Knowledge Discovery and the Aesthetics of Big Data
Simulating the Herschel Observatory

John Mulligan and Adolfo Carvalho

The authors, an astrophysics student and a humanities researcher, developed a simulation of William Herschel's visual experience during his observational runs, bringing to life archival data produced by William and his sister Caroline Herschel, who are credited with having invented modern cosmology. From a media studies perspective, the use of intensive computational resources to produce boring, accurate, real-time simulations of William Herschel's observations helps us to confront our conflation of visual complexity with reality in the era of big data. At the intersection of data science, the history of science and media studies, the project proposes the aesthetics of boredom as a means of dwelling with the sense of big data as "big" relative to modes of knowledge production.

However, seeing is believing, and the dynamism of big-data visualizations can make users credulous. The enormity of the data coupled with ready access to it depends on a process of abstraction that pushes users toward the manipulation of highly processed data and away from the critical questioning of provenance and form that has been central to recent digital humanities work [2]. This uncritical embrace of visual complexity is most visible in the organicist aesthetics that attend dynamically generated visualizations and the related and ideologically burdened [3] assumption that computationally intensive representations are objective because humans cannot draw them.

Counterintuitively, it is precisely technical expertise that can be the cause of practitioners becoming too credulous of their own models. The big data sets now available on real-world phenomena are so massively parameterizable that they put the best analysts in the difficult position of seeing their code working to produce credible outputs. We have experienced the consequences of credulous expertise getting high on its own supply in the 2016 United States elections, in which both campaign managers and news consumers used analytics asymmetrically adopted by the different parties [4] with dynamic models of deliberately awesome scale: "Overnight . . . the team's computers run 400,000 simulations of the fall campaign in what amounts to a massive stress-test of the possibilities on Nov. 8" [5]. Less money was spent on obtaining fresh data because "analytics were just as good for tracking which candidate was ahead and by how much in each state. Plus, the analytics were quicker and much cheaper" [6].

Highly interactive models therefore have surprising, real effects, in large part due to the logic of control on which interactivity relies [7]. In the above case, dynamic interfaces obscured the provenance of uneven data sets in order to provide individualized users incredible control over the unified data (Fig. 1). This interactivity represented electoral campaigns in the genre of choose-your-own-adventure games, which allow a user to interactively choose branching narrative possibilities [8], the effect being a misrepresentation of the electorate as an aggregate to be managed rather than a polity to

ARTISTS’ ARTICLE

Knowledge Discovery and the Aesthetics of Big Data
Simulating the Herschel Observatory

John Mulligan and Adolfo Carvalho

John Mulligan [researcher], Rice University, Center for Research Computing, Houston, TX 77005, U.S.A. Email: jcm10@rice.edu.
Adolfo Carvalho [astronomy PhD student], California Institute of Technology, Pasadena, CA 91125, U.S.A. Email: carvalho@caltech.edu.

See https://direct.mit.edu/leon/article-pdf/54/3/282/1925086/leon_a_01832.pdf by guest on 31 July 2021
be persuaded or mobilized. By giving us more control over more variables, interactive interfaces crowd out of view the infrastructure they run on as well as the variables that take work to change.

Because the drive to maximize interactivity biases interface design toward control and mastery over entities that should instead be cared for and liberated, we propose a counter-aesthetic of boredom as better suited to the materiality and scale of big data. The Romantic period provides a useful site for such aesthetic experimentation. The Romantics first properly mathematized and industrialized the sciences [9]; we have attempted to realize these processes by bringing to life the slow time of data collection in the observatory of Caroline and William Herschel, the sibling pair who invented modern cosmology. We argue that a boring visual simulation best interfaces with the Herschels’ historically big data sets and suggest that boring big data visualizations help us to apprehend postmodernity’s informational infrastructure. We close by comparing these boring visualizations with contemporary alternatives to visual boredom: the complementary efforts of artists who reject the visual as such and the aesthetics of “noise,” whose problematic tendencies we trace in Romanticism’s experiments with visual complexity.

DATA IS BORING

Astronomy is the paradigmatic science in the intertwined, Romantic-era processes of mathematization, industrialization and specialization, and the paradigmatic Romantic astronomers belonged to the Herschel family. The Royal Astronomical Society (RAS), founded in 1820, represented the first break of a specialized discipline from the Royal Society’s umbrella organization; its founders, including Charles Babbage and John Herschel, understood this newly specialized science as an enterprise of industrial knowledge production [10], and their observational procedures set the standard for precision measurement in the nineteenth century’s increasingly mathematized sciences [11].

Caroline Herschel, John’s aunt, became one of the earliest recipients of the RAS’s gold medal in 1828 for her numerical catalogue of 2,510 deep-sky, nonstellar objects discovered in partnership with William Herschel, John’s father and Caroline’s brother, during their 1783–1802 sky surveys [12]. Her collected data provided William and later John with a trove of fascinating, irregular objects to observe and theorize about. The scalar properties of the raw data they gathered from, added to and corrected in extant catalogues resemble those of today’s “big” data sets. As Michael Hoskin argues, William “demonstrate[d] the latent possibilities of stellar statistics” by creatively synthesizing their data to hypothesize about the “evolution” and structure of cosmic systems [13].

This data-driven approach necessitated a refusal of what we call “visualization”: Caroline translated the beautiful Atlas Coelestis (Fig. 2) of John Flamsteed (1646–1719) into tables of small, numerically determined “zones” (Fig. 3) that they could meticulously and serially comb over with their high-powered telescopes [14]. Their workflow necessitated this informationization of observation: William observed at such high magnifications that it was hard work to determine where and at what he was looking, and Caroline observed indoors, windows open in a lit, second-story room, with records and instruments at hand in order to hastily process the observations her brother shouted from the telescope’s scaffolding.

When we consider the years it took Caroline to preprocess their sweep parameters and post-process their observational data into its final format, we see that the data is relatively big enough to incentivize the alienation of intellectual labor. Indeed, it is no coincidence that astronomical data’s regularity
Fig. 2. Illustration of the constellation Virgo, from John Flamsteed’s Atlas Coelestis, 1729. RAS C 2/10.1. (Photo courtesy of the Royal Astronomical Society)

Fig. 3. Caroline Herschel’s “zones” cutting across the constellation Virgo. RAS C 2/1.2. (Photo courtesy of the Royal Astronomical Society)
and scale has provided some of the best test cases for artificial intelligence research [15]. For example, an in-progress monograph by Wolfgang Steinicke [16] promises to give new insights into their work based on his database of the Herschel's observations and discoveries.

The aesthetics and epistemology of data science are implicated by this problem of automation/alienation (versus Shoshana Zuboff’s “information,” which we might think of as abstraction and autonomization) [17]. Consider the confirmation-bias problematic we alluded to in the introduction. Confirmation bias is both the weakness and the strength of “data science” methods like “knowledge discovery” and “machine learning.” It allows users of knowledge-discovery engines to find what they are looking for, whether that be in the consumer-use case of recommendation engines that “help users . . . discover connections” [18] or the use of big medical data to find correlations [19], or the habit of natural scientists to call their stunning visual simulations “realistic” (which is just to say that they look right) [20].

Because of the meticulous records the Herschels kept, the standardized vertical “sweeping” procedure they typically followed to cover a band of the sky, and the paradigmatic clockwork regularity of the phenomena they recorded, it is possible to simulate William's visual experience of any sweep. We therefore built a plug-in for the open-source planetarium software Stellarium capable of replaying William's view over any observation sweep (Fig. 4). We then ran these simulations on a Mac Mini, while Python scripts wrote screenshots to Rice's Research Data Facility, compiled these screenshots into low-frame-rate, high-definition videos, and posted these videos with archival image assets and metadata to social media [21] using Rice's virtual machine infrastructure. This allows Internet users to play back any Herschel observation sweep [22].

We have therefore attempted to make this historical interface minimally interactive. The most exciting thing that happens in a simulation is when the telescope reaches its peak or nadir and reverses directions. What users are seeing is the extremely boring, highly disciplined and single-visioned work that went into producing the highly reproducible data of positional astronomy. What we are simulating in the aggregate is the tense, on-the-clock regulation of visual attention that attends the industrialization and mathematization of the sciences that properly began in the Romantic era.

As Susan Leigh Star has influentially argued, an aesthetics of boredom is well attuned to the infrastructural dimensions of what Jean-François Lyotard called informational “teleculture” [23]: “This article is in a way a call to study boring things. Many aspects of infrastructure are singularly unexciting. . . . It takes some digging to unearth the dramas inherent in system design creating, to restore narrative to what appears to be dead lists” [24]. Digital humanists particularly have used Star's theorization of infrastructure to foreground the materiality of scholarly practice in a computerized culture [25].

The aesthetic of boredom has also been a favorite of experimental filmmakers interested in postmodern temporalities, as seen, for example, in Andy Warhol's Empire (1964), Michael Snow's Wavelength (1967) and Douglas Gordon's 24 Hour Psycho (1993). Closest to our own work is Chantal Akerman’s 1972 La Chambre (The Apartment): Over the

Fig. 4. Screenshot of the simulator’s sweep display function, showing some sweeps’ coverage in the constellation Virgo.
course of 15 minutes, a camera pans back and forth over a studio apartment with a single human in it (Fig. 5). Present to the real time of the film in a way that abstracts us from ourselves, we find ourselves reduced to a gaze that expands just to take up time, inhabiting a minor form of life that feels itself unspool frame by frame.

With regard to pleasurable boredom, our simulator flirts with what Robert Pfaller has called “interpassivity”: a delegation of consumption, above all to works of art that enjoy themselves for us [26]. With its minimal user input, its parallel Python screen-capture scripts that observe the simulations for us and its singularly unengaging social media presence, our interface offers the passive pleasure of not obligating users to interact—after all, the Herschels already observed all of this for us. With regard to infrastructural boredom, the Herschels’ innovative alienation of intellectual labor is exemplified in the fact that these simulations are both boring and tense, too slow and too fast: Viewers who would discover objects with William must remain attentive. This kind of hypnotic worrying demystifies modernity’s spatio-temporal infrastructure and, we hope, helps users feel just what a grind big data has always been.

In its alienated pleasures and pressures, then, our simulator aligns with Marxist theorizations of boredom as an objectification of subjectivity’s overdetermination in administered society. Boredom is “the everyday made manifest,” where the “everyday” is the industrially woven fabric of modern life, “a purely quantitative time . . . which is experienced as abstract, linear, sequential, predictable and monotonous” [27]. Insofar as positional astronomy’s data originally calibrated modern clock-time [28], this simulator attunes us to the rhythms of the Jacquard looms that mass-produce the everyday, what William Blake called the Looms of Enitharmon, an informationized cosmology that the Herschels were well acquainted with.

**CHANNEL-CHANGING AND CHANNEL-OVERLOADING RESPONSES**

Our silent, high-definition and minimally interactive video simulations of the Herschels’ observational sweeps are intended as a response to the complexification of the visual in big-data interface design. It is worth noting, however, that some artists have creatively chosen to challenge the reliance on the visual as such by experimenting with other sensory channels. These channel-changing responses offer up creative uses of the embodied sensorium through computer-controlled feedback loops such as haptic feedback [29], auditory soundscapes [30], synesthetic experiments such as Francisco López’s audience-blindfolding performance pieces [31] and gustatory happenings such as Lynn McCabe’s quasi-cannibalistic dinners [32]. These turns to embodiment align with the rejection of modern vision as a vehicle for rationalization, idealization and individualization [33]; boredom’s affinities with embodied experience suggest the possibilities for creative integrations of visual boredom with nonvisual excitement.

We should, finally, contrast our own approach to reclaiming the visual via boredom with another: via the overloading of the visual channel in order to interfere with the rationalizations of Cartesian visuospatiality. Most notably, in the “glitch” and “noise” aesthetics of experimental digital media [34], we find an appeal to sensory authenticity that attempts...
to liberate percept from concept and in doing so embraces a vitalism inherent in apparently nonabstracted senses [35].

This channel-overloading turn to the authenticity of physical perception, however, is not new, and we can read its limitations in the fate of the Romantic embrace of visual complexity. Consider John Constable’s 1816 Cenotaph to the Memory of Sir Joshua Reynolds (Color Plate A), which foregrounds landscape as a playground for the viewer’s visual imagination by unfolding its complex terrain with the textured lighting of chiaroscuro; critics have split on the politics of this aesthetic. William Galperin argued that such plays of light resisted abstraction by allowing in the phenomenal materiality of “the visible just enough to justify both the painter’s blindnesses and his insights” [36]; James Heffernan, to whom Galperin was in part responding, had previously argued that such an apparently material resistance to visually mediated ideality is in fact isomorphic with the aesthetics of transcendental idealism insofar as chiaroscuro’s play of light and shadow reflects back to us the free play of our visual imagination [37].

While Heffernan’s reading is pessimistic, it cannily presents this complexity as a product of abstraction. This perspective is indispensable in our current situation, where informational infrastructures are geared toward naturalizing the phenomenal richness of big data, specifically through responsive interactivity. To fetishize data sets’ organically emergent properties is to remain willfully ignorant of their various dependencies (e.g. software, hardware, culture)—and thereby to participate in precisely those pernicious mechanisms, treated above, that put models into negative feedback loops with the rest of reality, feedback loops in which much more than the models break when their predictions fail.

Acknowledgments

Rice University’s Humanities Research Center and the Andrew W. Mellon Foundation supported this project financially. Rice’s Center for Research Computing supplied infrastructural and technical support. Thanks to Melissa Bailar, Holly Holmes, Wolfgang Steinicke, Alexander Wolf, Stephen Bradshaw, Justin Hunter, Clinton Heider, Niki Kasumi Clements, Annie Lowe and Sian Prosser.

References and Notes

20 A search on Google Scholar for “realistic” AND (“numerical model” OR “numerical simulation”) yields numerous pertinent examples of this usage.
21 www.twitter.com/herschelman. The project data and code-base are available at the Rice data repository: https://doi.org/10.25611/7E9s-9R96 (accessed 12 February 2021).

Manuscript received 20 January 2019.

**JOHN MULLIGAN** works in humanities computing for the Center for Research Computing at Rice University in Houston. He writes essays on art, literature, mathematics, science and sometimes society, as well as software around humanistic questions.

**ADOLFO CARVALHO** is an astronomy PhD student at Caltech, primarily studying young stars. He interrupted his studies at Caltech to pursue a Fulbright Research Fellowship at the Universidad de Chile in Santiago, where he studied proto-planetary disks imaged by the Atacama Large Millimeter and Submillimeter Array. Along with his research, he is interested in science communication and data visualization.

Downloaded from http://direct.mit.edu/leon/article-pdf/54/3/282/1925086/leon_a_01832.pdf by guest on 31 July 2021
John Constable, Cenotaph to the Memory of Sir Joshua Reynolds, erected in the grounds of Coleorton Hall, Leicestershire by the late Sir George Beaumont, 1836. (Bequeathed by Miss Isabel Constable as the gift of Maria Louisa Isabel and Lionel Bicknell Constable, 1888. © The National Gallery, London.) (See the article in this issue by John Mulligan and Adolfo Carvalho.)