A New Algorithmic Identity
Soft Biopolitics and the Modulation of Control

John Cheney-Lippold

Abstract
Marketing and web analytic companies have implemented sophisticated algorithms to observe, analyze, and identify users through large surveillance networks online. These computer algorithms have the capacity to infer categories of identity upon users based largely on their web-surfing habits. In this article I will first discuss the conceptual and theoretical work around code, outlining its use in an analysis of online categorization practices. The article will then approach the function of code at the level of the category, arguing that an analysis of coded computer algorithms enables a supplement to Foucauldian thinking around biopolitics and biopower, of what I call soft biopower and soft biopolitics. These new conceptual devices allow us to better understand the workings of biopower at the level of the category, of using computer code, statistics and surveillance to construct categories within populations according to users' surveilled internet history. Finally, the article will think through the nuanced ways that algorithmic inference works as a mode of control, of processes of identification that structure and regulate our lives online within the context of online marketing and algorithmic categorization.

Key words
algorithm ■ biopolitics ■ biopower ■ code ■ control ■ Deleuze ■ Foucault ■ internet

■ *Theory, Culture & Society* 2011 (SAGE, Los Angeles, London, New Delhi, and Singapore), Vol. 28(6): 164–181
DOI: 10.1177/0263276411424420

Introduction

LET ME begin this article with a hypothetical. You open up a new computer and fire up a web browser. You go to the washingtonpost.com, visit a couple of blogs at Wordpress and Tumblr, and go on the business social networking site linkedin.com. Maybe you take a break from the internet, go grab a cup of coffee, but return to watch some videos on Hulu, check US gossip at tmz.com, and look at the weather at wunderground.com. At this point you decide it might be best to go to work so you close your computer, get dressed, and go outside. While you may proceed with your day as if nothing has happened, something has changed about who you are online. You have been identified. Your IP address has been logged; you have a cookie file installed on your computer. And somewhere, in a database far, far away, you very well may have a gender, class, and race.

This small tour of web sites provides us with a mundane example of what I call in the title of this article a 'new algorithmic identity'. The networked infrastructure of the internet, with its technological capacity to track user movements across different web sites and servers, has given rise to an industry of web analytics firms that are actively amassing information on individuals and fine-tuning computer algorithms to make sense of that data. The product of many of these firms is a 'new algorithmic identity', an identity formation that works through mathematical algorithms to infer categories of identity on otherwise anonymous beings. It uses statistical commonality models to determine one's gender, class, or race in an automatic manner at the same time as it defines the actual meaning of gender, class, or race themselves. Ultimately, it moves the practice of identification into an entirely digital, and thus measureable, plane.

This article will examine the consequence of some of those practices aimed at understanding exactly what kind of user is visiting web sites, purchasing products, or consuming media. Online a category like gender is not determined by one's genitalia or even physical appearance. Nor is it entirely self-selected. Rather, categories of identity are being inferred upon individuals based on their web use. Code and algorithm are the engines behind such inference and are the axis from which I will think through this new construction of identity and category online. We are entering an online world where our identifications are largely made for us. A 'new algorithmic identity' is situated at a distance from traditional liberal politics, removed from civil discourse via the proprietary nature of many algorithms while simultaneously enjoying an unprecedented ubiquity in its reach to surveil and record data about users.

In this article I will first discuss the conceptual and theoretical work around code, outlining its use in an analysis of online categorization practices. The article will then approach the function of code at the level of the category, arguing that an analysis of coded computer algorithms enables a supplement to Foucauldian thinking around biopolitics and biopower, of what I call soft biopower and soft biopolitics. These new conceptual devices

will allow us to better understand the workings of biopower at the level of the category, of using computer code, statistics and surveillance to construct categories within populations according to users' surveilled internet history. Finally, the article will think through the nuanced ways that algorithmic inference works as a mode of control, of processes of identification that structure and regulate our lives online within the context of online marketing and algorithmic categorization.

Code and Categorization

Even spared of its technological complexity, code is still a tricky subject to wrap one's head around. To begin with, code as a concept is used quite indiscriminately. Some theorists see code as an architecture, of the digital walls and foundations that delimit our experience online. This thinking is centered around the work of Lawrence Lessig, for whom code constructs the space we have to use and access information. Code, in Lessig's mind, 'is law' (Lessig, 2006: 1). Rules are both intentionally and unintentionally written into the hardware and software of internet technologies that then create the infrastructure that determines how we as users can act in cyberspace.

But most contemporary uses of code as a concept require a more dynamic definition. Nigel Thrift (2005: 173) offers code not as law but 'new adaptive standards of conduct which can work in all kinds of situations at all kinds of scales'. And Alexander Galloway (2004) argues that this type of code, or his term 'protocol', also encompasses the standards for internet communication (TCP/IP and DNS). The structural ways that protocols enable and disable particular kinds of internet traffic become viable as they ensure standardization, an architectural metric system of sorts through which control follows from 'a set of technical procedures for defining, managing, modulating, and distributing information through a flexible yet robust delivery infrastructure' (Thacker, 2004: xv). Code as architecture works to structure the boundaries, as well as regulate the flows, of internet traffic.

A key part of this regulation is in the actual work that code does, of creating, ordering and ultimately giving meaning to digital artifacts. Code as a concept must also force us to think about the mundane processes of data differentiation online, specifically of determining how a variable like X or Y can come to be defined, what X and Y actually mean and, most importantly, how users experience X and Y within the architecture that code creates. Stephen Graham echoes this sentiment with his work on 'software-sorting', of 'directly, automatically and continuously allocating social or geographical access to all sorts of critical goods, services, life chances or mobility opportunities to certain social groups or geographical areas' (2005: 3). Code is part of a dynamic relationship to the real world, one that can 'automatically and continuously' affect life chances offered to users based on a pre-configured but also reflexive programmed logic. But rather than look at code as just an automated and adaptive gatekeeper of critical

information, we can explore its function much more constitutionally. An analysis that centers on code allows us to look at a list of lines of a computer language and see how it uses certain representations of the world (variables and data) to produce new value. A new value like X = male can then be used to suggest sorting exercises, like targeted content and advertisements, based entirely on assumptions around that value. Yet code can also construct meaning. While it can name X as male, it can also develop what 'male' may come to be defined as online.

But this new value is not corralled within an entirely digital realm. When a user changes her profile on Facebook to state her 'sex', there's both a delimiting action (of the set drop-down menu choices available to describe one's 'sex' of male or female) as well as a cultural discourse (what does being female online mean?) through which code speaks (Fuller, 2008). The nascent field of software studies looks at cultural practices that exceed the material world of code, of incorporating an analysis that takes into account human-machine interactions as well as the technology of machines themselves (Cramer, 2005). The study of code and software in general tries to go 'beyond the blip' to understand the implicit politics of computer code, 'to make visible the dynamics, structures, regimes, and drives' of the wide variety of programmed scripts that are littered across the internet (Fuller, 2003: 32). With this view we can see computer code not as a literal, rhetorical text from which we derive meaning but a complex set of relationships that tie together the coded systems of definition and organization that constitute our experience online. Codes are cultural objects embedded and integrated within a social system whose logic, rules, and explicit functioning work to determine the new conditions of possibilities of users' lives. How a variable like X comes to be defined, then, is not the result of objective fact but is rather a technologically-mediated and culturally-situated consequence of statistics and computer science.

Categorization

My argument will focus on the process of defining X, of the identification of particular groups within populations largely through the marketing logic of consumption. Within marketing, considerable research goes to identify the composition of a consumer audience. In the past, consumers were discriminated based on census-data laden geographies: rich people lived in a certain ZIP code, poor people lived in another, and businesses could provide coupons for class-differentiated products and services accordingly (Gandy, 1993). A move from demographic to psychographic categorizations followed, where clusters of consumer types were created to better understand the non-essentialist character of demographic-based consumption patterns (not all white people are interested in timeshares in Boca Raton, but those who are older and with money are) (Yankelovich and Meer, 2006; Arviddson, 2006). An important shift in marketing happened as marketers went online and were able to use data from search queries to create behavioral

understandings atop these clusters (Turrow, 2006). Purchase and other behavioral data were then lumped into how marketers see and interact with individuals through the construction of 'databases of intentions', where user information and search queries are aggregated to understand intentions – such as purchase intent – as well as general trends in social wants and needs (Battelle, 2005). Mathematical algorithms allowed marketers to make sense out of these data to better understand how to more effectively target services, advertisements, and content. Through algorithms, commonalities between data can be parsed and patterns within data then identified and labeled. And with the capacities of computers to surveil and capture user activity across the internet, consumer information was removed from the shackles of time-bound, decadal census data and began to float atop a constant stream of real-time web use that can be matched against existing behavior and identity models – like gender.

Such a shift focuses not on essential notions of identity but instead on pliable behavioral models, undergirded by algorithms, that allow for the creation of a cybernetic relationship to identification. New worlds online (of content or advertisements) are created according to newly identified user data and interactions. Continually updated new content can then be suggested to users based on the history of a user's interactions with the system, which are compiled into a particular identity for that user (like X = male). As the capacity of computers to aggregate user data increases and algorithms are improved upon to make disparate data more intelligible and useful, the ability for real-time cybernetic modeling to monitor, compute, and act becomes more efficient. So as more data is received about a certain user's behavior online, new coded computations can be done to change who the user is believed to be and what content that user might desire (the gender of the same user might change from male to female if enough user data, such as the addition of certain web sites that user visited, are presented to statistically identify that user with a different gender). In this constant feedback loop we encounter a form of control. But this is not a control that we can explain through references to code as law. It does not present us with static norms of conduct or of unsurpassable rules that prevent particular uses. Code also regulates the innards of the internet. It plays a constitutive role in the mundane activities – of what book I'm recommended at a particular moment or what kind of shampoo I'm advertised when I'm visiting a particular web page. This thinking falls into a variety of models of cybernetics, from Vincent Mosco's (1996) cybernetic commodity that details information about consumer transactions and creates a database of user behavior to Nikolas Rose's (1999) cybernetics of control that explores how normative assumptions of daily life become embedded within the circuits of our education, employment, and consumption practices.

But I argue we can even more effectively understand this process as modulation, of a Deleuzeian approach to control that relies on practices characterized by the terms of the societies of control (Deleuze, 1992). This move offers us a way of seeing how many mechanisms of control

have shifted to practices external to the body, where discipline has often become more or less unnecessary if control can be enacted through a series of guiding, determining, and persuasive mechanisms of power. Instead of constructing subjectivity exclusively through disciplinary power, with normative discourses prefiguring how we engage with and even talk about the world, a regulative approach allows for a distance between power and subject (Foucault, 2008). It configures life by tailoring its conditions of possibility. Regulation predicts our lives as users by tethering the potential for alternative futures to our previous actions as users based on consumption and research for consumption. Control in this formation becomes best represented by the concept of modulation, 'a self-deforming cast that will continuously change from one moment to the other, or like a sieve whose mesh will transmute from point to point' that sees perpetual training replacing serial encounters with singular, static institutions (Deleuze, 1992: 4). And modulation marks a continuous control over society that speaks to individuals in a sort of coded language, of creating not individuals but endlessly sub-dividable 'dividuals' (Deleuze, 1992). These dividuals become the axiom of control, the recipients through which power flows as subjectivity takes a deconstructed dive into the digital era. When we situate this process within questions around digital identity, dividuals can be seen as those data that are aggregated to form unified subjects, of connecting dividual parts through arbitrary closures at the moment of the compilation of a computer program or at the result of a database query.

Dividual pieces, onto which I conceptually map the raw data obtained by internet marketing surveillance networks, are made intelligible and thus constitute the digital subject through code and computer algorithms. The algorithmic process that I will focus on in this article looks at the web analytics firm Quantcast as it infers certain categorical identifications from seemingly meaningless web data.[1] As new information gets created about an individual through the tracking of her web presence, a user's identity becomes more defined within the system in accord with the logic of Quantcast's identification algorithms. If I would surf on Quantcast's member sites for several hours, my initially unknown gender would become more concrete in the face of Quantcast's algorithms, as exemplified in the hypothetical that begins this article. Dividual fragments flow across seemingly open and frictionless networks and into rigid database fields as part of the subsumption implicit in data mining (the practice of finding patterns within the chaos of raw data). As a user travels across these networks, algorithms can topologically striate her surfing data, allocating certain web artifacts into particular, algorithmically-defined categories like gender. The fact that user X visits the web site CNN.com might suggest that X could be categorized as male. And additional data could then buttress or resignify how X is categorized. As X visits more sites like CNN.com, X's maleness is statistically reinforced, adding confidence to the measure that X may be male. As X visits more sites that are unlike CNN.com, X's maleness

might be put into question or potentially resignified to another gender identity.

In this example, web data's apparent chaos and undefinability, of the limitless possibilities that any data set initially offers, is allocated into specific categories. The implicit disorder of data collected about an individual is organized, defined, and made valuable by algorithmically assigning meaning to user behavior – and in turn limiting the potential excess of meanings that raw data offer. By using these dividual fragments, a subject's identity can be articulated according to the programmed rationale of the algorithm. Ultimately, and after certain statistical checks of confidence, this articulation can result in a unified and static status: X is male. As described in a Quantcast white paper outlining the company's inference methodology, these static user identities help determine the gendered make-up of a site (Quantcast, 2008). By tracking users and algorithmically assigning categories according to their behavior, Quantcast may define the gender composition of a site on its network as 73 percent male and 27 percent female. But as more inferred-as-female users begin to visit the site, the site's gender identity has the chance to change in accord (for example to 72 percent male and 28 percent female). The indirect consequence of this dynamism comes as a site's altered gendered make-up then can play a constitutive part in reconfiguring the statistical groupings that Quantcast uses to infer gender upon its users. New, not-yet-gendered individuals that visit this site may still be identified as male but now with a statistical confidence potentially less than previously calculated.[2]

In requisite cybernetic fashion a user's ascribed gender can and may change as new user information arrives into the cybernetic system.This type of user identity does not approach identity formation from an essentialist framework. Rather, algorithms allow a shift to a more flexible and functional definition of the category, one that de-essentializes gender from its corporeal and societal forms and determinations while it also re-essentializes gender as a statistically-related, largely market research-driven category. Gender becomes a vector, a completely digital and math-based association that defines the meaning of maleness, femaleness, or whatever other gender (or category) a marketer requires. These categorizations are very much tied to the groupings of what a company like Quantcast finds useful, namely those that are derived from the world of market research (gender, age, income). And while offline data can be used to supplement algorithmic inference measures, the move toward a vector-based category marks a separation of the category from its normative domain and a reassociation of that category according to statistical correlation. Of course real world preconceptions around the meanings of categories feed into the foundational definitions that we have about certain genders, races, and classes online (Nakamura, 2002). In most cases, predetermined biases about the nature of group identity will be maintained online. But the capacity for cybernetic categorization to regulate certain categories' meaning according

to algorithm marks a move away from offline stereotypes and into a form of statistical stereotyping.

Rather than maintaining a particular and concrete relationship to maleness in the offline world, a cybernetic conception of gender would position maleness as a modulating category, based on a vector and composed of statistical assessments that define male group membership. New information can come to light within this algorithmic system and an algorithm can subsequently move to change both a category's meaning and the once-essential ideals ascribed to that category. Gender as a category still holds the capacity for discipline – of serving targeted advertisements and targeted content according to the inferred digital identity of a user – but it also is embodied by a new and flexible cybernetic categorization system. Instead of standards of maleness defining and disciplining bodies according to an ideal type of maleness, standards of maleness can be suggested to users based on one's presumed digital identity, from which the success of identification can be measured according to ad click-through rates, page views, and other assorted feedback mechanisms. The regulation of gender as a category then becomes wholly embedded within the logic of consumption, where categorical behaviors are statistically defined through a cybernetics of purchasing and research that marketers have deemed valuable for identification and categorization.

As the specifics of Quantcast's algorithms remain proprietary, we can use other examples of 'machine learning' to help us explicate this process in the public domain. Put simply, machine learning is the 'technical basis . . . used to extract information from the raw data in databases – information that is [then] expressed in a comprehensible form and can be used for a variety of purposes' such as prediction and inference (Witten and Frank, 2005: xxiii). Machines are programmed to learn patterns in data and use those correlative patterns to analyze and make assessments about new data. Yahoo researchers, for example, have described the capacity for an algorithm to infer demographic categories of identity (race, gender). Based on search query logs, user profile data, and US census data, these researchers found they can subsequently tailor results according to user categorizations. If a particular user issues the query 'wagner', an algorithm can offer differentiated results as determined by that user's gender categorization. Based on the observed web habits of 'typical' women and men, search results for 'wagner' can include sites about the composer 'Richard Wagner' for women while at the same time provide sites on 'Wagner USA' paint supplies for men (Weber and Castillo, 2010). Similarly, researchers in Israel have created an algorithm that predicts an author's gender based on lexical and syntactic features. Upon separating the written works of women and men, statistical correlations were run to create a vector-based categorization of two gender's written features, through which the algorithm was able to then identify the 'correct' gender of a new, non-categorized work with a success rate of 80 percent (Koppel et al., 2002). But undergirding all of this is the theoretical possibility that these statistical models can change

given new data. The chance that men, in the Yahoo example, stop clicking on paint supplies and start clicking on music sites could statistically suggest that Yahoo's definition of 'men' may not be appropriate. And the possibility that an inferred-as-male writer may not write like every man before him can make an algorithm statistically second guess its initial vector-based categorization. Categories always have the capacity for change, off and online, but the shift I want to bring to the world of identity is the continuous, data-centered manner that modulates both user experience (what content they see) as well as the categorical foundation of those identifications.

Soft Biopower and Soft Biopolitics

I argue that with cybernetic categorization comes a new analytical axis of power: the digital construction of categories of identity. A recent call to rethink biopolitics, to understand the historical context from which contemporary forms of control over populations interact, does well to push us away from more modernist conceptions of biopower and biopolitics and into a realm of contemporary, localized biopolitical analysis (Revel, 2009). My own analysis will focus on biopower as it works in the construction of identity, of the 'power over life' through the management of subjects at the level of the population (Foucault, 2003a: 135). The regulatory work of biopolitical technologies does much to understand the shift Foucault articulated from power of sovereignty to power over life (Foucault, 2008). And it allows us to assess the impact of somewhat ethically-ambiguous technologies of biopower. In what Foucault calls security mechanisms we can see the technologies installed to regulate the inherent randomness in any population of human beings so as to optimize a particular state of life (Foucault, 2003a: 246). I address these mechanisms through computer algorithm-based categorization that work to biopolitically manage and control groups of people.

Biopolitics' origin in 18th-century society worked to supplement existing modes of disciplinary power, of introducing forms of regulatory power that found a new type of control of populations at the level of the man-as-species (Foucault, 2003a). Using statistics, demographic assessments, and through an analysis of birth and death rates, government was able to situate itself in a relationship with subjects not only vis-à-vis individual bodies but vis-à-vis the population and sub-populations. Statistics, forecasting, and measurement became the preferred modes of regulation, of governmental intervention at a level of 'generality' – the level where parts of life are determined and through which a temporary homeostasis could be achieved (Foucault 2007, 2008). Biopolitics in traditional Foucauldian terms has been centered in the practices of achieving this equilibrium. I situate my intervention with the different processes of categorization that allow us to come to this equilibrium through 'a technology which brings together the mass effects of characteristic of a population, which tries to control the series of random events that can occur in a living mass, a technology

which tries to predict the probability of those events (by modifying it, if necessary) or at least compensate for their effects' (Foucault, 2003a: 249).

With the construction of categories we can see how biopolitical work controls those random events, of creating what Foucault calls 'caesuras', or breaks in the biological continuum of characteristics of life (Foucault, 2003a; Macey, 2009). Biopower works through a series of dividing practices, a process of objectification outlined in Louise Amoore's (2006: 339) analysis of border technology biometrics that act on populations through the categorizations of 'student', 'Muslim', 'woman', 'alien', 'immigrant', and 'illegal'. But the move I want to make is to analyze the role of what I term soft biopolitics, of how biopolitical processes give meaning to the categorizations that make up populations and through which we can look at the variable indirect processes of regulation. To explicate soft biopolitics is to better understand not just how populations are managed but the changing relationship that categories have to populations and, most importantly, what categories actually come to mean and be redefined as.

We can also take a grounded example with a biopolitical tinge to explore this process. Say you are categorized as a male through your use-patterns online. Medical services and health-related advertisements could be served to you based on that gendered categorization. This means that those who are not categorized as male may not experience the same advertisement and opportunities, enacting the consequences of what was already referred to as 'software-sorting' (Graham, 2005). But while the initial biopolitical relationship of subject and power would stop at this point, the introduction of the notion of cybernetic categorization can provide another perspective of Foucauldian power. Users are not categorized according to one-off census survey data but through a process of continual interaction with, and modification of, the categories through which biopolitics works. In marketing, these categorizations are constructed within the logic of consumer capitalism, where biopower can govern subjects at a distance, guarding their apparent autonomy while '[optimizing] systems of difference, in which the field is left open to fluctuating processes... [and] in which there is an environmental type of intervention instead of the internal subjugation of individuals' (Foucault, 2008: 259–60). Category construction in itself has become a biopolitical endeavor, where Foucauldian security dilemmas are resolved not just by enactment or change of policy but also through an alteration in how existing policy reaches subjects. Maleness can be constantly evaluated according to feedback data, through which the definitions of maleness can shift (to include or exclude transgendered folk, for just one example) according to the logic of the algorithm. Policies that rely on gendered categorizations then can be automatically and continuously reoriented to address new populations.

These algorithmic definitions then supplement existing discursive understandings of a category like gender – most fundamentally through the language of gender itself – with statistical conclusions of what an idea like maleness actually is. The exclusivity of gender's meaning then becomes

held within the logic of the categorical system as defined by a particular algorithm. The cybernetic definition and redefinition of gender then provides a form of elasticity to power. At this categorical level there exists a theoretical modulation that restages the relationship that biopower has to subjects. In our above example, medical advertisements targeted to men represent such power on two fronts. First, traditional biopolitical action of medical services helps manage life. The state or corporate power can biopolitically aim to reduce STD transfer among the male population by advertising condoms or safe sex behaviors. And second, the normalizing force of what it means to be categorized as male becomes a biopolitical activity. The category of gender in our example is a result of statistical analyses that regulate what Foucault calls the 'conduct of conduct', or more appropriately by discursively positioning maleness to indirectly, and largely unintentionally, control the conditions of possibilities afforded to users who are targeted as male (Foucault, 1982). It is this regulation of the randomness in gendered populations that optimizes through soft biopower a particular state of life as defined by algorithm.

The demands put on soft biopower – of updatable categories that can adapt to the dynamism of populations – require malleable, modulating categorical groupings in order to effectively define and redefine categories' meaning. The truth of what is gendered male requires us to dig beneath its potential self-evidence (scientific statistical probability models) and into a larger schema of power relations, of understanding the 'fundamental codes of a culture – those governing its language, its schemas of perception, its exchanges, its techniques, its values . . . [that] establish for every man. . . the empirical orders with which he will be dealing and within which he will be at home' (Foucault, 1973: xx). But Foucault denotes a middle space, between these coded limitations of our culture, on one hand, and philosophical and scientific interpretations of order on the other (for example, God created and classified the world). This middle space is where soft biopower lives, where the modulation of categorical identity changes according to a 'culture, imperceptibly deviating from the empirical orders prescribed for it by its primary codes . . . [that] frees itself sufficiently to discover that these orders are perhaps not the only possible ones or the best ones' (Foucault, 1973: xx).

By stepping through the concept of cybernetic categorization we can follow this important move etched by Foucault in his understanding of the social construction of categories. First, ontologies are embedded within a set of power relations. Second, the categories that are bred from those ontologies exercise a profound impact on how we as subjects encounter our world. And third, changes in our categorizing schematics are indebted to this fundamental coding of a culture, from which I find a strong parallel to the technological organization of subject identities through code and algorithm. The knowledge formation implicit in algorithmic work 'creates and causes to emerge new objects of knowledge and accumulates new bodies of information' that ultimately give rise to new categorization practices, in the

case of the shift from offline to online marketing, but also redefined categorical meanings in the case of cybernetic categorization (Foucault, 1980: 52). This redefinition is part of what others have argued is a shift toward a 'topological' approach to genealogy, one that identifies 'patterns of correlations' that lead to the formation of particular dispositions of unified heterogeneous elements (Collier, 2009). While Collier's treatment of a 'topology of power' takes these elements as a wide array of different and recombined techniques and technologies to create a unified state of power relations, the same logic can be applied to algorithm. Patterns of correlation can be found in technologies of algorithmic categorization, of recombining and unifying heterogeneous elements of data that have no inner necessity or coherence.

The softer versions of biopower and biopolitics supplement the discursive production of categories' meanings, as it is also through data and statistical analysis that conceptions of gender change – not just through discourse and its subsequent naturalization (Foucault, 1973).[3] In order to understand the regulatory process of biopolitics we need to place additional attention on the biopolitical construction of these categorizations. Here we can better define soft biopower and soft biopolitics. The former signifies the changing nature of categories that on their own regulate and manage life. Unlike conceptions of hard biopower that regulate life through the use of categorizations, soft biopower regulates how those categories themselves are determined to define life. And if we describe biopolitics as Foucault does, as 'the endeavor . . . to rationalize the problems presented to governmental practice by the phenomena characteristic of a group of living human beings constituted as a population', soft biopolitics constitutes the ways that biopower defines what a population is and determines how that population is discursively situated and developed (Foucault, 2003b: 73).

Control

I argue these defining practices of soft biopower and soft biopolitics act as mechanisms of regulatory control. For the remainder of this article I will understand control as 'operating through conditional access to circuits of consumption and civility', interpreting control's mark on subjects as a guiding mechanism that opens and closes particular conditions of possibility that users can encounter (Rose, 2000: 326). The feedback mechanism required in this guiding mechanism is the process of suggestion. I define suggestion as the opening (and consequent closing) of conditional access as determined by how the user is categorized online.[4]

As categorizations get constructed by firms like Quantcast, refined by algorithms that process online behavior, and continually improved upon as more and more time passes and more and more web artifacts are inputted into a database, advertisements and content are then suggested to users according to their perceived identities. In cybernetic categorization these groupings are always changeable, following the user and suggesting new

artifacts to be visited, seen, or consumed. Importantly, the cybernetic feedback loop exists only in so far as it productively aids in the maintenance of the system. The process of suggestion works to the extent that algorithmic inference adeptly categorizes.[5] If a particular suggestion is misguided, algorithms are programmed to interpret a mis-categorization and reassign a category of identity based on that user's newly observed behavior.

Historically thinking around control in the field of surveillance has focused on a direct relationship to the subject. David Lyon describes surveillance as 'the routine ways [that] focused attention is paid to personal details by organizations that want to influence, manage, or control certain persons or population groups' (Lyon, 2003a: 5). Lyon's explicit mention of control in a longer list including 'influence' and 'manage' presents a largely direct and causal description of control. Organizations surveil because they 'want to' determine discriminatory pricing (commercial surveillance), discourage crime (CCTV), or apprehend criminals (NSA surveillance programs).[6] I propose we approach Lyon's definition from a more subtle and indirect relationship of control. Through what I term cybernetic categorization, categories' meaning can be realigned according to the code and algorithmic models built to target content to particular consumers. The process of identification, at least in the online world, becomes mediated by what I term soft biopolitics, as user identities become tethered to a set of movable, statistically-defined categorizations that then can have influence in biopolitical decisions by states and corporations. This control resembles Nikolas Rose's (1999: 49–50) concept of 'government at a distance', of regulating how identity can be governed 'through the decisions and endeavours of non-political modes of authority [that are] distanced spatially, in that these technologies of government link a multitude of experts in distant sites to the calculation of those at a center'. Control then works at levels far past the purview of liberal individualism, situating subjects within networks of power that govern indirectly and without proximity. The individual user is incapable of really experiencing the effect that algorithms have in determining one's life as algorithms rarely, if ever, speak to the individual. Rather, individuals are seen by algorithm and surveillance networks as members of categories. So instead of positioning the individual as the locus of research, the addition of the level of the category can enable us to better understand the structuring of our lives under surveillance as well as the myriad ways control can work on both individuals and categories. And it is here that the potential for discourse around identity becomes problematic. The identifications that make us as subjects online are becoming more opaque and buried, away from our individual vantage points and removed from most forms of critical participation. They are increasingly finding mediation outside the realm of traditional political intervention and inside the black boxes of search engines and algorithmic inference systems (Becker and Stalder, 2009).

This system of regulative control makes a break with the theoretical lineage around control and technology that was begot in the world of cyberpunk, as we travel away from the liberal focus on the individual and

toward the structuring forms of suggestion at work when users that are gendered (or with any other categorical identity) visit a web page.[7] Such ubiquity demonstrates a capacity, and the increasing use of that capacity, to employ surveillance technologies to gather data about both users and populations in general. But this ubiquity also needs to take into account an anti-technological deterministic perspective that allows for technological failure. Surveillance technologies do not work as designed all the time, and we have seen with the example of CCTV that technological capacity does not ensure a human capacity to adeptly monitor and analyze that technology.[8] I agree with Wendy Hui Kyong Chun (2006) when she argues that the internet is neither a tool for freedom nor a tool for total control. Control is never complete, and neither is our freedom. At this present moment, web analytic firms have made it nearly impossible to not be incorporated into surveillance networks. We are always vulnerable to some degree of surveillance's gaze. But vulnerability, the historical point of departure when we define freedom in its most classical liberal sense, should then be thought not as the revocation of autonomy or one's suffocation under control (Chun, 2006).

Control has become a modulating exercise, much like Deleuze predicted, which constitutes an integral part of contemporary life as it works not just on the body nor just on the population, but in how we define ourselves and others (Deleuze, 1992). Deleuze is vital for exploring these processes for the very fact that he understood the shift from 'enclosed' environments of disciplinary society to the open terrain of the societies of control. Enclosure offers the idea of walls, of barriers to databases and surveillance technologies. Openness describes a freedom to action that at the same time is also vulnerable to surveillance and manipulation. And these open mechanisms of control, the automated categorization practices and the advertisements and content targeted to those categorizations effectively situate and define how we create and manage our own identities. Surveillance practices have increasingly moved from a set of inflexible disciplinary practices that operate at the level of the individual to the statistical regulation of categorical groupings: 'it is not the personal identity of the embodied individual but rather the actuarial or categorical profile of the collective which is of foremost concern' to new, unenclosed surveillance networks (Hier, 2003: 402).

Cybernetic categorization provides an elastic relationship to power, one that uses the capacity of suggestion to softly persuade users towards models of normalized behavior and identity through the constant redefinition of categories of identity. If a certain set of categories ceases to effectively regulate, another set can quickly be reassigned to a user, providing a seemingly seamless experience online that still exerts a force over who that user is. This force is not entirely benign but is instead something that tells us who we are, what we want, and who we should be. It is removed from traditional mechanisms for resistance and ultimately requires us to conceive of freedom, in whatever form, much more differently than previously

thought. We are effectively losing control in defining who we are online, or more specifically we are losing ownership over the meaning of the categories that constitute our identities. Algorithm ultimately exercises control over us by harnessing these forces through the creation of relationships between real-world surveillance data and machines capable of making statistically relevant inferences about what that data can mean. And the processes of soft biopower work in a similar fashion, allowing for a modularity of meaning that is always productive – in that it constantly creates new information – and always following and surveilling its subjects to ensure that its user data are effective. New cybernetic category constructions are the consequence of this modularity and ultimately allow for a 'free', but constantly conditioned, user.

Notes

1. Quantcast (www.quantcast.com) is a web analytics company that operates as a free service where member sites include HTML snippets of code into each HTML page on a web server. These snippets 'phone home' to Quantcast databases every time a user visits a site. This recording of user visits is then aggregated over time to create a history of a user's web use across various web sites.

2. Quantcast uses more than just web-use data to infer categories upon its users – 'An increasing variety of inputs can be used to continually validate and refine Quantcast inference models' – but the impact and extent of these data inputs are unknown. I will thus focus on web use exclusively (Quantcast Corporation, 2008: 1).

3. Soft as a prefix can be elaborated by linking biopower to two similar concepts from different literatures. In the more proximate field of computer science, soft computing developed as programmers moved from a world of perfect computation and into a fuzzier variety of close-enough, inexact solutions to problems (Zadeh, 1994). This mimics the shift from essential modes of identity online and toward fuzzy conceptions of statistical belief – a user is likely male based on probability. And in a theoretical parallel to diplomatic soft power, soft biopower works as a more flexible form of power that uses attraction and influence to achieve its goals. Where the brute force of biopolitical action (population control) is mediated through definition and redefinition of the targeted category, the brute force of hard state-power (war) is mediated through diplomatic mechanisms and arguments around mutual benefit (Nye, 2002). 'Hard' biopower acts by dividing populations and enacting policy to control subjects through those divisions; soft biopower acts by taking those divisions and modulating categories' meaning so as to best serve the rationale of hard biopower.

4. One example in particular of how one's conditions of possibility may be affected through targeting is what Cass Sunstein (2007) defines as 'The Daily Me', an individualized array of information that pertains to the perceived or provided interests of a user, theoretically decreasing the chance that the user will encounter news or information that may contradict his existing views about the world.

5. 'Adeptly categorizes' does not mean 'correctly categorizes'. The re-essentializing move of this categorization process relies on the fact that a categorization fits the behavior of a user, not that the user himself embodies that category.

6. For work on commercial surveillance see Lyon (2003b); for work on CCTV see Lomell (2004); for work on the NSA see Bamford (2001).

7. This liberal notion is most famously articulated in the work of Gibson (1984) and Dick (1968), both of whom defined cyberpunk's insistence on the individual's direct relationship with power.

8. The extent to which CCTV works is arguable, though analysis by Webster (2009) has offered a strong case that CCTV does little to prevent crime. The number of cameras in London alone, which has elicited a constant cry from privacy advocates, makes human observation of the technology near impossible.

References

Amoore, L. (2006) 'Biometric Borders: Governing Mobilities in the War on Terror', *Political Geography* 25: 336–351.

Arvidsson, A. (2006) *Brands: Meaning and Value in Media Culture*. New York: Routledge.

Bamford, J. (2001) *Body of Secrets*. New York: Random House.

Battelle, J. (2005) *The Search: How Google and its Rivals Rewrote the Rules of Business and Transformed Our Culture*. New York: Portfolio.

Becker, K. (2009) 'The Power of Classification: Culture, Context, Command, Control, Communications, Computing', in K. Becker and F. Stalder (eds) *Deep Search: The Politics of Search Beyond Google*. Munich: Studienverlag & Transaction.

Becker, K. and F. Stalder (2009) 'Introduction', in K. Becker and F. Stalder (eds) *Deep Search: The Politics of Search Beyond Google*. Munich: Studienverlag & Transaction.

Chun, W.H.K. (2006) *Control and Freedom: Power and Paranoia in the Age of Fiber Optics*. Cambridge, MA: MIT Press.

Collier, S. (2009) 'Topologies of Power: Foucault's Analysis of Political Government Beyond "Governmentality"', *Theory, Culture & Society* 26(6): 79–109.

Cramer, F. (2005) 'Words Made Flesh: Code, Culture, Imagination'. Available at: http://pzwart.wdka.hro.nl/mdr/research/fcramer/wordsmadeflesh (consulted January 2010).

Deleuze, G. (1992) 'Postscript on the Societies of Control', *October* 59: 3–7.

Dick, P. (1968) *Do Androids Dream of Electric Sheep?* New York: Random House.

Foucault, M. (1973) *The Order of Things: An Archaeology of the Human Sciences*. New York: Vintage Books.

Foucault, M. (1979) *Discipline and Punish: Birth of the Prison*. New York: Vintage Books.

Foucault, M. (1980) 'Two-Lectures', in C. Gordon (ed.) *Power/Knowledge: Selected Interviews and Other Writings, 1972–1977*. New York: Pantheon Books.

Foucault, M. (1982) 'Afterward: The Subject and the Power', in H. Dreyfus and P. Rabinow (eds) *Michel Foucault: Beyond Structuralism and Hermeneutics*. Chicago: University of Chicago Press.

Foucault, M. (2003a) *Society Must Be Defended: Lectures at the Collège de France, 1975–1976*. New York: Picador.

Foucault, M. (2003b) *Ethics: Subjectivity and Truth*. New York: New Press.

Foucault, M. (2007) *Security, Territory, Population: Lectures at the Collège de France, 1977–1978*. New York: Palgrave Macmillan.

Foucault, M. (2008) *The Birth of Biopolitics: Lectures at the Collège de France, 1978–1979*. New York: Palgrave Macmillan.

Fuller, M. (2003) *Behind the Blip: Essays on the Culture of Software*. Brooklyn: Autonomedia.

Fuller, M. (ed.) (2008) *Software Studies: A Lexicon*. Cambridge, MA: MIT Press.

Galloway, A. (2004) *Protocol: How Control Exists After Decentralization*. Cambridge, MA: MIT Press.

Gandy, O. (1993) *The Panoptic Sort: A Political Economy of Personal Information*. Boulder: Westview Press.

Gibson, W. (1984) *Neuromancer*. New York: Penguin.

Graham, S. (2005) 'Software-Sorted Geographies', *Progress in Human Geography* 29(5): 1–19.

Hier, S. (2003) 'Probing the Surveillant Assemblage: On the Dialectics of Surveillance Practices as Processes of Social Control', *Surveillance & Society* 1(3): 399–411.

Koppel, M., S. Argamon and A. Shimoni (2002) 'Automatically Categorizing Written Texts by Author Gender', *Lit Linguist Computing* 17(4): 401–412.

Lessig, L. (2006) *Code: Version 2.0*. New York: Basic Books.

Lomell, H.M. (2004) 'Targeting the Unwanted: Video Surveillance and Categorical Exclusion in Oslo, Norway', *Surveillance Studies* 2(2/3).

Lyon, D. (2003a) *Surveillance After September 11*. Cambridge: Polity.

Lyon, D. (2003b) *Surveillance as Social Sorting: Privacy, Risk and Digital Discrimination*. New York: Routledge.

Macey, D. (2009) 'Rethinking Biopolitics, Race and Power in the Wake of Foucault', *Theory, Culture & Society* 26(6): 186–205.

McNay, L. (2009) 'Self as Enterprise: Dilemmas of Control and Resistance in Foucault's *The Birth of Biopolitics*', *Theory, Culture & Society* 26(6): 55–77.

Mosco, V. (1996) *The Political Economy of Communication*. London: SAGE.

Nakamura, L. (2002) *Cybertypes: Race, Ethnicity, and Identity on the Internet*. New York: Routledge.

Nye, J. (2002) 'Soft Power', *Foreign Policy* 80: 153–171.

Quantcast Corporation (2008) 'Quantcast Methodology Overview: Delivering an Actionable Audience Service'. Available at: http://www.quantcast.com/white-papers/quantcast-methodology.pdf (consulted February 2010).

Revel, J. (2009) 'Identity, Nature, Life: Three Biopolitical Deconstructions', *Theory, Culture & Society* 26(6): 45–54.

Rose, N. (1996) *Inventing Ourselves: Psychology, Power and Personhood*. Cambridge: Cambridge University Press.

Rose, N. (1999) *Powers of Freedom: Reframing Political Thought*. Cambridge: Cambridge University Press.

Rose, N. (2000) 'Government and Control', *The British Journal of Criminology* 40: 321–339.

Sunstein, C. (2007) *Republic 2.0*. Princeton: Princeton University Press.

Thacker, E. (2004) 'Foreword: Protocol is as Protocol Does', in *Protocol: How Control Exists After Decentralization*, ed. A. Galloway. Cambridge, MA: MIT Press.

Thrift, N. (2005) *Knowing Capitalism*. Thousand Oaks, CA: SAGE.

Turrow, J. (2006) *Niche Envy: Marketing Discrimination in the Digital Age*. Cambridge, MA: MIT Press.

Weber, I. and C. Castillo (2010) 'The Demographics of Web Search', *Proceedings of 2010 Conference on Research and Development in Information Retrieval (SIGIR)*.

Webster, W. (2009) 'CCTV Policy in the UK: Reconsidering the Evidence Base', *Surveillance Studies* 6(1): 10–22.

Witten, I. and E. Frank (2005) *Data Mining: Practical Machine Learning Tools and Techniques*. San Francisco: Morgan Kaufmann.

Yankelovich, D. and D. Meer (2006) 'Rediscovering Market Segmentation', *Harvard Business Review* 1).

Zadeh, L. (1994) 'What Is BISC?'. Available at: http.cs.berkeley.edu/projects/Bisc/bisc.welcome.html (consulted September 2010).

John Cheney-Lippold is Assistant Professor of American Culture at the University of Michigan. [email: jchl@umich.edu]