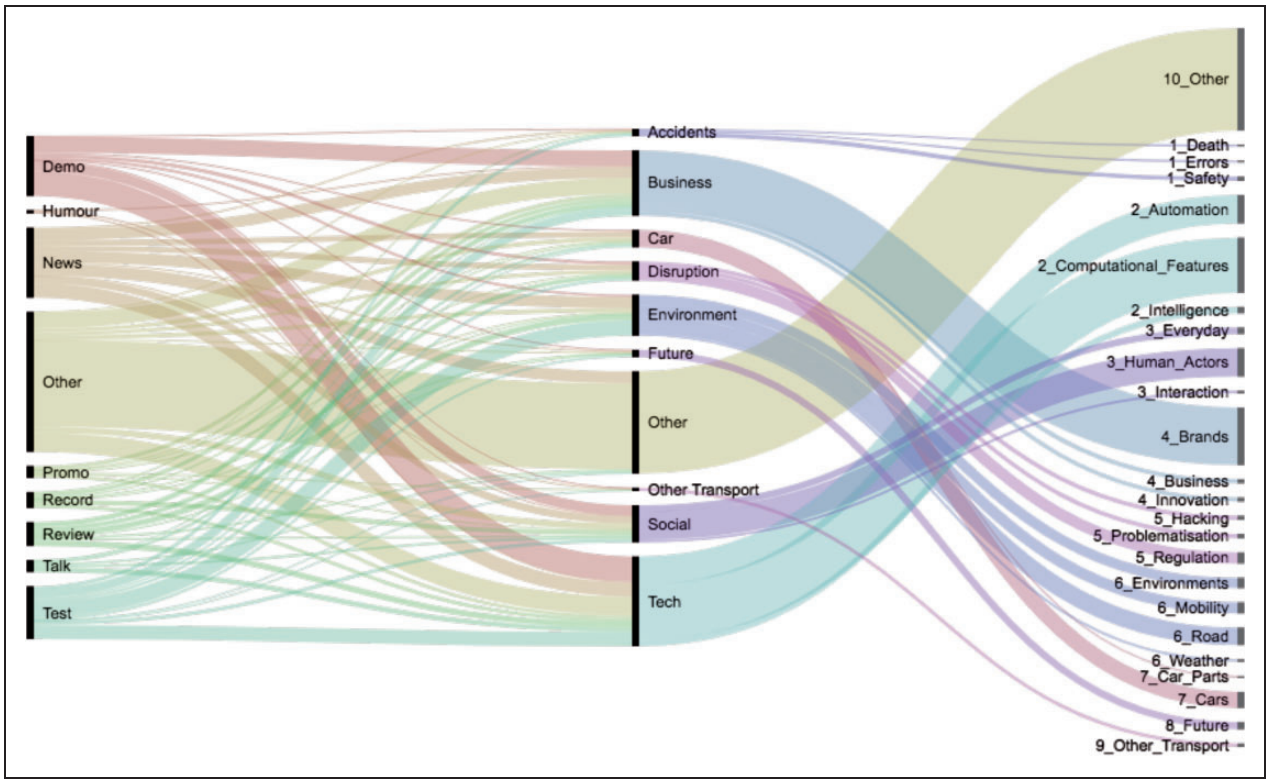

Figure 3. Correlating genre and feature; semi-automated lexicon analysis of 4052 driverless YouTube videos, Warwick/Siegen, April 2018.

our situational mapping to include media-specific elements, like channels.

Conclusion: From situational analysis to situational analytics

This article has identified some formative features of an interpretative approach to SA that I believe can make important contributions to the methodological refinement of computational social research. What are these features? First, SA explicitly recognises the participation of non-human elements, like technical architectures and media genres, in situations. As such, it enables us to study situations as distributed accomplishments, which are produced through the coming together of heterogeneous elements, including in situ occurrences (a car encounters a traffic sign on a road), demonstrational forms ('the test drive') and infrastructural effects (tweeting YouTube URLs). Second, SA offers an interpretative approach to data mapping with computational methods, making it possible to analyse situations marked by the *de-stabilisation* of routines, such as the introduction of new technology into the social environment. When using computational methods of data analysis, we tend to restrict ourselves to formal analysis, focusing on the detection of repeatable patterns across settings. This was also the case in our lexicon-based analysis. However, such a focus on the detection of regular patterns does not necessarily mean that we have to limit our analysis to ritualised, routine interactions. By adopting an empiricist approach-like SA, we can use computational methods to study less stable, disruptive, testing situations too.

Third, acknowledging the participation of digital infrastructure in situations enables us to address possible transformations of the situational fabric of digital societies. Keeping an open mind as to what constitutes a situation – and whether it is accomplished at all – allows us to analyse semi-situations, moment in which "aboutness" – the capacity of a situation to have a coherent referent – or accountability between actors is not necessarily realised. That is, the approach to SA put forward in this paper does *not* see it as its job to repair, or re-instate the aboutness of social life, and secure 'natural qualities' for the object of analysis by means of its analytic framework. The analyst's job is not to sustain naturalist definitions of social situations but to offer critical, empirically informed evaluations of the capacity of computationally mediated settings to surface situations. This also means that SA must expand its empirical scope and examine whether and

how distinctive features of situations – such as demonstration, problematisation, accountability – are being secured in computational environments in society.

But the approach presented here also moves beyond Clarke's SA in at least one decisive way: it proposes a way to automate and scale up the interpretative study of situations. Situational mapping can be done not only through reading and coding data but also with semi-automated lexicon-based analysis. The latter method involves a formalisation of SA, and this means that we face different kinds of challenges than situational analysts of ethnographic data: when applied to large data sets derived from online platforms, data analysis is inevitably inflected by media infrastructures in society: 'what makes a difference,' in our SA, is marked by infrastructural latencies, or what Amoore (2018) calls the subvisible –socio-technical architectures left implicit in digital interfaces. In computational settings it is notoriously opaque who or what participates in the situation, and accordingly, where it begins and ends. For this reason, situational analytics, in contrast to SA, almost inevitably moves beyond description, and involves *active curation not just of the data but of* the situation under study, through the establishment of relevance conditions: what is it necessary to take into account in order to grasp the situation qua situation? (Asdal and Moser, 2012). It is this need for curatorial work in order to actively surface situations from data with computational methods that I have in mind when calling for a move from SA to situational analytics.

To secure a space for interpretative analysis of situations in computational settings it is then not necessary to reject automated, formal data analysis as a social research methodology. Instead, it requires that we challenge *tendencies to naturalism* implicit in much contemporary work in computational social science, as well as in some forms of qualitative social research. Perhaps most of all, it requires that we commit to coming to terms, *methodologically speaking*, with the artificial – synthetic, heterogenous, compositional – nature of social phenomena in computationally-intensive societies.

Acknowledgements

Previous versions of this paper were presented at the EASST Conference (Lancaster University, 2018), the Oxford Internet Institute, the Centre de Sociologie de l'Innovation (MinesTech, Paris) and KCL's Digital Humanities Department, and I am grateful to colleagues at these institutions for organising these events, in particular Jonathan Gray.

Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Noortje Marres https://orcid.org/0000-0002-8237-6946

Notes

1. With thanks to Carolin Gerlitz, Britt Ross Winthereik and Hjalmar Bang Carlsen for their comments and suggestions without which I could not have written this article. Previous versions of this paper were presented at the EASST Conference (Lancaster University, 2018), the Oxford Internet Institute, the Centre de Sociologie de l'Innovation (MinesTech, Paris) and KCL's Digital Humanities Department, and Im grateful to colleagues at these institutions for organising these events, in particular Jonathan Gray.
2. I limit my discussion to social science applications of informatics and AI-based approaches to *textual analysis*, mostly for pragmatic reasons: it is an area with which I am familiar.
3. 3. Nguygen follows literature in socio-linguistics in setting as her objective the determination of stable and transferable features of social situations: 'the project of describing the varieties of language focuses on the constant features of the situational circumstances of language events, that can be consistently related to varieties in the language texts' (Gregory and Carroll, 1978: 10).
4. On this point, see also Marwick and Boyd (2010), who note that even the basic question of *who* constitutes the active and potentially active participants is ambiguous in many mediated settings. In most face-to-face situations the relevant participants are monitor-able by the interactants themselves, but in mediated settings one's audience is not so easily defined.
5. I use the -able to denote a potential for certain operations, rather than their actuality. As sociologist like GH Mead have long argued, the relevance of observation for social life is not limited to the actual monitoring of social life by actual actors, but as a possibility may inflect social life most decisively (Adkins and Lury, 2011).
6. Actor-network theory and ANT-inspired methodologies, such as controversy analysis and issue mapping, make a similar, empiricist assumption. This approach relies on network analysis to specify social ontology: who are the actors? what are the issues? where are they happening? (Marres, 2015). Situational analysis adds something to this: recognition of how the *settings* of social life inflect how it unfolds. Implementing situational analysis in computational environments, then, allows us to address the 'problem of the setting': the question of how the where of social life – its location – participates in the articulation of issues.
7. This term was suggested by Fabian Muniesa. pers. com.

8. As Hlajmar Bang Carlsen pointed out, the situation can be understood as a kind of meta- or infra-frame, that must be able to withstand disagreement at a lower level. This highlights situations are curat-able, which becomes more relevant in digital societies, where the boundaries of synthetic situations are not given, and neither is their composition.
9. Our initial video selection was thus theory-led and not in any way representative of the available population of self-driving video's on YouTube. It also means that the first, qualitative, part of our study, actually searched for 'testing situations' in the data, and only in the second part did we adopt the evaluative stance to establish whether testing situations involving self-driving cars are reported with YouTube.
10. A initial list of videos featuring street tests of intelligent vehicles was drawn from a variety of sources – collected news articles, colleagues' recommendations, the YouTube recommendation system.
11. '*Situational maps* lay out all the major human, nonhuman, discursive, historical, symbolic, cultural, political, and other elements in the research situation of concern' (Clarke and Charmaz, 2019: 15).
12. Transcription of the situational maps produced in sticky notes. Colours indicate whether elements contribute positively (green) or negatively (red) to articulating the introduction of self-driving cars into the street as a testing situation. The X-axis moves from rich (−1) vs. poor (+1) test environment, the Y-axis moves from radical innovation (top) to incremental innovation (bottom).
13. The Twitter data set was collected by the Academic Tech team at the Centre for Interdisciplinary Methodologies, University of Warwick using TCAT (Borra and Rieder, 2012). The data set for the relevant period contained 874.107 Tweets.
14. This script was coded by James Tripp, and has since been developed into the data tool Le-CAT, see https://warwick.ac.uk/fac/cross_fac/cim/tools
15. The category 'other' was added to the categories under 'feature' in the last stage of the visualisation of the Lexicon analysis results during the Siegen workshop in April 2018. I believe this reflects the proportion of videos for which no queries listed for any category obtained (no positive results), but this is not entirely consistent, as the type 'genre' also includes the category 'other' for which the lexicon does list specific query terms. Besides this inconsistency, an additional limitation of the exercise was that the lexicon contained a highly uneven number of queries for different categories, which at least partly accounts for differences in frequency.

References

Adkins L and Lury C (2011) Introduction: Special measures. *The Sociological Review* 59(2_suppl): 5–23.

Alaimo C and Kallinikos J (2019) Social media and the infrastructuring of sociality. *Research in the Sociology of Organizations* 62: 289–306.

Amoore L (2018) Cloud geographies: Computing, data, sovereignty. *Progress in Human Geography* 42(1): 4–24.

Ampofo L, Collister S, O'Loughlin B, et al. (2015) Text mining and social media: When quantitative meets qualitative and software meets people. In: Procter R and Halfpenny PJ (eds) Innovations in Digital Research Methods. London; Thousand Oaks, CA: SAGE, pp.161–192.

Arthurs J, Drakopoulou S and Gandini A (2018) Researching YouTube. *Convergence: The International Journal of Research into New Media Technologies* 24(1): 3–15.

Asdal K and Moser I (2012) Experiments in context and contexting. *Science Technology and Human Values* 37(4): 291–306.

Bechmann A and Bowker GC (2019) Unsupervised by any other name: Hidden layers of knowledge production in artificial intelligence on social media. *Big Data & Society* 6(1).

Boltanski L and Thévenot L (1999) The sociology of critical capacity. *European Journal of Social Theory* 2(3): 359–377.

Borra E and Rieder B (2012) Programmed method: Developing a toolset for capturing and analyzing tweets. *Aslib Proceedings* 66(3):

Brown B and Laurier E (2017) The trouble with autopilots: Assisted and autonomous driving on the social road. In: *Proceedings of the 2017 CHI conference on human factors in computing systems*, Denver, Colorado, pp.416–429.

Castelle M (2020) The social lives of generative adversarial networks. In: *Proceedings of the 2020 Conference on Fairness, Accountability and Transparency*, Barcelona, pp.413–413.

Clarke AE (2005) *Situational Analysis: Grounded Theory Mapping after the Postmodern Turn*. London; Thousand Oaks, CA: SAGE.

Clarke AE and Charmaz K (2019) Grounded Theory and Situational Analysis. In: Atkinson P, Delamont S, Cernat A, Sakshaug JW and Williams RA (eds) Sage Research Methods. London; Thousand Oaks, CA: Sage.

Cointet JP and Parasie S (2018) Ce que le big data fait à l'analyse sociologique des textes Un panorama critique des recherches contemporaines. *Revue Française de Sociologie* 59(3): 533–557.

Derksen M and Beaulieu A (2011) Social technology. In: Jarvie I and Zamora Bonilla J (eds) The Handbook of Philosophy of Social Science. London; Thousand Oaks, CA: Sage, pp.703–719.

Dieter M, Gerlitz C, Helmond A, et al. (2019) Multi-situated app studies: Methods and propositions. *Social Media+ Society* 5(2).

Goffman E (1964) The neglected situation. *American Anthropologist* (6): 133–136.

Gregory M and Carroll S (1978) *Language and Situation: Language Varieties and Their Social Contexts*. London; New York: Routledge.

Gross A (2016) Data Types and Functions: A Study of Framing Devices and Techniques. PhD Thesis, University of Warwick, UK.

Hine C (2000) *Virtual Ethnography*. London; Thousand Oaks, CA: Sage.

Kelleher JD and Tierney B (2018) *Data Science*. Cambridge, MA: MIT Press.

Kelly AH (2012) The experimental hut: Hosting vectors. *Journal of the Royal Anthropological Institute* 18: S145–S160.

Knorr Cetina K (2014) Scopic media and global coordination: the mediatization of face-to-face encounters. In: Lundby K (ed.) Mediatization of Communication. Berlin: de Gruyter, pp.39–62.

Knorr Cetina K (2009) The synthetic situation: Interactionism for a global world. *Symbolic Interaction* 32(1): 61–87.

Latour B (1993) *The Pasteurization of France*. Cambridge, MA: Harvard University Press.

Lazer D, Pentland A, Adamic L, et al. (2009) Computational social science. *Science* 323(5915): 721–723.

Lindgren S (2020) *Data Theory*. Cambridge: Polity.

Maguire J and Winthereik BR (2019) Digitalizing the state: Data centres and the power of exchange. *Ethnos*. DOI: 10.1080/00141844.2019.1660391

Marres N (2017) *Digital Sociology: The Reinvention of Social Research*. Cambridge: Polity.

Marres N (2015) Why map issues? On controversy analysis as a digital method. *Science, Technology & Human Values* 40(5): 655–686.

Marwick AE and Boyd D (2010) I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society* 13(1): 114–133.

Missche A and White H (1998) Between conversation and situation: Public switching dynamics across network domains. *Social Research* 65 (3): 695–724.

Nassauer A and Legewie NM (2018) Video data analysis: A methodological frame for a novel research trend. *Sociological Methods & Research*. DOI: 10.1177/0049124118769093.

Nguyen DP (2017) Text as social and cu*ltural data: A computational perspective on variation in text*. PhD Thesis, University of Twente, The Netherlands.

Rawls A (2008) Introduction. In: Garfinkel H and Rawls A (eds) *Toward a Sociological Theory of Information*. London; New York: Routledge, pp.1–100.

Rheinberger HJ (1997) *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford, CA: Stanford University Press.

Ruppert E, Law J and Savage M (2013) Reassembling social science methods: The challenge of digital devices. *Theory, Culture & Society* 30(4): 22–46.

Salganik M (2019) *Bit by Bit: Social Research in the Digital Age*. Princeton, NJ: Princeton University Press.

Savage M and Burrows R (2007) The coming crisis of empirical sociology. *Sociology* 41(5): 885–899.

Star SL (1999) The ethnography of infrastructure. *American Behavioral Scientist* 43(3): 377–391.

Stilgoe J (2018) Machine learning, social learning and the governance of self-driving cars. *Social Studies of Science* 48(1): 25–56.

Seaver N (2015) The nice thing about context is that everyone has it. *Media, Culture & Society* 37(7): 1101–1109.

Selbst AD, Boyd D, Friedler SA, et al. (2019) Fairness and abstraction in sociotechnical systems. In: *Proceedings of the Conference on Fairness, Transparency and Accountability (FAT), Atlanta USA*, pp.59–68.

Smith DE (1990) *The Conceptual Practices of Power: A Feminist Sociology of Knowledge*. Toronto: University of Toronto Press.

Törnberg A and Törnberg P (2016) Muslims in social media discourse: Combining topic modeling and critical discourse analysis. *Discourse, Context & Media* 13: 132–142.

Voigt R, Camp NP, Prabhakaran V, et al. (2017) Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences* 114(25): 6521–6526.

Woolgar S and Neyland D (2013) *Mundane governance: Ontology and accountability*. Oxford: Oxford University Press.

Zubiaga A, Liakata M and Procter R (2017) Exploiting context for rumour detection in social media. In: *International Conference on Social Informatics*, Oxford (UK), pp.109–123. New York: Springer.