

Beyond the “c” and the “x”: Learning with algorithms in massive open online courses (MOOCs)

Jeremy Knox¹

Published online: 13 February 2018

© The Author(s) 2018. This article is an open access publication

Abstract This article examines how algorithms are shaping student learning in massive open online courses (MOOCs). Following the dramatic rise of MOOC platform organisations in 2012, over 4,500 MOOCs have been offered to date, in increasingly diverse languages, and with a growing requirement for fees. However, discussions of *learning* in MOOCs remain polarised around the “xMOOC” and “cMOOC” designations. In this narrative, the more recent *extended* or *platform* MOOC (“xMOOC”) adopts a broadcast pedagogy, assuming a direct transmission of information to its largely passive audience (i.e. a teacher-centred approach), while the slightly older *connectivist* model (“cMOOC”) offers only a simplistic reversal of the hierarchy, posing students as highly motivated, self-directed and collaborative learners (i.e. a learner-centred approach). The online nature of both models generates data (e.g. on how many times a particular resource was viewed, or the ways in which participants communicated with each other) which MOOC providers use for analysis, albeit only after these data have been selectively processed. Central to many learning analytics approaches is the desire to predict students’ future behaviour. Educators need to be aware that MOOC learning is not just about teachers and students, but that it also involves *algorithms*: instructions which perform automated calculations on data. Education is becoming embroiled in an “algorithmic culture” that defines educational roles, forecasts attainment, and influences pedagogy. Established theories of learning appear wholly inadequate in addressing the agential role of algorithms in the educational domain of the MOOC. This article identifies and examines four key areas where algorithms influence the activities of the MOOC: (1) *data capture and discrimination*; (2) *calculated learners*; (3) *feedback and entanglement*; and (4) *learning with algorithms*. The

✉ Jeremy Knox
jeremy.knox@ed.ac.uk

¹ Centre for Research in Digital Education, University of Edinburgh, Edinburgh, UK

article concludes with a call for further research in these areas to surface a critical discourse around the use of algorithms in MOOC education and beyond.

Keywords massive open online course (MOOC) · algorithm · learning · student data · learning analytics

Résumé Au-delà du « c » et du « x » : apprendre avec des algorithmes dans les formations en ligne ouverte à tous (FLOT) – Cet article examine la façon dont les algorithmes influencent l'apprentissage dans les formations en ligne ouverte à tous (FLOT ou MOOC). À la suite de l'essor fulgurant en 2012 des organismes de plateformes FLOT, plus de 4500 de ces formations ont été proposées jusqu'ici, dans un nombre croissant de langues et avec une demande croissante de participation financière. Les débats sur l'*apprentissage* dans ces formations demeurent cependant polarisés autour des appellations xMOOC et cMOOC. Dans ce récit, le cours en ligne MOOC *étendu ou plateforme* (xMOOC), plus récent, adopte une pédagogie diffusée et fondée sur une transmission directe de l'information à un auditoire essentiellement passif (approche centrée sur l'enseignant), tandis que le modèle *connectiviste* (cMOOC) légèrement plus ancien opère uniquement un renversement simpliste de la hiérarchie, prenant les participants pour des apprenants très motivés, auto-dirigés et collaboratifs (approche centrée sur l'apprenant). La formule en ligne des deux modèles génère des données (par exemple fréquence de consultation d'une ressource donnée, modes de communication entre les participants), que les prestataires des MOOC exploitent à des fins d'analyse, néanmoins uniquement après avoir procédé à leur traitement sélectif. Au cœur de nombreuses approches d'analyse de l'apprentissage se trouve le désir de prédire le comportement futur des apprenants. Les éducateurs doivent être conscients du fait que l'apprentissage dans les cours en ligne ne dépend pas seulement des enseignants et des apprenants, mais aussi des *algorithmes* : les instructions qui effectuent des calculs automatisés sur les données. La formation devient de plus en plus mêlée à une « culture algorithmique » qui définit les rôles pédagogiques, la réalisation des prévisions, et qui influence la pédagogie. Les théories établies de l'apprentissage semblent totalement inadéquates pour aborder le rôle agentif des algorithmes dans le champ éducatif des MOOC. Cet article identifie et examine quatre principaux domaines d'influence des algorithmes sur les activités du cours en ligne : 1) *capture et discrimination de données*, 2) *apprenants évalués par le système*, 3) *retour d'information et enchevêtrement*, 4) *apprentissage avec les algorithmes*. L'article conclut par un appel à une recherche complémentaire dans ces domaines pour susciter un discours critique sur l'usage des algorithmes dans les cours en ligne et au-delà.

Introduction

Following the dramatic rise of massive open online course (MOOC) platform organisations in 2012, over 4,500 MOOCs have been offered to date, in increasingly diverse languages, and with a growing requirement for fees (OCR 2016). While their emergence has been shaped by media hype (see Adams 2012; Friedman 2013),

and shrouded by myth and paradox (Daniel 2012), two conflicting narratives have tended to dominate the discussion: those of the “cMOOC” and the “xMOOC”. While problematic as course categories (Bayne and Ross 2014), these designations correspond to the development and research of MOOCs in two chronological phases (Ebben and Murphy 2014), but also to different ideological and theoretical stances about the nature of learning, and the role of the educational institution (Knox 2015).

The earlier “cMOOCs” (or “connectivist” MOOCs) foreground “human agency, user participation, and creativity through a dynamic network of connections afforded by online technology” (Ebben and Murphy 2014, p. 333). These experimental courses position learning as the formation and utilisation of networks, underpinned by the proposed learning theory of “connectivism”, suggested to be distinct from more established concepts of behaviourism, cognitivism and social constructivism (Anderson and Dron 2011).¹ The educational content in cMOOCs is distributed amongst various social networking platforms, and is often generated by participants, necessitating highly motivated, self-directed individuals capable of navigating and evaluating diverse online resources. Much of the research emerging from this phase of the MOOC was concerned with the motivation of individuals and “aimed at understanding the under-performing participant” (Ebben and Murphy 2014, p. 334). In other words, successful learning in these courses is determined by the capacities of the individual alone, and the digital technologies of the cMOOC are largely considered as passive instruments for cohesive community networking.

Larger-scale MOOC platform organisations have subsequently surfaced, most notably *Coursera*, *edX*, *Udacity* and *FutureLearn*. Attracting considerably more participants than the distributed variety, these so-called “xMOOCs” (or “extended” MOOCs) have involved high-profile partnerships with elite universities, and operated on dedicated software platforms. Educational content in these courses has largely taken the form of streamed video lectures, broadcasting pre-recorded, centralised material to the entire class of participants. Thus, xMOOCs have tended to adopt a behaviourist pedagogy (Rodríguez 2013), and it is the suggested scalability of this approach (Anderson and Dron 2011) that has underpinned the often grand claims of global provision and universal access (Knox 2016). Correspondingly, xMOOC research has been dominated by the computational analysis of large amounts of user data. From its earliest incarnations, the xMOOC has been inextricably linked with the field of “learning analytics” (Ebben and Murphy 2014), which emerged aiming to not only provide new understandings of learning in MOOCs, but also answer “a multitude of questions about how humans learn and interact” (McKay 2013).

¹ The proposed learning theory of *connectivism* foregrounds the “network” as a literal model for learning, emphasising the ability to connect with sources of information using technology, rather than the need to retain knowledge (see Siemens 2005). The concept of *behaviourism* in education tends to posit that learning can be discerned through the observation of external behaviours, and that behaviours can be shaped through the reinforcement and repetition of engagement with external stimuli. *Cognitivism* situates learning within internal mental activities, rather than external behaviours, and seeks to change an individual’s internal way of thinking. *Social constructivism* shifts the focus on learning again, this time towards social interactions rather than the internal minds of individuals. In this view, knowledge is “constructed” through interactions with others.

To date, MOOC research has been categorised into three broad areas: (1) student profiling; (2) measurements of student progress and attainment; and (3) teaching methods (Breslow 2016), all drawing from the vast quantities of educational “big data” accompanying course offerings. Indeed, it is the sheer volume of participation that calls into question the value of “qualitative” methods in the MOOC domain. To examine “massive” numbers of participants in such “manual” ways would not only be too time-consuming and labour-intensive, but would also deny the supposedly new (and “objective”) insights that could come from the computational analysis of data sets no single human could undertake.

While further elaborations of the distinctions between c- and x-MOOC approaches can be found elsewhere (Knox 2015), the central thrust of this article is to highlight the tendency to overlook the significant role of technology in the learning process of MOOC students. So far, research has mainly focused on the collaborative construction of knowledge from communicative relations amongst peers or the behaviour of learners using a particular platform software. This frames the necessary technology as either a passive instrument for the self-directed networking of its human users, or a straightforward tool for universal access to the university offering the course. In attempting to foreground the significant influence of technology in the MOOC domain, this article focuses on one specific aspect: *algorithms*. These automated processes which perform calculations on data operate either within MOOC technologies themselves, or on user data in the form of research.

Critical of the later xMOOC offerings, Martin Weller defines a “Silicon Valley narrative” in the portrayal of MOOC technology:

There are several necessary elements ... firstly that a technological fix is both possible and in existence; secondly that external forces will change, or disrupt, an existing sector; thirdly that whole-sale revolution is required; lastly that the solution is provided by commerce (Weller 2015).

As we shall see, this vision of technological solutionism,² masking an undercurrent of proprietary ownership and profitability, also appears to encompass nascent developments in data computation and algorithmic educational design.

Algorithmic cultures

The critical study of algorithms is becoming established in software studies and digital sociology, often examining not only their technical functions, but also the ways in which they influence culture, politics and economics, becoming “powerful and consequential actors in a wide variety of domains” (Ziewitz 2015, p. 3). However, attempts to define algorithms have varied considerably, attesting to both

² *Solutionism* here refers to the work of Evgeny Morozov (see Morozov 2013). This is an underlying approach to the design and development of technology, typically involving the defining of a social “problem”, for which a technological system provides a corresponding “solution”. It is used as a critical term, in the sense that the simplistic definitions of complex social “problems” cause the development of inadequate “solutions”.

their ubiquitous and powerful influence over our lives, as well as to their simultaneous incomprehensibility and inscrutability (Ziewitz 2015). At a technical level, an algorithm might be understood simply as

encoded procedures for transforming input data into a desired output, based on specified calculations (Gillespie 2014, p. 167).

Nevertheless, their persistence in shaping various facets of contemporary society has invited suggestions of an “algorithmic culture” (Striphas 2015) or a “computational theocracy” (Bogost 2015). This signals much more broad and complex understandings of the algorithm, as involved in the proceduralisation of knowledge, and as a result, the formalising and delineating of social life. While not an uncontested idea in the social sciences, there is “broad agreement that algorithms are now increasingly involved in various forms of social ordering, governance and control” (Williamson 2014).

Relating specifically to digital technology, given the prominence of web search in ordering and privileging particular sources of knowledge, and the pervasiveness of social media in organising the communications and interactions between increasing numbers of people, the algorithms that underpin and control these services have garnered considerable attention from the academic community. Through web search, algorithms take on the work of culture: “the sorting, classifying and hierarchizing of people, places, objects and ideas” (Striphas 2015, p. 396), and through social media, they are “increasingly vital to how we organize human social interaction” (Gillespie 2014). For Ted Striphas, “*algorithms* are becoming decisive, and ... companies like Amazon, Google and Facebook are fast becoming, despite their populist rhetoric, the new apostles of culture” (Striphas 2015, p. 407, emphasis original).

This is a crucial insight for the discussion of algorithms in education, because the “Silicon Valley narrative” (Weller 2015) is also gaining traction as a set of powerful and plausible ideas about the benefits of the large-scale data mining and computational analysis of learner data. Significantly, it is the claim, not only of objectivity in the discovery of educational insights, but also of “public-” and “crowd-” based evidence that forms such a seductive narrative. Where the educational institution is framed as antiquated, elitist and “broken” (Weller 2015), the idea of student behaviour (in the form of data) driving pedagogical decisions and revolutionising educational research satisfies a simplistic vision of learner-centred solutionism.

Three key interrelated principles are central to developing a critical understanding of the ways in which algorithms challenge fundamental assumptions about knowledge and subjectivity in the context of education. First, algorithms must be understood to *produce* the conditions they purportedly represent, rather than *discover* an anterior reality or truth (Perrotta and Williamson 2016). In short, algorithms are not passive arbiters for objective insights. As Tarleton Gillespie demonstrates in an examination of the microblogging service Twitter (Gillespie 2011, 2014), the occurrence of “trending topics” is not in fact an already existing social phenomenon, but rather reflects intricately constructed “realities” produced through the workings of complex algorithms. However, the very premise of services

like Twitter is one in which such calculations are presented as direct and transparent representations of public life. The Twitter algorithms are thus involved in “curating a list whose legitimacy is based on the presumption that it has not been curated” (Gillespie 2011). This “aura of truth, objectivity, and accuracy” (Boyd and Crawford 2012, p. 663) that accompanies data analytics is of central concern in education, particularly where finite measurements of success appear to dominate the agenda. While the attraction of learning analytics may be the potential discovery of novel patterns of learning behaviour, it should also be recognised as involving a seamless alignment with “the logic of economic rationality and ‘accountability’ that pervades governance cultures in education” (Perrotta and Williamson 2016, p. 4).

Second, a recursive and non-deterministic relationship must be understood in the ways in which algorithms interact with educational practices and experiences. In other words, while some kind of (non-human) agency must be recognised in algorithms, it is not one which functions independently of the various human beings involved in their operation, from programmers to end users. The social and the algorithmic are entwined at every stage. Drawing on the work of sociologist Scott Lash, David Beer describes the power of algorithms, not in terms of “someone having power over someone else, but of the software making choices and connections in complex and unpredictable ways in order to shape the everyday experiences of the user” (Beer 2009, p. 997). However, this condition of power must not be understood as the external influence of a peripheral algorithm on the internal state of the (human) learner. Algorithms are already inextricably part of the cultures in which they operate, where “culture” and “technology” themselves are constantly shifting ideas, practices and materials (Striphas 2015). To acknowledge the non-deterministic relationship between algorithm and society, Ben Williamson proposes the “socioalgorithmic”, rather than “algorithmic power”, stressing that algorithms “are socially produced through mixtures of human and machine activities, as well as being socially productive” (Williamson 2014). As Rob Kitchin and Martin Dodge succinctly note, “algorithms are products of knowledge about the world” which “produce knowledge that is then applied, altering the world in a recursive fashion” (Kitchin and Dodge 2011, p. 248). Our assumptions about the process of learning are encoded into the procedural routines of analytics, whereby those same assumptions function to produce educational realities. The concern for education is that the assumed objectivity of data analytics “inevitably leads to reifying the outputs of those analyses as equally neutral, objective and natural phenomena” (Perrotta and Williamson 2016, p. 9). Furthermore, and crucially for educational concerns, this embedded relationship with algorithmic processes challenges established theories which locate learning exclusively within or amongst human beings (this aspect will be further addressed below).

Third, the concealment of algorithmic functioning presents particular concerns for education. Examining popular social media, Striphas contends that:

thanks to trade secret law, nondisclosure agreements and noncompete clauses, virtually none of us will ever know what is “under the hood” at Amazon, Google, Facebook or any number of other leading tech firms (Striphas 2015, p. 407).

While some educational analytics may be similarly proprietary, the prospect of laying bare the inner workings of algorithms is not just a matter of protecting competitive market advantage. If educational analytics are indeed capable of rendering unprecedented and valuable patterns of learning behaviour, there may be a necessary clandestine nature to the processes involved, particularly where assessment is concerned. As Gillespie suggests, “revealing the workings of their algorithm ... risks helping those who would game the system” (Gillespie 2011). Nevertheless, “knowing algorithms and their implications becomes an important methodological and political concern” (Ziewitz 2015, p. 4) in contemporary society, no less in education, particularly where critical understanding is the aim. If public education is to be subjected to unseen algorithmic operations, one might expect those processes to be as transparent as current assessment routines and criteria, for example.

However, such ideals are not necessarily achievable where highly complex calculations are concerned. Even if algorithms are made visible, the ability to understand them, and potentially manipulate their results, is not immediately apparent to all. Openness, despite the rhetoric, would not be a simple cure. As Gillespie contends, “[w]e don’t have a language for the unexpected associations algorithms make, beyond the intention (or even comprehension) of their designers” (Gillespie 2011). Furthermore, the mutable condition of algorithms is one in which they “are made and remade in every instance of their use because every click, every query, changes the tool incrementally” (Gillespie 2014, p. 172). Algorithmic literacy in such a scenario would be a persistent and ever-changing undertaking.

With such concealment and complexity, it has perhaps been easy for algorithms to slip silently into habit:

the sinking of software into our mundane routines, escalated by mundane technologies such as those found in the popular social networking sites, means that these new vital and intelligent power structures are on the inside of our everyday lives (Beer 2009, p. 995).

As Williamson has shown, the increasing prevalence of educational data science is now embedding algorithmic processing in the everyday, mundane practices of education (Williamson 2014, 2015a, b). With the aim of exposing the overlooked algorithms operating below the surface of MOOCs, the following section will review this developing relationship, drawing on established critical analyses of algorithms (Gillespie 2014). Four key areas where algorithms hold influence are identified and examined: (1) *data capture and discrimination*; (2) *calculated learners*; (3) *feedback and entanglement*; and (4) *learning with algorithms*.

MOOC research

Data analytics have, for some time, been highlighted in educational horizon scanning (Johnson et al. 2016), signalling the imminent and powerful mainstreaming of predictive and interventionist data science in formal education (Williamson 2014, 2015a, b). MOOCs have played a key role in these developments, offering a

“massive research agenda for mining data” (Ebben and Murphy 2014, p. 343) and a fertile space for the field of learning analytics. The premier conference in this emerging field, Learning Analytics and Knowledge (LAK), has seen increasing contributions that deal with MOOC data specifically,³ and the inclusion of dedicated panels for MOOC research.⁴ Following from the introduction to critical algorithm studies, the point of this section is not to argue that algorithms are “corrupting” education provided through MOOCs, or indeed that they are “influencing” an authentic or ordinary educational practice from the outside. Rather, the point is to highlight the already inseparable role of algorithms immanent to the project of the MOOC, and to show that such an arrangement challenges established ideas about learning in this high-profile educational space.

Data capture and discrimination

Given that, at a technical level, an algorithm is a set of instructions, data are required for it to function. There are two principal concerns here for the MOOC. First, algorithmic processing can only provide insights about the learning process according to the kind of data available to it. The ease of access lauded in MOOC platform promotion is the very factor that conditions the kind of data a course can generate. Delivered through a web browser and centralised in a single platform, MOOC data can only derive from user interactions with course materials and basic web profiling. This may present a very limited scope for understanding behaviour, compared to the broad range of performance that accompanies any learning experience. The conclusions of data analytics, therefore, tend to be drawn based on relatively little information. While an algorithmic approach can tend to promote confidence in such “sufficient approximation” (Gillespie 2014, p. 174), the desire for more data appears pervasive, particularly in the field of learning analytics (Hildebrandt 2016, 2017). In predicting future MOOC directions, Robin Middlehurst foresees a broader convergence and integration of technologies across different platforms and devices (Middlehurst 2016). It is unlikely that such developments are pursued exclusively with user experience in mind, and they are certain to capture and align learner data in ever more persistent ways. Highlighting the ethical dilemmas of data collection, Mireille Hildebrandt warns of questionable insights deriving from non-contextual data, potentially contributing to results of education-specific information (Hildebrandt 2016). The future of learning analytics exists in tension with the future of student privacy. While concerns for data discretion are pressing, the specifically algorithmic processing of such data adds additional, highly complex dimensions to the dilemma. Whatever part of the learning process is not algorithmically readable will not become part of the learning analytics process, and therefore will not be part of the resulting “performance

³ The third LAK conference, held in 2013, included 3 papers mentioning MOOC data specifically (see <http://dl.acm.org/citation.cfm?id=2460296>), while the number in the sixth LAK conference, held in 2016, already totalled 19 (see <http://dl.acm.org/citation.cfm?id=2883851&picked=prox>) [both links accessed 5 January 2018].

⁴ For example, panels on “discussion analysis” (see <http://dl.acm.org/citation.cfm?id=2883851&picked=prox>) [accessed 5 January 2018].

matrix” (Hildebrandt 2016). Prominence given to specific kinds of analysis, for example that of analysing clickstream data from video lectures, results in particular educational activities and behaviours being privileged above others, on the grounds that data and analytics practices become easily available. If this kind of learning analytics is perceived to be successful, it encourages the use of video lectures in course design, and privileges this kind of activity as of central importance to learning.

Second, it would be disingenuous to separate algorithm and data on the basis that the former is involved in simply processing the latter. The critique of “raw” data is well established (see Boellstorff and Maurer 2015), calling into question the idea that data exist in an unrefined and natural form, merely waiting for algorithmic analysis to “make sense” of them. As Gillespie points out in what he terms “patterns of inclusion” (Gillespie 2014, p. 168), algorithms can only function if the data have been captured and categorised in a specific way that is “readable” by the software routine. The underlying point here is that, before the algorithm even begins to provide its calculative insights, the data have already undergone significant processing, such as selection and exclusion, ordering, and “correcting” incomplete or erroneous records. As Gillespie notes, this is not a simple or straightforward process, but one that already involves decisions about what counts as meaningful and relevant data (Gillespie 2014).

One prominent example of algorithmic analysis in MOOCs is the Coursera analytics dashboard (see Coursera 2014). Typical of many “dashboard” interfaces related to data analytics, the Coursera version provides a number of statistics and visualisations, including profiles of class enrolments and detailed analysis of assessment activities. However, despite whatever potential value there might be in such calculations (see Coursera 2014), the pre-processing of data may be a significant factor in what is finally presented. The explanation of the Coursera analytics dashboard reveals complex organising of data that takes place before the generation of visualisations and outputs, including the generation of “intermediate tables” of aggregated data (Coursera 2014). This in itself is an important analytical dimension, pertaining to the procedures required for data to be readable by algorithmic processes that will make educational judgements. As Gillespie describes, this is an activity that is selective, rather than indiscriminate: “information must be collected, readied for the algorithm, and sometimes excluded or demoted” (Gillespie 2014, p. 169). The discussion of the implications of the dashboard analytics is continued below.

Calculated learners

Gillespie defines “calculated publics” as the ways in which social media algorithms contribute to the construction of groups, communities and social affiliations that would not otherwise have existed: “[w]hen Amazon recommends a book that ‘customers like you’ bought, it is invoking and claiming to know a public with which we are invited to feel an affinity” (Gillespie 2014, p. 188). Yet, such associations are calculated by hidden, automated processes, and present a very different sense of social grouping to that conventionally understood as “social

networking”: that it is the human users of the software that are forming communities and driving the social exchange.

It is this same emphasis on “user behaviour” that has dominated MOOC research, and has similarly overlooked the role of algorithms in the construction of learner profiles and groups. The early phase of MOOC research frequently focused on the identification of learner “roles” (see for example: Breslow et al. 2013; Kizilcec et al. 2013; Perna et al. 2013). This constitutes a similar process of “calculated publics”, categorising MOOC participants into particular groupings and establishing these categories as tangible evidence for future research and practice. These associations have been shown to be determined directly and necessarily in relation to the features of the platform software, rather than deriving as exclusive characteristics of human behaviour (Knox 2016). However, it is the processes by which these affiliations are made that are important here: algorithmic methods that calculate “types” of MOOC learning, and group participants into associative categories.

The trend is still apparent. Rebecca Ferguson and Doug Clow identify “seven distinct patterns of engagement: Samplers, Strong Starters, Returners, Mid-way Dropouts, Nearly There, Late Completers and Keen Completers” (Ferguson and Clow 2015, p. 1). Tobias Hecking et al. identify user roles by comparing patterns of social and semantic exchange, and modelling discussion in terms of “information-seeking and corresponding information-giving posts” (Hecking et al. 2016, p. 198). They identify a “dominant” role of “regular relations” and two smaller roles of “information-seekers” and “information-providers” (ibid., p. 205). With a similar aim, Oleksandra Poquet and Shane Dawson analyse the formation of distinct networks of MOOC learners (2016). While Ferguson and Clow (2015) and Poquet and Dawson (2016) use “cluster analysis”, Hecking et al. (2016) use “blockmodelling”; both are algorithmic routines for producing categories of MOOC learners (for a more detailed analysis of clustering, see Perrotta and Williamson 2016). Ferguson and Clow are clear about the practical application of such “publics”, “clusters identified here can help inform a range of strategies for intervention and improvement” (Ferguson and Clow 2015, p. 7). Similarly, Hecking et al. recommend that “the design of asynchronous communication in online courses should consider better adaptivity to different needs of different user roles” (Hecking et al. 2016, p. 206).

These are calls for the direct crafting of future MOOC technology, and the explicit modifying of pedagogical practice according to social and linguistic structures that have been exposed through algorithmic processing. In this way, the cycles of MOOC delivery and research entail an “algorithmic presentation of publics back to themselves that shapes a public’s sense of itself” (Williamson 2014). While the automated “personalization” of MOOC technology has only been identified as a future development (Middlehurst 2016), at present the concretisation of particular “roles” for learning can be fed back through responsive pedagogical practices. The rendering of such “learner functions” may indeed be useful ways of understanding emerging educational practices in the MOOC phenomenon. However, these “roles” should be accompanied by more of a critical discourse around the ways in which they are generated. As Gillespie cautions, “the questions that

appear to sort us most sufficiently ... are likely to grow in significance as public measures. And to some degree, we are invited to formalize ourselves into these knowable categories” (Gillespie 2014, p. 174).

Feedback and entanglement

Key to understanding the role of algorithms in MOOCs are the feedback mechanisms through which research shapes futures technology development and pedagogical practice. The analytics dashboard offered by Coursera is one pertinent example of this potential in the MOOC:

The visualizations and metrics we present help instructors understand their learners and make informed decisions. By building user-friendly tools, we are making data a part of the everyday act of teaching (Coursera 2014).

Such dashboard visualisations are a key illustration of the ways in which the determining capacities of algorithms are hidden beneath the surface of MOOC technologies. Importantly, in the scenario described above, the “decisions” granted to the MOOC instructor are not related to the ways in which the dashboard visualisations have been calculated and displayed, but are rather conditioned as responses to the end results. What is overlooked in this archetypal “blackboxing” of technology is the fact that decisions have already been made with regard to how the data have been captured and processed, judgements already encoded into the algorithmic processes behind the dashboard. This concealed processing signals a significant problem for the role of the teacher as MOOCs continue to develop: teachers may not necessarily be aware of the specific mechanisms through which such “trustworthy”, and “authoritative” knowledge about their students has been generated. Without awareness, teachers appear to be required to act without full understanding of the educational contexts in which they are working. Becoming aware may impose significant requirements for “algorithmic literacy” upon already-demanding workloads.

Furthermore, this constitutes a questionable avoidance of responsibility for the pedagogical implications of revealing learner data on the part of MOOC dashboard providers, given that it is not just “data” which have become part of MOOC teaching in this scenario, but also the decisions of software engineers and their algorithms. As Gillespie asserts, “evaluations performed by algorithms always depend on inscribed assumptions about what matters, and how what matters can be identified” (Gillespie 2014, p. 177). Teaching interventions have already occurred in the production of data dashboards, well before the MOOC instructor looks at the visualisations and prepares to intervene. Further on, Coursera seems to acknowledge their role in the process, if only minimally: “We want to do far more to pull insight-needles out of the data-haystack, directing instructors’ attention to the most important patterns and points of interest” (Coursera 2014). It is this “directing attention” that is significant however, and is more than simply an impartial signalling: it is the co-construction of “importance” in the context of MOOC learners, of authoritative knowledge about this prominent education domain. The central point here is not to reject the idea of algorithmically generated data

dashboards as useful pedagogic devices, but rather to reject the idea that they provide a transparent window to objectively significant truths about the learning process.

The entanglements of algorithmic practices go further than feedback to potential MOOC teachers. At a much more profound level, the operative routines of algorithms are shaped by MOOC learners, but also assumed into normative forms of participation. As Gillespie argues, “[t]here is a case to be made that the working logics of ... algorithms not only shape user practices, but lead users to internalize their norms and priorities” (Gillespie 2014, p. 187). In the same manner in which users of social media actively influence the ways in which “search” and “recommender” algorithms function through their online activity, but also come to alter their own behaviour as a result, MOOC learners also participate in a co-constitutive arrangement. At present, however, algorithms are not operating within MOOC platforms to the extent that we see in social media, where user behaviour is often accompanied by simultaneous feedback from algorithmic processes, for example, watching videos in YouTube, and simultaneously being “recommended” additional video content. (For a fuller account of YouTube algorithms in the context of MOOCs, see Knox 2014.) This makes the feedback process less immediate, and perhaps less intense. However, cycles of MOOC development and research are facilitating this relationship, and as we have seen with the identification of “roles”, participants will have more opportunities to assume and internalise predetermined learner subjectivities.

Central to many learning analytics approaches is the desire to predict students’ future behaviour (for example Kennedy et al. 2015). This is a process that binds current MOOC participant activity to future cohorts of learners in increasingly concrete ways. This relates to the ways in which algorithms operate by “learning” and inferencing from large data sets: rendering patterns of behaviour from existing users that become models for the categorisation of future users. Majority behaviour now may define how judgements are made about future MOOC learners. This relates to what Gillespie terms “cycles of anticipation”, where “the perceptual or interpretive habits of some users are taken to be universal, contemporary habits are imagined to be timeless” (Gillespie 2014, p. 174). Such a view reveals the commitment to an objective, anterior “truth” to the learning process, transcendent of circumstance, that underpins the data science approach. There may be numerous future educational scenarios where the specific, contemporaneous context of learning is a much better measure than the behavioural activities of individuals from years in the past.

Learning *with* algorithms

Looking to a future with more embedded learning analytics, Hildebrandt proposed the idea of “learning as machine”, suggesting:

human beings increasingly live in a world saturated with data-driven applications that are more or less capable of machine learning. Since this will require human beings to anticipate how their intelligent environment

learns, I ... argue that – to some extent – humans will engage in “learning as if a machine” (Hildebrandt 2016).

As has been argued, it may be that MOOC participants are already embedded in an algorithmic culture that is increasingly shaping the learning process towards something that resembles the machinic. However, as has also been argued, this would not be a determinist relationship, and human agency would certainly be part of the recursive entanglement. Nevertheless, whatever the particulars of those arrangements might be in their specific contexts, the key insight here is that established theories of learning appear wholly inadequate in addressing the agential role of algorithms in the educational domain of the MOOC.

Individual learning “behaviour” may not be attributed solely to MOOC participants as they respond to educational resources, but might also involve algorithmic decision making that is fed back from data-intensive research. Similarly, it may not be entirely accurate to attribute the “social construction” of learning exclusively to the human beings actively participating in the MOOC community or network: algorithms also play a part in the calculating of groups and communicative practices of participants. Where this review of algorithmic activity in MOOCs differs significantly from social media studies is in the immediacy and potency of feedback mechanisms. While social media recommendation and search algorithms tend to provide a concurrent form of interaction and feedback for end users, MOOC algorithms operate largely in the sphere of research. While this creates a less direct and intense relationship with MOOC learners, the field as whole is grounded in a conceptual and practical commitment to “algorithmic education”. Given the xMOOC associations with Silicon Valley, it may be that social media-type algorithms begin to populate MOOC platforms in the future, and this will increase the need for critical algorithmic research in education, reflecting the analytic areas outlined in this article. Furthermore, the dedication to algorithms is not to be underestimated. Just as social media algorithms are engineered to suit the (ultimately economic) aims of their providers, rather than necessarily the experience of their end users (Gillespie 2011), MOOCs themselves may be grounded in institutional concerns rather than those of their learners. As Maureen Ebben and Julien Murphy contend, “the case could be made that edX is more about running a massive data collection experiment than about providing an education” (Ebben and Murphy 2014, p. 342).

Nevertheless, there is a present need to rethink the dominant assumptions about learning in MOOCs, in order to accommodate the ways in which algorithms intervene and shape the behaviour and communication of learners. However, as we have seen, such an approach may not be as straightforward as simply understanding how they function alongside, or within, what we might classify as behavioural, cognitive or constructivist “learning”. The much more profound question to address here is: *what happens to the very concept of “human” learning when fundamental insights about it are only intelligible to algorithmic processes?* The idea that data analytics offers novel and extraordinary educational insights is habitual, although implicit, in much of its promotion. The explanation of the Coursera analytics dashboard, for example, suggests:

The streams of data coming in from learners can give instructors an unprecedentedly detailed view into how learning happens, where they can make improvements, and future pedagogical directions to explore (Coursera 2014).

It is telling that the algorithms themselves are overlooked in this description, which appears to focus exclusively on the supposedly “raw” data streams. Nevertheless, it is the claim of “unprecedented” insight that is most significant here. If “how learning happens” at this scale is not discernible to humans alone, then the ability to know whether it has taken place is no longer in “our” hands. This positions algorithms as indispensable requirements for the future of the MOOC, and signals the reliance on automated data-intensive processes to deal with such educational activity. Gillespie warns of our increasing dependence on algorithms, grounded in the desire for simple, neutral calculations, free from human intrusion (Gillespie 2011). The habit of algorithmic intervention can only become more established in a world where we cannot understand what learning is without them. This prospect is both exciting and alarming for education. Where algorithms “are designed to work without human intervention ... and they work with information on a scale that is hard to comprehend (at least without other algorithmic tools)” (Gillespie 2014, p. 192), the potential for radically new, more-than-human educational insights appears tantalisingly on the horizon. However, by the very same description, learning seems to be pulled further from away from where we have always assumed it to be: within and amongst human beings. While the full implications of this shift are beyond the scope of this article, the implicit challenge to established theories of learning must be noted. If we continue to perceive learning as the social construction of knowledge, this theoretical foundation limits the scope of enquiry to the actions and responses of human beings, despite the claim that automated algorithms are making decisions and influencing behaviours amongst and in between social communication. In the age of algorithms, theories of learning need to be developed to take account of the more-than-human condition of agency. This work might look to concepts such as the “cognisphere”: “globally interconnected cognitive systems in which humans are increasingly embedded” (Hayles 2006, p. 161).

Conclusions

This article has reviewed the intervention of algorithms in the phenomenon of the MOOC. Dominated by the educational concepts and practices embodied in the “c-” and “x-” MOOC designations, learning in these courses has tended to be understood as either “behaviourist” or generally “constructivist”. However, both these assumptions appear to overlook the influence of technology on the learning process, and specifically the role of algorithms in the MOOC project. The critical study of algorithms has been outlined, drawing from fields outside of educational research, such as software studies and digital sociology. These perspectives call for a critical understanding beyond the functioning of algorithms, towards the ways in

which they influence cultural practices and individual subjectivities. Three key principles were highlighted: (1) the production of educational realities rather than the discovery of objective truths; (2) the recursive and co-constitutive relationship between algorithms and their human users; and (3) the educational dilemmas of concealing the working of algorithms.

Four areas of algorithmic influence in MOOC research and practice were identified and examined. First, “data capture and discrimination” stressed the limitations of the data produced by MOOC platforms, and the importance of data management. This suggested that it is not just the specific instructions encapsulated by the algorithm that are important to study, but also the selective processing that happens to data in order to make them recognisable to algorithmic routines. Data processing was highlighted in the example of the Coursera analytic dashboard.

Second, the notion of “calculated learners” examined the tendencies in MOOC research to categorise and group participants according to patterns in platform data. Such “roles” were shown to arise, not exclusively from learner behaviour, but also from algorithmic processes that “calculate” affiliations according to social networks and communicative practices. This practice embroils computational data processing in the formation of individual and group learner identities. More research is needed to understand the educational implications of algorithmically calculating groups of learners, where individuals are imbued with particular characteristics derived from cluster analysis or blockmodelling. Future work with MOOCs should recognise the ways in which learner roles are constructed through combinations of user behaviour and algorithmic process, rather than basing pedagogical and course-design decisions on the assumption of innate learning characteristics.

Third, “feedback and entanglement” outlined the ways in which data analytics research influences pedagogic practices and future MOOC design. This section suggested that MOOC learners might internalise the outputs from algorithms, and adopt “calculated” roles and learning practices in their educational activity. Significant here is the interest, evident in the field of learning analytics, in predicting future behaviour and forecasting learner success. As Williamson suggests:

algorithms are not only social inventions capable of reinforcing existing forms of social order and organization, but have a powerfully productive part to play in predicting and even pre-empting future events, actions, and realities (Williamson 2014).

Prediction must be recognised as a crucial part of the entanglement of algorithms in learner practices, establishing participant roles which subsequently frame future MOOC activity and engagement. MOOC teachers need to develop more awareness of the kind of calculations that algorithms are making behind the slick interfaces of course dashboards. Simply responding to reported data runs the risk of making crucial pedagogical decisions without understanding the rules that have been coded into the dashboard systems, and thus their relation to individual student contexts. There is also a need for MOOC organisations to work more collaboratively with teachers and educators, not only to share the inner workings of algorithmic processes within their software, but also to respond to educational perspectives concerning what kind of data should be used in pedagogical decision-making.

Finally, in outlining a condition of “learning *with* algorithms”, this article suggested that current assumptions about learning in the MOOC – a constructivist or “connectivist” form in cMOOCs, and a behaviourist form in xMOOCs – are inadequate. Theories of learning in the MOOC must account for the role of algorithms in constructing social and communicative roles, as well as learning behaviours. Educational research could draw influence from work in software studies and critical algorithm studies, which has highlighted the broader implications of an algorithmically infused culture, involving “the enfolding of human thought, conduct, organization and expression into the logic of big data and large-scale computation” (Striphas 2015, p. 396). There are significant implications for a “computational turn” in education, and continued research needs to develop a critical discourse around the use of algorithms, particularly in the high-profile domain of the MOOC.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adams, S. (2012). Is Coursera the beginning of the end for traditional higher education? *Forbes* (blog post 17 July). Retrieved January 5, 2018, from <https://www.forbes.com/sites/susanadams/2012/07/17/is-coursera-the-beginning-of-the-end-for-traditional-higher-education/#6cc70b2e12ae>.
- Anderson, T., & Dron, J. (2011). Three generations of distance education pedagogy. *International Review of Research in Open and Distance Learning*, 12(3), 80–97.
- Bayne, S., & Ross, J. (2014). *The pedagogy of the massive open online course: The UK view*. HEA report. Hestlington: The Higher Education Academy (HEA). Retrieved January 15, 2018, from https://www.heacademy.ac.uk/system/files/hea_edinburgh_mooc_web_240314_1.pdf.
- Beer, D. (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media and Society*, 11(6), 985–1002.
- Boellstorff, T., & Maurer, B. (Eds.). (2015). *Data, now bigger and better!*. Chicago: Prickly Paradigm Press.
- Bogot, I. (2015). The cathedral of computation. *The Atlantic*, 15 January. Retrieved January 5, 2018, from <http://www.theatlantic.com/technology/archive/2015/01/the-cathedral-of-computation/384300>.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication and Society*, 15(5), 662–679.
- Breslow, L. (2016). MOOC research: Some of what we know and avenues for the future. In E. De Corte, L. Engwall, & U. Teichler (Eds.), *From books to MOOCs? Emerging models of learning and teaching in higher education*. London: Portland Press.
- Breslow, L., Pritchard, D. E., DeBoer, J., Stump, G. S., Ho, A. D., & Seaton, D. T. (2013). Studying learning in the worldwide classroom: Research into edX’s first MOOC. *Research and Practice in Assessment*, 8(2), 13–25. Retrieved January 5, 2018, from <http://www.rpajournal.com/dev/wp-content/uploads/2013/05/SF2.pdf>.
- Coursera. (2014). Bringing data to teaching. *Coursera building* (blog post 19 November). Retrieved January 5, 2018, from <https://building.coursera.org/blog/2014/11/19/bringing-data-to-teaching/>.
- Daniel, S. J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of Interactive Media in Education*, Perspective (3), Article 18. <http://doi.org/10.5334/2012-18>.

- Ebben, M., & Murphy, J. S. (2014). Unpacking MOOC scholarly discourse: A review of nascent MOOC scholarship. *Learning, Media and Technology*, 39(3), 328–345.
- Ferguson, R., & Clow, D. (2015). Examining engagement: Analysing learner subpopulations in massive open online courses (MOOCs). In *Proceedings of the 5th international learning analytics and knowledge conference (LAK'15)*, March 16–20, 2015, Poughkeepsie, NY (pp. 51–58). Poughkeepsie, NY: Association for Computing Machinery (ACM).
- Friedman, T. (2013). Revolution hits the universities. *New York Times*, 27 January. Retrieved February 5, 2013, from <http://www.nytimes.com/2013/01/27/opinion/sunday/friedman-revolution-hits-the-universities.html>.
- Gillespie, T. (2011). Can an algorithm be wrong? Twitter Trends, the specter of censorship, and our faith in the algorithms around us. *Culture Digitally* (blog post 19 October). Retrieved June 3, 2016, from <http://culturedigitally.org/2011/10/can-an-algorithm-be-wrong/>.
- Gillespie, T. (2014). The relevance of algorithms. In T. Gillespie, P. Boczkowski, & K. Foot (Eds.), *Media technologies* (pp. 167–194). Cambridge, MA: MIT Press.
- Hayles, N. K. (2006). Unfinished work: From cyborg to cognisphere. *Theory, Culture and Society*, 23(7–8), 159–166.
- Hecking, T., Chounta, I. A., & Hoppe, H. U. (2016). Investigating social and semantic user roles in MOOC discussion forums. In *Proceedings of the 5th international learning analytics and knowledge conference (LAK'15)* (pp. 198–207). Poughkeepsie, NY: Association for Computing Machinery (ACM).
- Hildebrandt, M. (2016). Learning as a machine. In Keynote presented on April 27 at the 6th international learning analytics and knowledge conference (LAK'16), April 25–29, 2016, Edinburgh.
- Hildebrandt, M. (2017). Learning as a machine: Crossovers between humans and machines (an updated version of the 2016 Keynote). *Journal of Learning Analytics*, 4(1), 6–23. Retrieved January 5, 2017, from https://works.bepress.com/mireille_hildebrandt/.
- Johnson, L., Adams Becker, S., Cummins, M., Estrada, V., Freeman, A., & Hall, C. (2016). *NMC horizon report: 2016 Higher education edition*. Austin, TX: The New Media Consortium (NMC). Retrieved January 5, 2018, from <http://cdn.nmc.org/media/2016-nmc-horizon-report-he-EN.pdf>.
- Kennedy, G., Coffrin, C., & de Barba, P. (2015). Predicting success: How learners' prior knowledge, skills and activities predict MOOC performance. In *Proceedings of the 5th international learning analytics and knowledge conference (LAK'15)* (pp. 136–140). Poughkeepsie, NY: Association for Computing Machinery (ACM).
- Kitchin, R., & Dodge, M. (2011). *Code/space: Software and everyday life*. Cambridge, MA: MIT Press.
- Kizilcec, R. F., Piech, C., & Schneider, E. (2013). Deconstructing disengagement: Analyzing learner subpopulations in massive open online courses. In *Proceedings of the 3rd international learning analytics and knowledge conference (LAK'13)*, April 8–12, 2013, Leuven, Belgium (pp. 170–179). Poughkeepsie, NY: Association for Computing Machinery (ACM).
- Knox, J. (2014). Active algorithms: Sociomaterial spaces in the e-learning and digital cultures MOOC. *Campus Virtuales*, 3(1), 42–55.
- Knox, J. (2015). Massive open online courses (MOOCs). In M. A. Peters (Ed.), *Encyclopaedia of educational theory and philosophy* (online resource). Singapore: Springer. Retrieved January 5, 2018, from http://link.springer.com/referenceworkentry/10.1007/978-981-287-532-7_219-1.
- Knox, J. (2016). *Posthumanism and the massive open online course: Contaminating the subject of global education*. Abingdon: Routledge.
- McKay, R. F. (2013). Learning analytics at Stanford takes a leap forward with MOOCs. *Stanford Online*, 11 April. Retrieved March 3, 2016, from <http://online.stanford.edu/news/2013/04/11/learning-analytics-stanford-takes-huge-leap-forward-moocs>.
- Middlehurst, R. (2016). Looking to the edges: Future perspectives. In E. De Corte, L. Engwall & U. Teichler (Eds.), *From books to MOOCs? Emerging models of learning and teaching in higher education* (pp. 93–102). London: Portland Press. Retrieved January 5, 2018, from http://www.portlandpresspublishing.com/sites/default/files/Editorial/Wenner/PPL_Wenner_Ch08.pdf.
- Morozov, E. (2013). *To save everything click here: Technology, solutionism and the urge to fix problems that don't exist*. London: Penguin.
- Online Course Report, OCR. (2016). *State of the MOOC 2016: A year of massive landscape change for massive open online courses* (online resource). Retrieved January 5, 2018, from <http://www.onlinecourereport.com/state-of-the-mooc-2016-a-year-of-massive-landscape-change-for-massive-online-courses/>.

- Perna, L., Ruby, A., Boruch, R., Wang, N., Scull, J., Evans, C., & Ahmad, S. (2013). The life cycle of a million MOOC users. In *Presented at the MOOC research initiative conference*, December 5–6, 2013, University of Texas at Arlington. Retrieved January 5, 2018, from http://www.gse.upenn.edu/pdf/ahead/perna_ruby_boruch_mooocs_dec2013.pdf.
- Perrotta, C. & Williamson, B. (2016). The social life of Learning Analytics: Cluster analysis and the “performance” of algorithmic education. *Learning, Media and Technology*, 17 May (online article). Retrieved January 5, 2018, from <http://www.tandfonline.com/doi/full/10.1080/17439884.2016.1182927>.
- Poquet, O., & Dawson, S. (2016). Untangling MOOC learner networks. In *Proceedings of the 6th international learning analytics and knowledge conference (LAK'16)* (pp. 208–212). Poughkeepsie, NY: Association for Computing Machinery (ACM).
- Rodriguez, O. (2013). The concept of openness behind c and x-MOOCs (Massive Open Online Courses). *Open Praxis*, 5(1), 67–73.
- Siemens, G. (2005). Connectivism: A learning theory for the digital age. *International Journal of Instructional Technology and Distance Learning*, 2(1), 3–10. Retrieved January 22, 2018, from http://www.itdl.org/Journal/Jan_05/article01.htm.
- Striphas, T. (2015). Algorithmic culture. *European Journal of Cultural Studies*, 18(4–5), 395–412. Retrieved January 5, 2018, from <http://journals.sagepub.com/doi/pdf/10.1177/1367549415577392>.
- Weller, M. (2015). MOOCs and the Silicon Valley narrative. *Journal of Interactive Media in Education* (1), Article 5. <http://dx.doi.org/10.5334/jime.am>.
- Williamson, B. (2014). Organizing algorithms, calculated publics in digitally-mediated education. *DML central* (blog post 4 August). Retrieved January 5, 2018, from <http://dmlcentral.net/organizing-algorithms-calculated-publics-in-digitally-mediated-education/>.
- Williamson, B. (Ed.). (2015a). *Coding/learning: Software and digital data in education*. Stirling: University of Stirling.
- Williamson, B. (2015b). Learning in the digital microlaboratory of educational data science. *DML central* (blog post 30 November). Retrieved January 5, 2018, from <http://dmlcentral.net/learning-in-the-digital-microlaboratory-of-educational-data-science/>.
- Ziewitz, M. (2015). Governing algorithms: Myth, mess, and methods. *Science, Technology and Human Values*, 41(1), 3–16.

The author

Jeremy Knox is a Lecturer in Digital Education at the University of Edinburgh, and a Core Member of the Centre for Research in Digital Education (<http://www.de.ed.ac.uk/>). His research interests include critical posthumanism, data studies and algorithm studies, and the implications of such thinking for education and educational research, with a specific focus on the digital. Jeremy has published numerous critical perspectives on open education, and is the author of the recent book *Posthumanism and the Massive Open Online Course: Contaminating the Subject of Global Education* (Routledge 2016; <https://www.routledge.com/Posthumanism-and-the-Massive-Open-Online-Course-Contaminating-the-Subject/Knox/p/book/9781138940826>).